

ESTUDIOS

TRATADO SOBRE EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL DE LA UNIÓN EUROPEA

LORENZO COTINO HUESO
PERE SIMÓN CASTELLANO
DIRECTORES

INCLUYE LIBRO
ELECTRÓNICO

III ARANZADI

TRATADO SOBRE EL REGLAMENTO DE
INTELIGENCIA ARTIFICIAL DE LA UNIÓN EUROPEA

Directores

LORENZO COTINO HUESO

Catedrático de Derecho Constitucional de la Universitat de València. Valgrai

PERE SIMÓN CASTELLANO

*Profesor Titular de Derecho Constitucional Universidad Internacional de la Rioja –
UNIR*

TRATADO SOBRE EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL DE LA UNIÓN EUROPEA

Autores

ALESSANDRO MANTELERO	IGNACIO ALAMILLO DOMINGO
JACQUES ZILLER	EDUARD CHAVELI DONET
JUAN GUSTAVO CORVARÁN	PERE SIMÓN CASTELLANO
MARÍA VICTORIA CARRO	MARÍA LOZA CORERA
LORENZO COTINO HUESO	FRANCISCA RAMÓN FERNÁNDEZ
ALFONSO ORTEGA GIMÉNEZ	WILMA ARELLANO TOLEDO
ÁNGEL GÓMEZ DE ÁGREDA	ANTONIO MERCHÁN MURILLO
JESÚS JIMÉNEZ LÓPEZ	ESTRELLA DAVID GUTIÉRREZ
LEIRE ESCAJEDO	GUILLERMO LAZCOZ MORATINOS
MIGUEL ÁNGEL PRESNO LINERA	ANA ABA CATOIRA
LUIS MIGUEL GONZÁLEZ DE LA GARZA	MARCO EMILIO SÁNCHEZ ACEVEDO
FERNANDO MIRÓ LLINARES	IDOIA SALAZAR
MARIO SANTISTEBAN GALARZA	MIGUEL ÁNGEL LIÉBANAS
INIGO DE MIGUEL BERIAIN	JOSÉ ANTONIO CASTILLO
GAL·LA BARRACHINA NAVARRO	AGUSTÍ CERRILLO I MARTÍNEZ
ANDRÉS BOIX PALOP	JUAN CARLOS HERNÁNDEZ PEÑA
ROSA CERNADA BADÍA	F. JAVIER SEMPERE
VICENTE ÁLVAREZ GARCÍA	AURELIO LÓPEZ-TARRUELLA MARTÍNEZ
ADRIÁN PALMA ORTIGOSA	GABRIELE VESTRI

III ARANZADI

© Lorenzo Cotino Hueso, Pere Simón Castellano (Dirs.) y otros, 2024
© Editorial Aranzadi, S.A.U.

Editorial Aranzadi, S.A.U.
C/ Collado Mediano, 9
28231 Las Rozas (Madrid)
Tel: 91 602 01 82
e-mail: clienteslaley@aranzadilaley.es
<https://www.aranzadilaley.es>

Primera edición: octubre 2024

Depósito Legal: M-21792-2024
ISBN versión impresa con complemento electrónico: 978-84-1162-931-7
ISBN versión electrónica: 978-84-1162-930-0

Diseño, Preimpresión e Impresión: Editorial Aranzadi, S.A.U.
Printed in Spain

© Editorial Aranzadi, S.A.U. Todos los derechos reservados. A los efectos del art. 32 del Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba la Ley de Propiedad Intelectual, Editorial Aranzadi, S.A.U., se opone expresamente a cualquier utilización del contenido de esta publicación sin su expresa autorización, lo cual incluye especialmente cualquier reproducción, modificación, registro, copia, explotación, distribución, comunicación, transmisión, envío, reutilización, publicación, tratamiento o cualquier otra utilización total o parcial en cualquier modo, medio o formato de esta publicación.

Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra solo puede ser realizada con la autorización de sus titulares, salvo excepción prevista por la Ley. Dirijase a Cedro (Centro Español de Derechos Reprográficos, www.cedro.org) si necesita fotocopiar o escanear algún fragmento de esta obra.

El editor y los autores no asumirán ningún tipo de responsabilidad que pueda derivarse frente a terceros como consecuencia de la utilización total o parcial de cualquier modo y en cualquier medio o formato de esta publicación (reproducción, modificación, registro, copia, explotación, distribución, comunicación pública, transformación, publicación, reutilización, etc.) que no haya sido expresa y previamente autorizada.

El editor y los autores no aceptarán responsabilidades por las posibles consecuencias ocasionadas a las personas naturales o jurídicas que actúen o dejen de actuar como resultado de alguna información contenida en esta publicación.

EDITORIAL ARANZADI no será responsable de las opiniones vertidas por los autores de los contenidos, así como en foros, chats, u cualesquiera otras herramientas de participación. Igualmente, EDITORIAL ARANZADI se exime de las posibles vulneraciones de derechos de propiedad intelectual y que sean imputables a dichos autores.

EDITORIAL ARANZADI queda eximida de cualquier responsabilidad por los daños y perjuicios de toda naturaleza que puedan deberse a la falta de veracidad, exactitud, exhaustividad y/o actualidad de los contenidos transmitidos, difundidos, almacenados, puestos a disposición o recibidos, obtenidos o a los que se haya accedido a través de sus PRODUCTOS. Ni tampoco por los Contenidos prestados u ofertados por terceras personas o entidades.

EDITORIAL ARANZADI se reserva el derecho de eliminación de aquellos contenidos que resulten inveraces, inexactos y contrarios a la ley, la moral, el orden público y las buenas costumbres.

Nota de la Editorial: El texto de las resoluciones judiciales contenido en las publicaciones y productos de Editorial Aranzadi, S.A.U., es suministrado por el Centro de Documentación Judicial del Consejo General del Poder Judicial (Cendoj), excepto aquellas que puntualmente nos han sido proporcionadas por parte de los gabinetes de comunicación de los órganos judiciales colegiados. El Cendoj es el único organismo legalmente facultado para la recopilación de dichas resoluciones. El tratamiento de los datos de carácter personal contenidos en dichas resoluciones es realizado directamente por el citado organismo, desde julio de 2003, con sus propios criterios en cumplimiento de la normativa vigente sobre el particular, siendo por tanto de su exclusiva responsabilidad cualquier error o incidencia en esta materia.

Autores según orden de aparición, con sus créditos

Alessandro Mantelero

Profesor titular de Derecho Civil en el Politécnico de Turín y titular de la Cátedra Jean Monnet de Sociedades Digitales Mediterráneas y Derecho

Jacques Ziller

Catedrático de derecho público y de la Unión europea, Universidades Paris-1 Panthéon Sorbonne y Pavia

Juan Gustavo Corvarán

Director del Laboratorio de Innovación e Inteligencia Artificial de la Facultad de Derecho de la Universidad de Buenos Aires

María Victoria Carro

Doctoranda, Universidad de Génova. Directora de investigación, UBA IALAB

Lorenzo Cotino Hueso

Catedrático de Derecho Constitucional de la Universitat de València. Valgrai

Alfonso Ortega Giménez

Profesor Titular de Derecho internacional privado de la Universidad Miguel Hernández de Elche (Alicante)

Ángel Gómez de Ágreda

Doctor Universidad Politécnica de Madrid. Ministerio de Defensa de España. Odiseia

Jesús Jiménez López

Director del Consejo de Transparencia y Protección de datos de Andalucía

Leire Escajedo

Profesora Titular de Derecho Constitucional Universidad del País Vasco/EHU

Miguel Ángel Presno Linera

Catedrático de Derecho Constitucional de la Universidad de Oviedo

Luis Miguel González de la Garza

Profesor Titular de Derecho Constitucional UNED

Fernando Miró Llinares

Catedrático de Derecho Penal y Director Centro CRIMINA. Universidad Miguel Hernández de Elche

Mario Santisteban Galarza

Universidad del País Vasco

Inigo De Miguel Beriain

Ikerbasque research profesor. Investigador Universidad del País Vasco/Euskal Herriko Unibertsitatea. Miembro del Comité de Bioética de España

Gal·la Barrachina Navarro

Universitat de València

Andrés Boix Palop

Profesor Titular Derecho Administrativo Universitat de València

Rosa Cernada Badía

Profesora de Derecho Administrativo

Universidad Católica de Valencia San Vicente Mártir

Vicente Álvarez García

Catedrático de Derecho Administrativo Universidad de Extremadura

Adrián Palma Ortigosa

Profesor Ayudante Doctor del Departamento de Derecho Administrativo de la Universitat de València

Ignacio Alamillo Domingo

Doctor en Derecho

Eduard Chaveli Donet

Abogado especialista en Derecho Digital. Head of Consulting Strategy en Govertis, Part of Telefónica Tech

Pere Simón Castellano

Profesor Titular de Derecho Constitucional Universidad Internacional de la Rioja - UNIR

María Loza Corera

Doctora en Derecho. Lead Advisor en Govertis parte de Telefónica Tech

Profesora de la Universidad Internacional de La Rioja

Francisca Ramón Fernández

Catedrática de Derecho Civil de la Universitat Politècnica de València

Wilma Arellano Toledo

Doctora por la Universidad Complutense de Madrid. OdiseIA

Antonio Merchán Murillo

Doctor. Abogado. OdiseIA. Profesor ayudante doctor (acreditado a Prof. Titular) en la Universidad de Cádiz

Estrella David Gutiérrez

Profesora ayudante doctora de Derecho Constitucional Universidad Complutense de Madrid

Guillermo Lazcoz Moratinos

Centro de Investigación Biomédica en Red (CIBERER - ISCIII)

Instituto de Investigación Sanitaria Fundación Jiménez Díaz (IIS-FJD)

Ana Aba Catoira

Profesora Titular de Derecho Constitucional. Universidad de A Coruña

Marco Emilio Sánchez Acevedo

Abogado. Profesor Doctor e investigador Universidad Católica de Colombia

Idoia Salazar

Doctora. Profesora de la Universidad CEU San Pablo. Presidenta de Odiseia

Miguel Ángel Liébanas

Criminólogo experto en Sistemas Inteligentes. Odiseia. CEO de Human Trends

José Antonio Castillo

Doctor. Investigador Juan de la Cierva y Delegado de Protección de Datos de la Universidad de Granada

Agustí Cerrillo i Martínez

Catedrático de Derecho Administrativo de la Universitat Oberta de Catalunya

Juan Carlos Hernández Peña

Profesor Titular de Derecho administrativo

Universidad de Navarra

F. Javier Sempere

Director de Supervisión y Protección de Datos del Consejo General del Poder Judicial. Doctorando por CEU Escuela Internacional de Doctorado (CEINDO)

Aurelio López-Tarruella Martínez

Profesor Titular Derecho internacional privado Universidad de Alicante

Gabriele Vestri

Doctor en Derecho, Fundador y Presidente del Observatorio Sector Público e Inteligencia Artificial

Contenido general obra (por capítulos y autores)

Presentación e introducción a la obra

El Reglamento de inteligencia artificial y su contextualización mundial, desde Iberoamérica y en Europa

El Reglamento de inteligencia artificial: la respuesta del legislador europeo a los retos de la inteligencia artificial, por Alessandro Mantelero

El Convenio del Consejo de Europa de inteligencia artificial frente al Reglamento de la Unión Europea: dos instrumentos jurídicos muy diversos, por Jacques Ziller

El Reglamento de inteligencia artificial desde fuera de la Unión Europea: impulsos reguladores desde otras partes del mundo y una visión desde Iberoamérica, por Juan Gustavo Corvarán y María Victoria Carro

«Inteligencia artificial», ámbito territorial y alcance del Reglamento y su relación con la protección de datos

¿Qué es «inteligencia artificial» para el Reglamento? Análisis, delimitación y aplicaciones prácticas, por Lorenzo Cotino Hueso

El ámbito de aplicación territorial del Reglamento de inteligencia artificial, por Alfonso Ortega Giménez

La exclusión de los sistemas inteligencia artificial de seguridad nacional, defensa y militares del Reglamento y el Derecho aplicable, por Ángel Gómez de Ágreda

El Reglamento de inteligencia artificial y el Reglamento general de protección de datos, por Jesús Jiménez López

La inteligencia artificial prohibida o inaceptable para el Reglamento (artículo 5)

El reconocimiento biométrico en el Reglamento de inteligencia artificial: exenciones, prohibiciones y especialidades de alto riesgo, por Leire Escajedo

La prohibición de sistemas de inteligencia artificial que evalúan y clasifican a las personas a partir de datos que no guardan relación con el contexto donde se generaron y que provocan discriminaciones, por Miguel Ángel Presno Linera

El contenido de las llamadas «técnicas subliminales» y las vulnerabilidades de grupo específico de personas en el Reglamento de inteligencia artificial, por Luis Miguel González de la Garza

El resto de sistemas de inteligencia artificial prohibidos o inaceptables en el Reglamento, por Pere Simó Castellanos

Los sistemas de inteligencia artificial de alto riesgo: delimitación y análisis de algunos ámbitos

Alcance y delimitación de los sistemas de alto riesgo en el Reglamento de inteligencia artificial, por Lorenzo Cotino Hueso

La regulación de los sistemas policiales predictivos en el Reglamento Inteligencia Artificial, por Fernando Miró Llinares y Mario Santisteban Galarza

La aplicabilidad del Reglamento de inteligencia artificial al ámbito salud y especialidades respecto de su cumplimiento, por Inigo De Miguel Beriain

La aplicabilidad del Reglamento europeo de inteligencia artificial al ámbito de la Administración pública y servicios públicos y especialidades respecto de su cumplimiento: Especial atención a Anexo III y actuación administrativa y particularidades cumplimiento, por Gal·la Barrachina Navarro y Andrés Boix Palop

Grandes plataformas y sistemas de inteligencia artificial destinados a la influencia política: la intersección entre la «Ley de Servicios Digitales» y el Reglamento de inteligencia artificial desde la perspectiva del riesgo, por Rosa Cernada Badía

Régimen general aplicable a los sistemas de inteligencia artificial de alto riesgo

La aplicación de las normas armonizadas y de las especificaciones comunes en el ámbito de la inteligencia artificial (artículos 40 y 41 Reglamento), por Vicente Álvarez García

La evaluación de la conformidad en el diseño y producción de sistemas basados en IA en el contexto del «Nuevo Marco Legislativo», por Adrián Palma Ortigosa

Régimen general de obligaciones de proveedores y responsables del despliegue en el Reglamento de inteligencia artificial, por Adrián Palma Ortigosa

Sujetos y agentes en evaluaciones de conformidad (organismos notificados), por Ignacio Alamillo Domingo

Las obligaciones de los proveedores e implantadores de sistemas de alto riesgo

La evaluación de impacto de derechos fundamentales por quienes despliegan sistemas de inteligencia artificial en el Reglamento, por Eduard Chaveli Donet

Los sistemas de gestión de riesgos como obligación específica para los sistemas de inteligencia artificial de alto riesgo en el artículo 9 del Reglamento, por Pere Simón Castellano

Datos y gobernanza de datos y conexiones con principios protección de datos en el artículo 10 del Reglamento, por María Loza Corera

Sistemas de gestión de calidad, documentación técnica y conservación en el Reglamento, por Francisca Ramón Fernández

La obligación de conservar registros de los sistemas de alto riesgo en el Reglamento de inteligencia artificial, por Wilma Arellano Toledo y Antonio Merchán Murillo

La obligación de los proveedores de transparencia y comunicación de información a los implementadores en el artículo 13 del Reglamento, por Estrella David Gutiérrez

La vigilancia o supervisión humana en el artículo 14 del Reglamento de inteligencia artificial: ¿un mero requisito obligatorio para los sistemas de alto riesgo?, por Guillermo Lazcoz Moratinos

Precisión y solidez de los sistemas de inteligencia artificial de alto riesgo en el artículo 15 del Reglamento, por Ana Aba Catoira

Ciberseguridad en sistemas de inteligencia artificial de alto riesgo en el artículo 15 del Reglamento, por Marco Emilio Sánchez Acevedo

Vigilancia poscomercialización en los sistemas de inteligencia artificial de alto riesgo en el Reglamento. Descripción, medidas y casos de uso, por Idoia y Miguel Ángel Liébanas

Inteligencia artificial de uso general, sistemas que no son de alto riesgo y los sistemas del artículo 50

Inteligencia artificial de uso general, modelos fundacionales (y «Chat GPT») en el Reglamento de inteligencia artificial, por José Antonio Castillo

Códigos de conducta, sellos o certificaciones para los sistemas de inteligencia artificial que no son de alto riesgo (artículo 95 del Reglamento, por Lorenzo Cotino Hueso

El artículo 50 del Reglamento y las obligaciones de transparencia de los proveedores y responsables del despliegue de determinados sistemas de inteligencia artificial, por Agustí Cerrillo i Martínez

Sandbox, gobernanza, vigilancia, régimen sancionador, derechos y confidencialidad en el Reglamento

Sandbox, espacios controlados y pruebas en condiciones reales de sistemas de inteligencia artificial en el Reglamento. Medidas para PYMES, startups y micro empresas, por Lorenzo Cotino Hueso

La gobernanza y vigilancia del Reglamento de inteligencia artificial: autoridades de vigilancia del mercado, Comisión y las diversas entidades, por Juan Carlos Hernández Peña

El régimen sancionador en el Reglamento de inteligencia artificial, por F. Javier Sempere

Derecho a presentar una reclamación y derecho a una explicación. Vías de recurso para los particulares en el reglamento de inteligencia artificial, por Aurelio López-Tarruella Martínez

Acceso a documentación y confidencialidad en el Reglamento de inteligencia artificial, por Gabriele Vestri

Índice general

	<u>Página</u>
AUTORES SEGÚN ORDEN DE APARICIÓN, CON SUS CRÉDITOS.	7
CONTENIDO GENERAL OBRA (POR CAPÍTULOS Y AUTORES).	11
PRESENTACIÓN E INTRODUCCIÓN	47
 EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL Y SU CONTEXTUALIZACIÓN MUNDIAL, DESDE IBEROAMÉRICA Y EN EUROPA 	
EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL: LA RES- PUESTA DEL LEGISLADOR EUROPEO A LOS RETOS DE LA INTELIGENCIA ARTIFICIAL	53
I. INTRODUCCIÓN.....	53
II. LA PERSPECTIVA EUROPEA	53
III. LA MODULACIÓN DEL LLAMADO ENFOQUE BASADO EN EL RIESGO EN UNA LEGISLACIÓN DE PRIMERA GE- NERACIÓN	58
IV. CONCLUSIONES.....	65
 EL CONVENIO DEL CONSEJO DE EUROPA DE INTELIGEN- CIA ARTIFICIAL FRENTE AL REGLAMENTO DE LA UNIÓN EUROPEA: DOS INSTRUMENTOS JURÍDICOS MUY DIVER- SOS	 67
I. EL CONTENIDO DEL PROYECTO DE CONVENIO MAR- CO DEL CONSEJO DE EUROPA.....	69

	<u>Página</u>
II. LAS RAZONES DE UN TRATADO DEL CONSEJO DE EUROPA SOBRE INTELIGENCIA ARTIFICIAL.....	72
III. EL INSTRUMENTO DEL CONVENIO MARCO DE FRENTE AL INSTRUMENTO DEL REGLAMENTO	75
IV. LOS LÍMITES DERIVADOS DE LAS RESPECTIVAS COMPETENCIAS DEL CONSEJO DE EUROPA Y DE LA UNIÓN EUROPEA.....	77
V. LA NECESIDAD DE RATIFICAR EL TRATADO DEL CONSEJO DE EUROPA DE FRENTE A LA APLICABILIDAD DIRECTA DEL REGLAMENTO DE LA UNIÓN EUROPEA	82
VI. A MODO DE CONCLUSIÓN.....	83
EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL DESDE FUERA DE LA UNIÓN EUROPEA: IMPULSOS REGULADORES DESDE OTRAS PARTES DEL MUNDO Y UNA VISIÓN DESDE IBEROAMÉRICA.....	85
I. INTRODUCCIÓN.....	85
II. SALIR DE LA OLLA A TIEMPO Y OTROS DESAFÍOS DE LA REGULACIÓN	87
1. Unidad vs. fragmentación. ¿Enfoque «horizontal» o «vertical»?	89
2. Segundo. Obligatoriedad vs. voluntariedad y el fortalecimiento de la colaboración	92
3. Tercero. Un enfoque basado en riesgos	95
III. IMPULSOS REGULADORES DESDE OTRAS PARTES DEL MUNDO.....	97
2. Canadá	99
IV. UNA PERSPECTIVA DESDE LATINOAMÉRICA.....	100
1. México	100
2. Chile	102
3. Uruguay	103
4. Brasil	104
5. Colombia	106

	<u>Página</u>
6. Argentina.....	107
V. CONCLUSIÓN.....	108

«INTELIGENCIA ARTIFICIAL», ÁMBITO TERRITORIAL Y ALCANCE DEL REGLAMENTO Y SU RELACIÓN CON LA PROTECCIÓN DE DATOS

¿QUÉ ES «INTELIGENCIA ARTIFICIAL» PARA EL REGLAMENTO? ANÁLISIS, DELIMITACIÓN Y APLICACIONES PRÁCTICAS.....	113
I. LA IMPORTANCIA DEL CONCEPTO DE INTELIGENCIA ARTIFICIAL EN EL REGLAMENTO	113
II. LAS DIVERSAS DEFINICIONES DE INTELIGENCIA ARTIFICIAL SE HAN BARAJADO.....	114
III. LA DEFINICIÓN DEL ARTÍCULO 3 REGLAMENTO Y SUS COMPONENTES: TÉCNICAS, AUTONOMÍA, ADAPTACIÓN, ENTRADAS Y SALIDAS Y CONTEXTO	117
IV. EJEMPLOS DE SISTEMAS QUE SÍ QUE SON O NO INTELIGENCIA ARTIFICIAL.....	120
EL ÁMBITO DE APLICACIÓN TERRITORIAL DEL REGLAMENTO DE INTELIGENCIA ARTIFICIAL.....	123
I. PLANTEAMIENTO	123
II. APLICACIÓN DEL ARTÍCULO 2.1 POR LOS OPERADORES ECONÓMICOS.....	126
III. APLICACIÓN DEL ARTÍCULO 2.1 POR LAS AUTORIDADES NACIONALES COMPETENTES	129
1. Necesidad de cooperación internacional	129
2. Armonización de estándares regulatorios	130
3. Interacción con el Derecho internacional	131

	<u>Página</u>
IV. APLICACIÓN DEL ARTÍCULO 2.1. POR LOS TRIBUNALES DE JUSTICIA	131
V. APLICACIÓN EXTRATERRITORIAL DEL REGLAMENTO EUROPEO DE INTELIGENCIA ARTIFICIAL.....	133
VI. REFLEXIÓN FINAL	136
LA EXCLUSIÓN DE LOS SISTEMAS INTELIGENCIA ARTIFICIAL DE SEGURIDAD NACIONAL, DEFENSA Y MILITARES DEL REGLAMENTO Y EL DERECHO APLICABLE	137
I. INTRODUCCIÓN.....	137
II. CÓDIGOS ÉTICOS PARA INTELIGENCIA ARTIFICIAL DE USO GENERAL	139
III. USO DUAL DE LAS TECNOLOGÍAS.....	140
IV. EL DECÁLOGO DE PRINCIPIOS ÉTICOS DEL CCW PARA SISTEMAS DE ARMAS AUTÓNOMOS LETALES	141
V. DIFERENCIAS SIGNIFICATIVAS ENTRE LOS USOS CIVILES Y MILITARES DE LA INTELIGENCIA ARTIFICIAL.....	143
VI. APLICABILIDAD DEL DERECHO INTERNACIONAL A LOS SISTEMAS DOTADOS DE INTELIGENCIA ARTIFICIAL.....	144
VII. RESPONSABILIDAD Y CONTROL HUMANO SIGNIFICATIVO	145
VIII. USO DE LA INTELIGENCIA ARTIFICIAL EN SEGURIDAD Y DEFENSA SEGÚN EL CONVENIO MARCO SOBRE INTELIGENCIA ARTIFICIAL, DERECHOS HUMANOS, DEMOCRACIA Y ESTADO DE DERECHO	147
IX. CONCLUSIONES.....	147
EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL Y EL REGLAMENTO GENERAL DE PROTECCIÓN DE DATOS	149
I. INTRODUCCIÓN.....	149
II. SISTEMAS DE INTELIGENCIA ARTIFICIAL Y TRATAMIENTO DE DATOS PERSONALES	150

	<u>Página</u>
1. Sistemas de Inteligencia Artificial	150
3. Tratamiento de datos personales y SIAs	153
4. Identificación de responsables de tratamiento de datos personales en los SIAs	156
III. LA RELACIÓN DE REGLAMENTO DE INTELIGENCIA ARTIFICIAL Y EL REGLAMENTO GENERAL DE PROTECCIÓN DE DATOS	161
1. Necesidad de un marco de relación entre ambos cuerpos normativos	161
2. Aplicación del Reglamento sin perjuicio de la aplicación del RGPD	164
3. El Reglamento como «lex specialis» con fines de policía	165
4. Supuestos específicos de base jurídica de tratamiento de datos personales en el Reglamento	166
5. Autoridades independientes de control en materia de protección de datos personales	170
6. Vigilancia humana y decisiones individuales automatizadas: artículo 22 RGPD y el Reglamento	172
7. El derecho a una explicación. Decisiones individuales en el contexto de determinados SIA de AR y Artículo 22 RGPD	173
8. Colaboración del Reglamento en el cumplimiento del RGPD	176
9. Limitaciones a la colaboración por el espacio inicial de regulación	178
IV. REFLEXIONES FINALES	178

LA INTELIGENCIA ARTIFICIAL PROHIBIDA O INACEPTABLE PARA EL REGLAMENTO (ARTÍCULO 5)

EL RECONOCIMIENTO BIOMÉTRICO EN EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL: EXENCIONES, PROHIBICIONES Y ESPECIALIDADES DE ALTO RIESGO.....	183
I. EL RECONOCIMIENTO BIOMÉTRICO EN EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL: ASPECTOS CLAVE DE SU REGULACIÓN.....	183
1. El enfoque desde el que se regula el reconocimiento biométrico en el Reglamento	183
2. Antecedentes relevantes: la distancia entre el Libro Blanco y las Resoluciones del PE sobre algunas modalidades de reconocimiento biométrico.....	186
3. Esquema regulatorio de las tecnologías de reconocimiento biométrico en el Reglamento.....	188
II. ALGUNOS CONCEPTOS CLAVE EN LA TIPIFICACIÓN DE LAS MODALIDADES DE RECONOCIMIENTO BIOMÉTRICO EN EL ENUNCIADO DEL REGLAMENTO: DATOS BIOMÉTRICOS Y VERIFICACIÓN BIOMÉTRICA.....	190
1. ¿Una nueva noción de «datos biométricos» para abarcar con claridad las biometrías no singularizantes?.....	190
1.1. <i>La situación previa a la aprobación del Reglamento: la noción de datos biométricos del RGPD versus la noción científico-técnica.....</i>	<i>190</i>
1.2. <i>La necesidad de una noción funcional que abarque los datos biométricos con y sin potencial identificante, sean personales o no</i>	<i>192</i>
1.3. <i>Evolución de la noción de datos biométricos en el proceso de elaboración del Reglamento</i>	<i>195</i>
2. Mínimo común de todos los sistemas de reconocimiento biométrico.....	196
2.1. <i>Sistemas de reconocimiento biométrico.....</i>	<i>196</i>
2.2. <i>Fuentes de datos, datos en crudo y patrones biométricos.....</i>	<i>197</i>

	<u>Página</u>
2.3. <i>Utilidades biométricas que identifican, utilidades que no</i>	198
2.4. <i>Importantes limitaciones de los sistemas de reconocimiento biométrico: científico-técnicas, de arquitectura y ético sociales</i>	201
3. La verificación biométrica, ¿fuera del Reglamento?	204
4. La identificación no remota, ¿en el limbo?	206
5. ¿Queda abarcado el cribado biométrico?	208
III. SISTEMAS DE RECONOCIMIENTO BIOMÉTRICO AFECTADOS POR LAS PROHIBICIONES Y RESTRICCIONES DEL ARTÍCULO 5 DE REGLAMENTO: EVALUACIÓN SOCIAL, PREDICCIÓN DE PELIGROSIDAD CRIMINAL, AMPLIACIÓN DE BASES DE RECONOCIMIENTO FACIAL, INFERENCIA DE EMOCIONES EN ALGUNOS CONTEXTOS Y CATEGORIZACIONES SOBRE DATOS ESPECIALMENTE PROTEGIDOS EN EL ART. 9 RGPD	208
1. La sistemática del artículo 5 en lo que se refiere al reconocimiento biométrico	208
1.1. <i>La propuesta inicial de la Comisión</i>	209
1.2. <i>Toma de postura del Parlamento</i>	209
1.3. <i>Toma de postura del Consejo y texto finalmente aprobado</i>	210
2. Biometrías afectadas por las prohibiciones de las letras C, D y E del art. 5.1. Reglamento: evaluaciones de la personalidad con finalidades de puntuación ciudadana, predicción de riesgo de cometer delitos y ampliación de las bases de reconocimiento facial	211
3. Reconocimiento biométrico de emociones en los lugares de trabajo y en los centros educativos, exceptuando los casos en que se persigan motivos médicos o de seguridad (art. 5.1. f). Concepto de reconocimiento de emociones en el Reglamento	212
4. Categorización biométrica con el fin deducir o inferir su raza, opiniones políticas, afiliación sindical, convicciones religiosas o filosóficas, vida sexual u orientación sexual (art.5.1 G)	215

IV.	SISTEMAS DE RECONOCIMIENTO BIOMÉTRICO AFECTADOS POR LAS PROHIBICIONES Y RESTRICCIONES DEL ARTÍCULO 5 REGLAMENTO (Y II): LA IDENTIFICACIÓN BIOMÉTRICA REMOTA «EN TIEMPO REAL» EN ESPACIOS DE ACCESO PÚBLICO CON FINES DE GARANTÍA DEL CUMPLIMIENTO DEL DERECHO	216
1.	Preocupación global por la identificación biométrica remota en espacios de acceso público	216
2.	Interpretación de los conceptos clave de la prohibición del art. 5.1.h).....	218
2.1.	<i>Identificación biométrica remota en tiempo real y diferido</i>	218
2.2.	<i>Escenario operativo y misión: espacios de acceso público y fines de garantía del cumplimiento del Derecho</i>	219
2.3.	<i>Excepcionable en la medida que el uso sea estrictamente necesario para alcanzar uno o varios de los siguientes objetivos</i>	220
3.	Los apartados 2 a 8.....	221
VI.	LOS GRANDES SISTEMAS DE RECONOCIMIENTO OPERATIVOS ANTES DE LA ENTRADA EN VIGOR DEL REGLAMENTO (ART. 111 Y ANEXO X).....	225
1.	El art. 111 y el anexo X del Reglamento	226
2.	Los Reglamentos de Interoperabilidad	227
3.	Sistemas de reconocimiento biométrico comprendidos en el anexo X: algunos datos relevantes	230
VII.	REFLEXIONES FINALES.....	233
	LA PROHIBICIÓN DE SISTEMAS DE INTELIGENCIA ARTIFICIAL QUE EVALÚAN Y CLASIFICAN A LAS PERSONAS A PARTIR DE DATOS QUE NO GUARDAN RELACIÓN CON EL CONTEXTO DONDE SE GENERARON Y QUE PROVOCAN DISCRIMINACIONES	237
I.	INTRODUCCIÓN.....	237
II.	EL SISTEMA DE CRÉDITO SOCIAL CHINO	240

	<u>Página</u>
III. EL DESARROLLO DE LOS SISTEMAS DE CALIFICACIÓN COMO UNA VÍA DE EXPANSIÓN DEL CAPITALISMO DE VIGILANCIA.....	242
IV. LA PROHIBICIÓN DE DETERMINADOS SISTEMAS QUE EVALÚAN O CLASIFICAN A LAS PERSONAS FÍSICAS.....	247
EL CONTENIDO DE LAS LLAMADAS «TÉCNICAS SUBLIMINALES» Y LAS VULNERABILIDADES DE GRUPO ESPECÍFICO DE PERSONAS EN EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL.....	251
I. INTRODUCCIÓN.....	251
II. EVOLUCIÓN DE LA TRAMITACIÓN Y CONTENIDO.....	252
III. ¿EN QUÉ CONSISTEN LAS TÉCNICAS SUBLIMINALES?..	260
IV. UNA INTELIGENCIA ARTIFICIAL QUE LO PROCESA TODO	267
V. TECNOLOGÍAS ADICTIVAS. LA ACTUACIÓN DE LA IA SOBRE COLECTIVOS, LOS MENORES, LOS JÓVENES Y OTROS COLECTIVOS.....	270
VI. CONCLUSIONES.....	273
EL RESTO DE LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL PROHIBIDOS O INACEPTABLES EN EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL	275
I. INTRODUCCIÓN.....	275
II. DEFINICIONES Y TERMINOLOGÍA: TECNOLOGÍAS DE CATEGORIZACIÓN BIOMÉTRICA, EMOCIONES, DATOS BIOMÉTRICOS Y RECONOCIMIENTO FACIAL	277
III. PRÁCTICAS INACEPTABLES: OBJETO Y CONTENIDO DE LA PROHIBICIÓN	280
1. Reconocimiento fácil vía «scraping» o extracción no selectiva de imágenes faciales en Internet y circuito cerrado de televisión.....	281
2. La prohibición del uso de sistemas de inteligencia artificial para inferir emociones en el Reglamento	282

<p>3. El uso de sistemas de inteligencia artificial de categorización biométrica que clasifiquen individualmente a las personas físicas sobre la base de sus datos biométricos para deducir o inferir su raza, opiniones políticas, afiliación sindical, convicciones religiosas o filosóficas, vida sexual u orientación sexual.....</p> <p>IV. LA COEXISTENCIA DE LA REGULACIÓN PREVISTA EN EL REGLAMENTO CON LA NORMATIVA DE PROTECCIÓN DE DATOS: UNA REGULACIÓN SUPERPUESTA Y, ¿COMPATIBLE?.....</p>	<p>283</p> <p>285</p>
--	-------------------------------------

LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL DE ALTO RIESGO: DELIMITACIÓN Y ANÁLISIS DE ALGUNOS ÁMBITOS

<p>ALCANCE Y DELIMITACIÓN DE LOS SISTEMAS DE ALTO RIESGO EN EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL.....</p> <p>I. LA CONDICIÓN DE SISTEMA DE ALTO RIESGO ES ESENCIAL PARA EL REGLAMENTO</p> <p>II. UNA ADVERTENCIA: LA REGULACIÓN DE SISTEMAS COMO DE ALTO RIESGO POR EL REGLAMENTO NO IMPLICA SU HABILITACIÓN LEGAL.....</p> <p>III. LOS SISTEMAS INTELIGENCIA ARTIFICIAL EN PRODUCTOS PELIGROSOS DEL ANEXO I.....</p> <p>1. El sistema inteligencia artificial como componente de seguridad o producto de productos sometidos a una «evaluación de conformidad» por un tercero.....</p> <p>2. El proveedor debe conocer razonablemente si su producto puede ser del Anexo I.....</p> <p>IV. LOS SISTEMAS DE ALTO RIESGO QUE PERSIGUEN LOS FINES DEL ANEXO III</p>	<p>291</p> <p>291</p> <p>293</p> <p>294</p> <p>294</p> <p>297</p> <p>298</p>
--	---

	<u>Página</u>
1. Sistemas que tengan una influencia sustancial en la toma de decisiones para las finalidades del anexo III....	298
2. Será siempre de alto riesgo la elaboración de perfiles con inteligencia artificial para fines del Anexo III.....	302
3. Las finalidades de los sistemas de alto riesgo del Anexo III	303
4. Presunción de que el sistema inteligencia artificial que persigue fines del Anexo III sí que es de alto riesgo. Especiales obligaciones y actuaciones.....	309
5. Cuando el implementador altera un sistema y pasa a tener una finalidad de alto riesgo.....	310
V. EL PAPEL DE LA COMISIÓN, CRITERIOS, ACTOS DELEGADOS, ACTUALIZACIÓN Y MODIFICACIÓN DE LOS SISTEMAS DE ALTO RIESGO, EN ESPECIAL, DEL ANEXO III	311
VI. RECAPITULACIÓN Y CONCLUSIONES.....	312
LA REGULACIÓN DE LOS SISTEMAS POLICIALES PREDICTIVOS EN EL REGLAMENTO INTELIGENCIA ARTIFICIAL	315
I. INTRODUCCIÓN.....	315
II. SISTEMAS POLICIALES PREDICTIVOS E INTELIGENCIA ARTIFICIAL.....	317
1. Organización policial en tiempos de digitalización y la «mal llamada» policía predictiva.....	317
2. Los riesgos éticos de la policía predictiva (y los que añaden el uso de inteligencia artificial)	320
III. LA REGULACIÓN DE LA POLICÍA PREDICTIVA EN EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL	326
1. Evolución de la regulación del uso de sistemas policiales predictivos en el proceso legislativo del Reglamento de inteligencia artificial	326
2. Sistemas predictivos policiales prohibidos en el Reglamento de inteligencia artificial	330
3. Policía predictiva «de alto riesgo» y sus implicaciones.	333

	<u>Página</u>
4. Conclusiones	334
 LA APLICABILIDAD DEL REGLAMENTO DE INTELIGENCIA ARTIFICIAL AL ÁMBITO SALUD Y ESPECIALIDADES RESPECTO DE SU CUMPLIMIENTO	 337
I. INTRODUCCIÓN.....	337
II. REGULACIÓN DE LOS DISPOSITIVOS SANITARIOS EN EL REGLAMENTO Y SU CONSIDERACIÓN DE ALTO RIESGO POR EL ANEXO I O III	338
1. Análisis preliminar: la regulación de los dispositivos sanitarios que incorporan inteligencia artificial en el Reglamento	338
2. Sistemas que son de alto riesgo de acuerdo con lo dispuesto en el Anexo III.....	338
3. Sistemas que pueden ser o no de alto riesgo según lo dispuesto en el Anexo I	340
III. LA REGULACIÓN DE LOS PRODUCTOS SANITARIOS: LAS ESTIPULACIONES DEL MDR Y EL IVDR	341
1. Los productos sanitarios. Una caracterización	341
2. Clases de productos sanitarios y exigencias de supervisión inherentes a cada tipo según el MDR.....	343
3. El IVDR y los dispositivos de diagnóstico in vitro.....	346
4. Las excepciones al régimen general de los sistemas incluidos en el Anexo III	346
IV. RECAPITULACIÓN.....	348
 LA APLICABILIDAD DEL REGLAMENTO DE INTELIGENCIA ARTIFICIAL AL ÁMBITO DE LA ADMINISTRACIÓN PÚBLICA Y SERVICIOS PÚBLICOS Y ESPECIALIDADES RESPECTO DE SU CUMPLIMIENTO: ESPECIAL ATENCIÓN A ANEXO III Y ACTUACIÓN ADMINISTRATIVA Y PARTICULARIDADES CUMPLIMIENTO	 351
I. INTRODUCCIÓN: EL PLANTEAMIENTO DEL REGLAMENTO Y LA PROYECCIÓN DE SUS CONTROLES Y GARANTÍAS SOBRE LA ACTUACIÓN DE LOS PODERES PÚBLICOS.....	351

	<u>Página</u>
1. Panorámica general: orientación básica y aplicación a la acción de los poderes públicos del Reglamento.....	351
2. Integración del control de la actividad automatizada y algorítmica, y del uso de inteligencia artificial, por parte de los poderes públicos con las normas de Derecho interno y algunos de sus problemas y carencias	356
II. EVOLUCIÓN, TRAMITACIÓN Y CONTENIDO FINAL DE LOS PRECEPTOS DEL REGLAMENTO QUE AFECTAN EN MAYOR MEDIDA A LOS PODERES PÚBLICOS	359
1. Consideraciones generales que pueden proyectarse singularmente sobre los poderes públicos en su uso de inteligencia artificial	359
2. Afección a las Administraciones públicas y a los poderes públicos de los usos prohibidos de inteligencia artificial establecidos en el art. 5 Reglamento	361
3. Sobre las precauciones del art. 6 Reglamento respecto a los usos de alto riesgo en relación a las obligaciones que se derivan de los mismos para los poderes públicos	363
4. Proyección de la delimitación de los usos de alto riesgo según el Anexo III sobre el empleo de inteligencia artificial por parte de los poderes públicos.....	365
5. Proyección de normas del Reglamento sobre actuaciones administrativas	366
III. ALGUNAS CONCLUSIONES SOBRE LA APLICACIÓN DEL REGLAMENTO AL SECTOR PÚBLICO	369
GRANDES PLATAFORMAS Y SISTEMAS DE INTELIGENCIA ARTIFICIAL DESTINADOS A LA INFLUENCIA POLÍTICA: LA INTERSECCIÓN ENTRE LA «LEY DE SERVICIOS DIGITALES» Y EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL DESDE LA PERSPECTIVA DEL RIESGO	373
I. INTRODUCCIÓN. FUNDAMENTO DEL TRATAMIENTO ESPECÍFICO DE LAS PLATAFORMAS DIGITALES Y LOS SISTEMAS DE INFLUENCIA POLÍTICA EN EL REGLAMENTO	373

	<u>Página</u>
II. UNA BREVE MIRADA AL « <i>ITER LEGIS</i> » DEL REGLAMENTO EN LA REGULACIÓN DE LAS GRANDES PLATAFORMAS Y SISTEMAS DE INFLUENCIA POLÍTICA	376
1. Atención específica del Reglamento a los sistemas de inteligencia artificial para la influencia política	377
2. Los sistemas de recomendación algorítmica: tratamiento en la propuesta de inteligencia artificial versión Parlamento y su fundamentación	379
III. LA LÓGICA DEL RIESGO EN LAS PLATAFORMAS DE GRAN TAMAÑO: LA COMPLEMENTARIEDAD ENTRE LA DSA Y EL REGLAMENTO.....	380
IV. ESPECIALIDADES DE LA APLICACIÓN DEL REGLAMENTO EN LAS GRANDES PLATAFORMAS Y SISTEMAS DE IA PARA LA INFLUENCIA POLÍTICA	384
1. Garantías en materia de gestión de riesgos	385
2. Garantías en materia de transparencia algorítmica: explicabilidad vs. opacidad	387
3. Garantías procedimentales	390
4. Garantías orgánicas y gobernanza digital	391
V. CONCLUSIONES	393

RÉGIMEN GENERAL APLICABLE A LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL DE ALTO RIESGO

LA APLICACIÓN DE LAS NORMAS ARMONIZADAS Y DE LAS ESPECIFICACIONES COMUNES EN EL ÁMBITO DE LA INTELIGENCIA ARTIFICIAL (ARTÍCULOS 40 Y 41 REGLAMENTO)..	397
I. INTRODUCCIÓN	397
1. La regulación de la inteligencia artificial a través de la técnica armonizadora del nuevo enfoque	397
2. Una breve introducción a los elementos básicos de la técnica armonizadora del nuevo enfoque aplicados a la inteligencia artificial	399

	<u>Página</u>
II. LAS NORMAS ARMONIZADAS	401
1. Una cuestión previa básica: las normas armonizadas tienen la naturaleza jurídica de Derecho comunitario...	401
2. La distinción entre las normas europeas y las normas armonizadas europeas	402
3. La intervención de los organismos europeos de normalización y de la Comisión en el procedimiento de elaboración de las normas armonizadas	404
4. Las previsiones sobre las normas armonizadas durante el proceso de tramitación de la propuesta de Reglamento efectuadas por la Comisión, por el Consejo y por el Parlamento Europeo.....	407
5. Las normas armonizadas en el texto definitivo del Reglamento.....	410
6. Los problemas de la aplicación de las técnicas normalizadoras a la regulación de la inteligencia artificial en la Unión Europea	411
7. Los resultados actuales del trabajo de normalización en inteligencia artificial	415
III. LAS ESPECIFICACIONES COMUNES	417
1. Unas ideas iniciales sobre su concepto	417
2. La evolución de la regulación de la figura de las especificaciones comunes durante la tramitación de la propuesta de Reglamento: desde el proyecto de la Comisión hasta las enmiendas del Parlamento Europeo, pasando por el texto transaccional del Consejo.....	418
3. Los elementos esenciales que configuran las especificaciones comunes en el Reglamento	421
4. Otras soluciones técnicas equivalentes a las ofrecidas por las normas armonizadas y por las especificaciones comunes.....	425
 LA EVALUACIÓN DE LA CONFORMIDAD EN EL DISEÑO Y PRODUCCIÓN DE SISTEMAS BASADOS EN INTELIGENCIA ARTIFICIAL EN EL CONTEXTO DEL «NUEVO MARCO LEGISLATIVO».	 427

	<u>Página</u>
I. INTRODUCCIÓN.....	427
II. EVOLUCIÓN, TRAMITACIÓN Y CONTENIDO FINAL DE LOS ARTÍCULOS DEL REGLAMENTO DE LA INTELIGENCIA ARTIFICIAL IMPLICADOS.....	428
III. LA EVALUACIÓN DE LA CONFORMIDAD EN LA LEGISLACIÓN DE LA UNIÓN EUROPEA.....	429
IV. LAS FORMAS DE EVALUACIÓN DE LA CONFORMIDAD EN EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL....	431
1. La autoevaluación de conformidad.....	431
2. La evaluación realizada por un organismo notificado ...	432
V. LA EVALUACIÓN DE LA CONFORMIDAD EN FUNCIÓN DEL TIPO DE SISTEMA DE INTELIGENCIA ARTIFICIAL..	434
1 Sistemas de Inteligencia artificial cuya finalidad es considerada de alto riesgo (finalidades de alto riesgo).....	435
2. Sistemas de Inteligencia artificial de productos o componentes de seguridad de productos considerados de alto riesgo	436
VI. LOS ORGANISMOS NOTIFICADOS ENCARGADOS DE REALIZAR LA EVALUACIÓN DE LA CONFORMIDAD DE LOS DIFERENTES SISTEMAS DE INTELIGENCIA ARTIFICIAL	440
VII. SUPUESTOS DONDE NO ES NECESARIO REALIZAR LA EVALUACIÓN DE LA CONFORMIDAD O EXISTEN PRESUNCIONES DE CONFORMIDAD DE SU CUMPLIMIENTO	442
1. Autorización previa de puesta en el mercado de sistemas de inteligencia artificial.....	442
2. La puesta en el mercado del sistema de inteligencia artificial sin autorización previa.....	443
3. Las exenciones de evaluación de la conformidad de los productos sometidos a legislación de armonización	444
4. Las presunciones de Conformidad	444
VIII. REFLEXIONES SOBRE LA REGULACIÓN DE LA EVALUACIÓN DE LA CONFORMIDAD ESTABLECIDA EN EL REGLAMENTO	444

	<u>Página</u>
IX. CONCLUSIONES.....	445
RÉGIMEN GENERAL DE OBLIGACIONES DE PROVEEDORES Y RESPONSABLES DEL DESPLIEGUE EN EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL	447
I. INTRODUCCIÓN.....	447
II. EVOLUCIÓN, TRAMITACIÓN Y CONTENIDO FINAL DE LOS ARTÍCULOS DEL REGLAMENTO IMPLICADOS	448
III. LOS OPERADORES PRESENTES DURANTE LA CADENA DE VALOR DE LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL EN EL REGLAMENTO.....	448
1. El proveedor y sus obligaciones	448
2. El responsable del despliegue y sus obligaciones	450
2.1. <i>Obligaciones relacionadas con el cumplimiento de los requisitos esenciales del Reglamento.....</i>	<i>450</i>
2.2. <i>Otras obligaciones derivadas del cumplimiento del Reglamento.....</i>	<i>451</i>
2.3. <i>Obligaciones específicas cuando se utilice un sistema de inteligencia artificial con finalidad de identificación biométrica.....</i>	<i>453</i>
2.4. <i>La obligación de realizar una evaluación de impacto relativa a los derechos fundamentales</i>	<i>454</i>
3. El representante autorizado y sus obligaciones	456
4. El importador y sus obligaciones	457
5. El distribuidor y sus obligaciones.....	458
6. Posibles alteraciones de las responsabilidades de los operadores.....	460
IV. LAS AUTORIDADES NOTIFICANTES.....	461
1. Concepto de autoridad notificante	461
2. Actividades de la autoridad notificante	463
V. MEDIDAS DIRIGIDAS A LOS PROVEEDORES Y USUARIOS A PEQUEÑA ESCALA.....	464
VI. LA NOTIFICACIÓN DE INCIDENTES GRAVES	466

	<u>Página</u>
1. Concepto de incidente grave	466
2. ¿Quién, ante quién y cuándo se ha de notificar?	467
3. Actuaciones posteriores a la notificación del incidente.	469
VII. CONCLUSIONES.....	470
SUJETOS Y AGENTES EN EVALUACIONES DE CONFORMIDAD (ORGANISMOS NOTIFICADOS).....	473
I. INTRODUCCIÓN.....	473
II. EL PROCEDIMIENTO DE NOTIFICACIÓN	475
1. La solicitud de designación	476
2. El procedimiento de notificación.....	477
3. Identificación y publicidad de los organismos notificados.....	479
4. Cambios en la notificación.....	479
5. Cuestionamiento de la competencia de los organismos notificados	481
III. LA ACTUACIÓN DE LOS ORGANISMOS NOTIFICADOS.	482
1. Los requisitos aplicables a los organismos notificados	482
2. Obligaciones operacionales de los organismos notificados. Coordinación por la comisión	489
3. Expedición y validez de certificados.....	490
4. Obligaciones de información de los organismos notificados.....	491
IV. RECAPITULACIÓN.....	492
 LAS OBLIGACIONES DE LOS PROVEEDORES E IMPLANTADORES DE SISTEMAS DE ALTO RIESGO	
 LA EVALUACIÓN DE IMPACTO DE DERECHOS FUNDAMENTALES POR QUIENES DESPLIEGAN SISTEMAS DE INTELIGENCIA ARTIFICIAL EN EL REGLAMENTO.....	495

	<u>Página</u>
I. INTRODUCCIÓN.....	495
II. LAS EVALUACIONES DE IMPACTO COMO HERRAMIENTA DE PONDERACIÓN DE DERECHOS FUNDAMENTALES	496
1. Diferencias entre análisis de riesgos y evaluaciones de impacto. El ejemplo de la protección de datos.....	498
2. Diferencias entre los riesgos a tratar del sistema de gestión de riesgos del artículo 9 vs evaluación de impacto de derechos fundamentales del artículo 27 Reglamento	500
3. Tipologías de evaluaciones de impacto.....	503
III. ANTECEDENTES DE EVALUACIONES DE DERECHOS FUNDAMENTALES DE SISTEMAS DE IA PREVIOS AL REGLAMENTO	505
2. Análisis de las FRIA en el Reglamento	512
A. <i>Alcance subjetivo: ¿Quién es el obligado a llevarla a cabo? y quienes intervienen en la misma?</i>	512
B. <i>Equipo que debe de intervenir en la misma.....</i>	513
C. <i>¿Cuándo se realiza?</i>	516
D. <i>¿Sobre qué se realiza? Alcance sustantivo. Una preEIA.</i>	517
V. PASOS A REALIZAR EN UNA FRIA	519
VI. EVALUACIONES DE IMPACTO EN DERECHOS FUNDAMENTALES Y EVALUACIONES DE IMPACTO EN PROTECCIÓN DE DATOS.....	529
VII. RECAPITULACIÓN Y CONCLUSIONES.....	531
LOS SISTEMAS DE GESTIÓN DE RIESGOS COMO OBLIGACIÓN ESPECÍFICA PARA LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL DE ALTO RIESGO EN EL ARTÍCULO 9 DEL REGLAMENTO	535
I. QUÉ ES UN SISTEMA DE GESTIÓN DE RIESGOS. AUTONOMÍA CONCEPTUAL RESPECTO DE FIGURAS AFINES	535

	<u>Página</u>
1. Identificación de riesgos (o fase de apreciación)	536
2. Evaluación de riesgos de derechos y su diferenciación del análisis y gestión de riesgos del proveedor	537
3. Mitigación de riesgos, monitoreo y revisión	545
II. EVOLUCIÓN DEL SIGNIFICADO, CONTENIDO Y DESTINATARIOS DE LA OBLIGACIÓN DE CONTAR CON UN SISTEMA DE GESTIÓN DE RIESGOS (ARTÍCULO 9 REGLAMENTO)	548
1. Qué es un sistema de gestión de riesgos según el Reglamento. El contenido de la obligación	552
2. Sujetos obligados. ¿Quién está obligado a contar con un sistema de gestión de riesgos? Cuadro resumen de las obligaciones vinculadas a los sistemas de riesgo alto	556
III. RECAPITULACIÓN Y CONCLUSIONES	562
DATOS Y GOBERNANZA DE DATOS Y CONEXIONES CON PRINCIPIOS PROTECCIÓN DE DATOS EN EL ARTÍCULO 10 DEL REGLAMENTO	565
I. INTRODUCCIÓN	565
II. GOBERNANZA DE DATOS	566
1. Concepto de gobernanza de datos	566
2. Contexto europeo	568
3. Concepto de gobernanza en el ámbito de la inteligencia artificial	570
III. EVOLUCIÓN, TRAMITACIÓN Y CONTENIDO FINAL DEL ARTÍCULO 10	572
1. Roles involucrados	572
2. Obligaciones	574
IV. APROXIMACIÓN CRÍTICA	579
V. CONFLUENCIA DE LA NORMATIVA DE PROTECCIÓN DE DATO	588
VI. CONCLUSIONES	591

	<u>Página</u>
SISTEMAS DE GESTIÓN DE CALIDAD, DOCUMENTACIÓN TÉCNICA Y CONSERVACIÓN EN EL REGLAMENTO.....	595
I. INTRODUCCIÓN.....	595
II. EL ARTÍCULO 17 DEL REGLAMENTO SOBRE SISTEMA DE GESTIÓN DE LA CALIDAD.....	600
III. EL ARTÍCULO 11 DEL REGLAMENTO SOBRE DOCUMENTACIÓN TÉCNICA CON ANEXOS.....	604
IV. EL ARTÍCULO 18 DEL REGLAMENTO SOBRE CONSERVACIÓN DE LA DOCUMENTACIÓN	609
V. EL INICIAL ARTÍCULO 50 DEL REGLAMENTO SOBRE CONSERVACIÓN DE LOS DOCUMENTOS.....	609
VI. CONCLUSIONES.....	610
LA OBLIGACIÓN DE CONSERVAR REGISTROS DE LOS SISTEMAS DE ALTO RIESGO EN EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL	613
I. INTRODUCCIÓN.....	613
II. ALGUNAS NOCIONES PREVIAS RELACIONADAS CON LAS OBLIGACIONES EN MATERIA DE REGISTROS.....	614
III. LAS OBLIGACIONES EN MATERIA DE REGISTROS EN EL REGLAMENTO: EVOLUCIÓN, TRAMITACIÓN Y CONTENIDO FINAL.....	619
IV. REFLEXIONES Y ANÁLISIS FINAL	626
TRANSPARENCIA Y COMUNICACIÓN DE INFORMACIÓN A LOS RESPONSABLES DEL DESPLIEGUE EN EL ARTÍCULO 13 REGLAMENTO DE INTELIGENCIA ARTIFICIAL	627
I. INTRODUCCIÓN: APROXIMACIÓN GENERAL AL ARTÍCULO 13 DEL REGLAMENTO.....	627
1. El proceso legislativo de conformación del artículo 13 y sus retos interpretativos	628
2. Objetivos y consideraciones metodológicas	630

	<u>Página</u>
II. DESCRIPCIÓN SISTEMÁTICA DEL ARTÍCULO 13: DIMENSIONES DE LA TRANSPARENCIA.....	634
III. TRANSPARENCIA, INTERPRETABILIDAD Y EXPLICABILIDAD: SU TRATAMIENTO (A)SISTEMÁTICO EN EL ARTÍCULO 13.....	645
1. Significado y tipos de transparencia en el reglamento y en el artículo 13	648
2. La interpretabilidad en el artículo 13: ¿desincentivación de los modelos de caja negra?	651
3. La explicabilidad en el artículo 13: un enfoque ambiguo y limitado	656
IV. ÁMBITOS SUBJETIVO Y FORMAL DEL ARTÍCULO 13	663
1. Sujetos y los fines de la transparencia en el artículo 13: los grandes ausentes en el reglamento	663
V. EL CONTENIDO MATERIAL DE LA OBLIGACIÓN DE TRANSPARENCIA.....	669
1. Información sobre la idoneidad funcional y otras propiedades	672
2. Información sobre la corrección funcional o rendimiento predictivo: la «precisión» y sus «métricas»	674
VI. VALORACIÓN FINAL DEL ARTÍCULO 13 DEL REGLAMENTO	678
1. Una mejorable técnica redaccional	678
2. Una mejorable articulación de las relaciones entre la transparencia, la interpretabilidad y la explicabilidad: ¿desincentivación de los modelos de caja negra?	678
3. Indefinición del «tipo y nivel de transparencia adecuados»	679
4. Un contenido de «minimis» de la transparencia material y sin concreción de su alcance	679
5. Normalización y transparencia de los sistemas de alto riesgo	680
6. Los «grandes olvidados» del Reglamento	680

	<u>Página</u>
LA VIGILANCIA O SUPERVISIÓN HUMANA EN EL ARTÍCULO 14 DEL REGLAMENTO DE INTELIGENCIA ARTIFICIAL: ¿UN MERO REQUISITO OBLIGATORIO PARA LOS SISTEMAS DE ALTO RIESGO?	681
I. INTRODUCCIÓN.....	681
II. LA VIGILANCIA O SUPERVISIÓN HUMANA EN LA PROPUESTA DE LA COMISIÓN DE ABRIL DE 2021	683
1. Supervisión humana para prevenir o reducir riesgos	683
2. Supervisión humana efectiva desde el diseño	684
3. ¿Cómo lograr una supervisión humana efectiva?	686
4. El rol del responsable del despliegue en la supervisión humana	687
III. LA TRAYECTORIA DE LA SUPERVISIÓN HUMANA EN EL PROCEDIMIENTO LEGISLATIVO ORDINARIO.....	688
1. Voces que reclaman el derecho a la intervención humana (y otras garantías) para la toma de decisiones basada en sistemas de alto riesgo.....	689
2. Humanos, ¿qué humanos?.....	691
IV. LA VERSIÓN FINAL DE LA SUPERVISIÓN HUMANA EN EL REGLAMENTO DE UN VISTAZO	692
V. ALGUNAS REFLEXIONES ACERCA DE QUÉ PODEMOS ESPERAR DE LA SUPERVISIÓN HUMANA EN EL REGLAMENTO	693
1. ¿Pueden los seres humanos cumplir la finalidad normativa de la supervisión humana en el Reglamento?	694
2. ¿Es necesario, en virtud del Reglamento, que haya seres humanos in the loop en la toma de decisiones con inteligencia artificial de alto riesgo para garantizar la supervisión humana efectiva exigida?	695
3. Más allá de la supervisión humana, ¿tenemos un Reglamento centrado en el ser humano?.....	697
VI. CONCLUSIONES.....	699

	<u>Página</u>
PRECISIÓN Y SOLIDEZ DE LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL DE ALTO RIESGO EN EL ARTÍCULO 15 DEL REGLAMENTO	701
I. INTRODUCCIÓN A LA PRECISIÓN Y LA SOLIDEZ EN LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL DE ALTO RIESGO	701
II. EVOLUCIÓN, TRAMITACIÓN Y CONTENIDO FINAL DEL ARTÍCULO 15	707
III. LAS EXIGENCIAS DE UN NIVEL ADECUADO DE PRECISIÓN Y SOLIDEZ	710
1. Métricas y rendimiento del sistema	712
2. Evaluación de la precisión y la solidez para garantizar la calidad del sistema	718
IV. CONCLUSIONES	719
 CIBERSEGURIDAD EN SISTEMAS DE INTELIGENCIA ARTIFICIAL DE ALTO RIESGO EN EL ARTÍCULO 15 DEL REGLAMENTO	 721
I. ¿LA CIBERSEGURIDAD ES PARA TODOS LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL? LA OBLIGACIÓN DE CIBERSEGURIDAD DESDE SU PROPUESTA HASTA LA APROBACIÓN FINAL	722
1. Evolución, tramitación y contenido final de la ciberseguridad como requisito en los sistemas de inteligencia artificial en el marco de la propuesta adoptada	724
2. La ciberseguridad en sistemas de inteligencia artificial catalogados alto riesgo	731
II. EL MARCO EUROPEO DE CERTIFICACIÓN DE LA CIBERSEGURIDAD COMO INSTRUMENTO DE GARANTÍA DE LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL DE ALTO RIESGO	734
1. Certificaciones de ciberseguridad en sistemas de inteligencia artificial de alto riesgo ¿Obligatorias o voluntarias?	736

	<u>Página</u>
2. Obligaciones de ciberseguridad en los sistemas de inteligencia artificial de alto riesgo	737
2.1. <i>La ciberseguridad en sistemas de inteligencia artificial de alto riesgo implementados por autoridades</i>	738
2.2. <i>La ciberseguridad en los sistemas de inteligencia artificial de alto riesgo que hagan parte de actividades críticas o servicios esenciales</i>	739
2.3. <i>Evaluaciones de conformidad como instrumento para la ciberseguridad en los sistemas de inteligencia artificial catalogados como de alto riesgo</i>	740
III. CONCLUSIONES	741
VIGILANCIA POSCOMERCIALIZACIÓN EN LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL DE ALTO RIESGO EN EL REGLAMENTO. DESCRIPCIÓN, MEDIDAS Y CASOS DE USO	743
I. LA VIGILANCIA POSCOMERCIALIZACIÓN EN EL REGLAMENTO	743
1. Introducción	743
2. ¿Qué es un plan de vigilancia postcomercialización y qué incluye?	744
3. ¿Por qué es necesario un plan de vigilancia post-comercialización?	747
4. El plan de vigilancia postcomercialización en el Reglamento	748
5. Quién debe realizar el sistema de vigilancia postcomercialización	749
II. CÓMO ABORDAR EL REQUISITO DE LA VIGILANCIA POSCOMERCIALIZACIÓN	750
1. Plan de Vigilancia y Sistema de Vigilancia. Elementos clave de la vigilancia	750
3. El concepto de Indicador	752
4. Medidas a desarrollar en el Plan de Vigilancia	753
5. Validez del Sistema y el Plan de Vigilancia	753
III. CONCLUSIONES	754

**INTELIGENCIA ARTIFICIAL DE USO GENERAL,
SISTEMAS QUE NO SON DE ALTO RIESGO Y LOS
SISTEMAS DEL ARTÍCULO 50**

INTELIGENCIA ARTIFICIAL DE USO GENERAL, MODELOS FUNDACIONALES (Y «CHAT GPT») EN EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL	757
I. INTRODUCCIÓN.....	757
II. INTELIGENCIA ARTIFICIAL GENERAL, DE USO GENERAL, MODELOS FUNDACIONALES E INTELIGENCIA ARTIFICIAL GENERATIVA.....	759
III. ¿QUÉ ES Y QUÉ NO ES INTELIGENCIA ARTIFICIAL DE USO GENERAL EN EL REGLAMENTO? MODELOS DE USO GENERAL CON RIESGO SISTÉMICO Y EXCLUSIÓN DE LOS MODELOS ESPECÍFICAMENTE DESTINADOS A INVESTIGACIÓN	761
IV. ALGUNOS RETOS NORMATIVOS DE LA INTELIGENCIA ARTIFICIAL GENERATIVA.....	764
V. APLICABILIDAD DEL REGLAMENTO COMO NORMA GENERAL Y EVOLUCIÓN NORMATIVA DEL TRATAMIENTO DE LA INTELIGENCIA ARTIFICIAL DE USO GENERAL	769
VI. OBLIGACIONES DE LOS PROVEEDORES DE INTELIGENCIA ARTIFICIAL DE USO GENERAL EN EL REGLAMENTO	771
VII. SUPERVISIÓN Y SEGUIMIENTO, Y RÉGIMEN SANCIONADOR.....	773
VIII. CONCLUSIONES.....	775
CÓDIGOS DE CONDUCTA, SELLOS O CERTIFICACIONES PARA LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL QUE NO SON DE ALTO RIESGO (ARTÍCULO 95 DEL REGLAMENTO	779

I.	LA REGULACIÓN EN EL ARTÍCULO 95 DE CÓDIGOS DE CONDUCTA PARA SISTEMAS QUE NO SON DE ALTO RIESGO	779
II.	FINALMENTE EL REGLAMENTO NO HA INCLUIDO UNOS PRINCIPIOS OBLIGATORIOS PARA TODO TIPO DE SISTEMA INTELIGENCIA ARTIFICIAL.....	781
III.	LA INSERCIÓN DE LA INTELIGENCIA ARTIFICIAL EN LOS MODELOS DE CERTIFICACIÓN Y SELLOS TECNOLÓGICOS EN LA UE. ¿UN SELLO ESPAÑOL DE INTELIGENCIA ARTIFICIAL?	784
IV.	LAS VARIADAS INICIATIVAS Y HERRAMIENTAS DE CERTIFICACIÓN O SELLOS DE INTELIGENCIA ARTIFICIAL EUROPEAS E INTERNACIONALES	786
V.	PARA CONCLUIR.....	791
	EL ARTÍCULO 50 DEL REGLAMENTO Y LAS OBLIGACIONES DE TRANSPARENCIA DE LOS PROVEEDORES Y RESPONSABLES DEL DESPLIEGUE DE DETERMINADOS SISTEMAS DE INTELIGENCIA ARTIFICIAL	793
I.	LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL DE RIESGO LIMITADO.....	793
II.	EVOLUCIÓN, TRAMITACIÓN Y CONTENIDO FINAL DEL ARTÍCULO 50 REGLAMENTO.....	794
III.	EL ALCANCE DE LAS OBLIGACIONES DE TRANSPARENCIA PREVISTAS EN EL ARTÍCULO 50 REGLAMENTO	796
IV.	SISTEMAS DE INTELIGENCIA ARTIFICIAL QUE INTERACTÚEN DIRECTAMENTE CON PERSONAS FÍSICAS ...	797
	1. Sistemas incluidos.....	797
	2. Alcance de las obligaciones.....	798
V.	SISTEMAS DE INTELIGENCIA ARTIFICIAL QUE GENEREN CONTENIDOS SINTÉTICOS.....	799
	1. Sistemas incluidos.....	799

	<u>Página</u>
2. Alcance de las obligaciones	800
VI. SISTEMAS DE INTELIGENCIA ARTIFICIAL DE RECONOCIMIENTO DE EMOCIONES	801
1. Sistemas incluidos	801
2. Alcance de las obligaciones	805
VII. SISTEMAS DE INTELIGENCIA ARTIFICIAL DE CATEGORIZACIÓN BIOMÉTRICA	806
1. Sistemas incluidos	806
2. Alcance de las obligaciones	808
VIII. SISTEMAS DE INTELIGENCIA ARTIFICIAL QUE GENEREN O MANIPULE CONTENIDOS QUE CONSTITUYAN UNA ULTRAFALSIFICACIÓN	808
1. Sistemas incluidos	808
2. Alcance de las obligaciones	811
IX. RECAPITULACIÓN.....	812

SANDBOX, GOBERNANZA, VIGILANCIA, RÉGIMEN SANCIONADOR, DERECHOS Y CONFIDENCIALIDAD EN EL REGLAMENTO

SANDBOX, ESPACIOS CONTROLADOS Y PRUEBAS EN CONDICIONES REALES DE SISTEMAS DE INTELIGENCIA ARTIFICIAL EN EL REGLAMENTO. MEDIDAS PARA PYMES, STARTUPS Y MICRO EMPRESAS	817
I. LAS «MEDIDAS DE APOYO A LA INNOVACIÓN» DEL CAPÍTULO VI.....	817
II. ORIGEN Y CONCEPTO DE SANDBOX Y ESPACIOS CONTROLADOS.....	818
III. EXPERIENCIAS DE SANDBOX DE INTELIGENCIA ARTIFICIAL.....	820

	<u>Página</u>
IV. LAS VENTAJAS QUE IMPLICAN LOS SANDBOX DE INTELIGENCIA ARTIFICIAL DESDE LOS DIFERENTES PUNTOS DE VISTA.....	823
V. EL MARCO NORMATIVO DE UN SANDBOX DE INTELIGENCIA ARTIFICIAL BAJO EL REGLAMENTO	827
VI. EXCEPCIONALIDAD DEL RÉGIMEN JURÍDICO Y RESPONSABILIDAD DE LOS PARTICIPANTES. TEORÍA, REALIDAD Y REGLAMENTO.....	830
VII. AUTORIDADES DEL SANDBOX, SELECCIÓN DE PARTICIPANTES, DURACIÓN, DESARROLLO Y OBTENCIÓN DE EVALUACIÓN DE CONFORMIDAD	834
1. La autoridad competente del sandbox y su cooperación con otras autoridades nacionales y europeas	834
2. Selección y admisión de participantes y duración del sandbox.....	835
3. Desarrollo, finalización y logro de una evaluación de conformidad en el sandbox	837
VIII. LA PROTECCIÓN DATOS EN EL CONTEXTO DE UN SANDBOX DE INTELIGENCIA ARTIFICIAL.....	837
IX. PRUEBAS EN CONDICIONES DE SISTEMAS INTELIGENCIA ARTIFICIAL DE ALTO RIESGO.....	839
X. PYMES, STARTUPS Y MICROEMPRESAS EN EL REGLAMENTO	841
XI. CONCLUSIONES Y RECAPITULACIÓN.....	842
LA GOBERNANZA Y VIGILANCIA DEL REGLAMENTO DE INTELIGENCIA ARTIFICIAL: AUTORIDADES DE VIGILANCIA DEL MERCADO, COMISIÓN Y LAS DIVERSAS ENTIDADES	845
I. INTRODUCCIÓN.....	845
II. EVOLUCIÓN, TRAMITACIÓN Y CONTENIDO FINAL	846
III. EL MODELO DE GOBERNANZA DE LA UE DEL REGLAMENTO	847

	<u>Página</u>
IV. COMISIÓN EUROPEA. ATRIBUCIONES Y FUNCIONES....	849
V. OFICINA DE INTELIGENCIA ARTIFICIAL. NATURALEZA, ESTRUCTURA Y FUNCIONES	852
VI. COMPETENCIAS DE LOS ESTADOS MIEMBROS Y AUTORIDADES NACIONALES COMPETENTES	856
1. Autoridades de vigilancia del mercado	858
2. Autoridad notificante	859
VII. COMITÉ EUROPEO DE INTELIGENCIA ARTIFICIAL Y LOS SUBGRUPOS PERMANENTES.....	860
1. Comité Europeo de Inteligencia Artificial. Estructura y atribuciones	860
2. Subgrupos permanentes de vigilancia del mercado y autoridades notificantes	863
VIII. OTROS ENTES DE ASESORAMIENTO, APOYO Y COLABORACIÓN	864
1. Foro Consultivo	864
2. Grupo de expertos científicos independientes	865
3. Centro europeo para la transparencia algorítmica y las estructuras de apoyo a las pruebas de inteligencia artificial	866
IX. PROCEDIMIENTOS DE SALVAGUARDIA, VIGILANCIA DEL MERCADO Y CONTROL DE LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL EN LA UNIÓN.....	867
1. Procedimiento aplicable a los sistemas de inteligencia artificial que presenten un riesgo a nivel nacional	868
2. Procedimiento de salvaguardia de la Unión	870
3. Procedimiento respecto de sistemas de inteligencia artificial conformes que presenten un riesgo	870
4. Procedimiento en caso de incumplimiento formal	871
5. Procedimiento aplicable a los sistemas calificados por los proveedores como no de alto riesgo en aplicación del Anexo III	871

	<u>Página</u>
EL RÉGIMEN SANCIONADOR EN EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL	873
I. INTRODUCCIÓN.....	873
II. EL ARTÍCULO 71 DEL REGLAMENTO: NECESIDAD DE DESARROLLO LEGISLATIVO E INTERACCIÓN CON OTRAS NORMATIVAS Y FIGURAS ALTERNATIVAS A LA MULTA Y ESPECIFICACIONES PARA LAS ADMINISTRACIONES.....	874
III. PRESCRIPCIÓN Y TIPIFICACIÓN DE INFRACCIONES Y CUANTÍA DE LAS MULTAS EN EL ARTÍCULO 71	877
IV. CRITERIOS PARA DETERMINAR LA CUANTÍA DE LA MULTA EN EL ARTÍCULO 71.....	881
V. PROPUESTAS Y CAMBIOS DEL PARLAMENTO Y EN LAS VERSIONES FINALES.....	885
1. Cambios propuestos en el Parlamento	885
2. Las novedades en los textos finales	887
VI. EL ARTÍCULO 72 SOBRE LA POTESTAD SANCIONADORA DEL SUPERVISOR EUROPEO DE PROTECCIÓN DE DATOS.....	889
VII. CONCLUSIONES.....	891
 DERECHO A PRESENTAR UNA RECLAMACIÓN Y DERECHO A UNA EXPLICACIÓN. VÍAS DE RECURSO PARA LOS PARTICULARES EN EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL	 893
I. INTRODUCCIÓN.....	893
II. EL DERECHO A PRESENTAR UNA RECLAMACIÓN ANTE UNA AUTORIDAD DE VIGILANCIA DEL MERCADO.....	894
1. Evolución del texto de la disposición en los trabajos preparatorios	895
2. Diferencias entre la regulación del derecho de reclamación en el Reglamento y la establecida en el RGPD y otras leyes digitales europeas	897

	<u>Página</u>
3. Regulación del derecho a presentar una reclamación en el Reglamento	901
III. DERECHO A UNA EXPLICACIÓN DE LA TOMA DE DECISIONES INDIVIDUALES	904
1. Evolución del texto de la disposición en los trabajos preparatorios	904
2. El principio de explicabilidad de los sistemas de inteligencia artificial	906
3. Condiciones para el ejercicio del derecho a una explicación	908
4. La relación entre el derecho a una explicación del artículo 86 y el RGPD	912
5. Límites al derecho a una explicación: los derechos de propiedad intelectual y la información confidencial	914
IV. EL PAPEL DE DENUNCIANTES Y ASOCIACIONES DE INTERESES COLECTIVOS	916
V. CONCLUSIONES	918
ACCESO A DOCUMENTACIÓN Y CONFIDENCIALIDAD EN EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL	919
I. INTRODUCCIÓN.....	919
II. ANÁLISIS DEL CONTENIDO DEL ARTÍCULO 77 DEL REGLAMENTO	921
III. ANÁLISIS DEL CONTENIDO DEL ARTÍCULO 78 DEL REGLAMENTO	924
IV. CONCLUSIONES	927

Presentación e introducción

LORENZO COTINO Y PERE SIMÓN

En la actualidad, la inteligencia artificial (en adelante, IA) se ha consolidado como una de las tecnologías más disruptivas e influyentes de nuestro tiempo, por su ingente capacidad de rápida transformación de sectores productivos, desde la medicina hasta la seguridad, y, también, entre muchas otras revoluciones, por su posible integración en objetos, productos o servicios digitales. Al menos en la Unión Europea (en adelante, UE) se percibe desde hace años la necesidad de un marco regulatorio general robusto y coherente que acompañe y encauce el desarrollo de esta tecnología. La Unión Europea, para bien o para mal, ha sido la pionera en establecer esta regulación general tanto para sí misma como para intentar influir en el resto del mundo con lo que se ha llamado el *efecto Bruselas*, algo que en cierta medida consiguió con la aprobación del Reglamento General de Protección de Datos (en adelante, RGPD).

El Reglamento de Inteligencia Artificial de la Unión Europea (que en este tratado haremos referencia como RIA)¹ ha seguido un largo y costoso procedimiento legislativo², del que cabe ahora destacar especialmente la propuesta de la Comisión Europea de 2021³, la posición del Consejo de la Unión de diciembre de 2022.⁴ También son muy relevantes las enmiendas del Parlamento en junio de 2023⁵. El acuerdo político fue de 8 de diciembre de 2023 y desde entonces se ha ido concretando el texto y sus traducciones en las diversas fases de su adopción final hasta su publicación.

El RIA ha encajado la regulación de la IA en el ámbito de la seguridad y garantía de los productos, las normas de armonización y el modelo del llamado «nuevo marco legislativo». Se trata del marco por el que se establecen unas bases comunes para la comercialización, evaluación y vigilancia de productos en la Unión Europea⁶.

1. Su nombre propiamente es Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial).
2. Puede seguirse el mismo en los siguientes enlaces de referencia. Parlamento Europeo: [https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2021/0106\(COD\)&l=en](https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2021/0106(COD)&l=en) Y *Legislative Train Schedule*, también del Parlamento: <https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-regulation-on-artificial-intelligence>
3. <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex%3A52021PC0206>
4. <https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/es/pdf>
5. https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_ES.html
6. Las tres textos legales que conforman el Nuevo Marco Legislativo son: el Reglamento (CE) n.º 765/2008 del Parlamento Europeo y del Consejo por el que se establecen los

Todo sea dicho, es un modelo con el que los juristas en general no estamos muy familiarizados. Los pasos para abordar la IA en la UE se iniciaron hace más de siete años con intervención de Comisión, Parlamento y Consejo. Se busca una IA «made in Europe» caracterizada por estar diseñada éticamente para el respeto de los derechos y los principios democráticos.⁷

La propuesta de la Comisión Europea pretendía, de un lado, garantizar que los sistemas de IA utilizados en la UE e introducidos en el mercado europeo sean seguros y respeten los derechos de los ciudadanos; del otro, estimular la inversión y la innovación en el ámbito de la IA en Europa. La propuesta ha sido objeto de debate, estudio y enmienda, siendo hitos notables en la tramitación, que ha seguido el procedimiento legislativo ordinario, el acuerdo provisional del Consejo y el Parlamento europeo en fecha de 8 de diciembre de 2023, la aprobación por parte de los Comités de Mercado Interno y Libertades Civiles de 13 de febrero de 2024, la Resolución del Parlamento Europeo, de 13 de marzo de 2024 (corrección de errores de 16 de abril de 2024), y la aprobación definitiva por parte del Consejo de la Unión Europea con fecha de 21 de mayo de 2024.⁸

El texto definitivo del Reglamento pretende armonizar las normas sobre inteligencia artificial, y lo hace con un enfoque basado en el riesgo, estableciendo un marco de obligaciones y requisitos distintos en función del nivel de riesgo de la tecnología de IA aplicable y su uso concreto. La técnica del enfoque armonizador exige, por ejemplo, unas obligaciones reforzadas para los sistemas de inteligencia artificial calificados como de alto riesgo; lo que deberá concretarse con la elaboración de estándares técnicos y el establecimiento de controles *ad hoc* para su aplicación.

Aunque el RIA entra en vigor en 2024, ciertamente cuenta con una aplicación y obligatoriedad muy escalonada, que llega incluso a los seis años. En cualquier caso, es posible prever un enorme impacto en el mercado y la sociedad. Pronto seremos testigos de las prohibiciones de tecnologías y usos concretos, de la implantación y actividad de las oficinas y autoridades europeas y nacionales de supervisión de la IA. Ya en febrero de 2024 se creó la nueva Oficina de la IA. Veremos la incoación y resolución de procedimientos sancionadores, de la aprobación de estándares técnicos y de la actividad de los organismos de certificación, de la proliferación de sistemas de gestión de riesgos específicos, de la aparición de códigos de conducta voluntarios sectoriales, entre muchas otras cuestiones que serán reflejo de la norma que es objeto de estudio en el presente trabajo.

Esta obra proporciona una comprensión sistemática, exhaustiva y detallada del nuevo RIA. La aproximación que los directores hemos querido realizar es, a nuestro

requisitos de acreditación y vigilancia del mercado de los productos; la Decisión n.º 768/2008/CE del Parlamento Europeo y del Consejo sobre un marco común para la comercialización de los productos y; el Reglamento (UE) 2019/1020 del Parlamento Europeo y del Consejo relativo a la vigilancia del mercado y la conformidad de los productos.

7. Un análisis exhaustivo de los pasos y políticas de la UE en la materia hasta 2019, Cotino Hueso, L., «Ética en el diseño para el desarrollo de una inteligencia artificial, robótica y big data confiables y su utilidad desde el derecho» en *Revista Catalana de Derecho Público* n.º 58 (junio 2019). <http://revistes.eapc.gencat.cat/index.php/rcdp/issue/view/n58> <http://dx.doi.org/10.2436/rcdp.i58.2019.3303>
8. <https://data.consilium.europa.eu/doc/document/PE-24-2024-INIT/es/pdf>

modo de ver, realmente original, al huir de forma deliberada del tradicional modelo de comentarios a una Ley o Reglamento, ordenados por orden de aparición en el articulado de la norma. Podríamos haber realizado un breve comentario para cada artículo, pero tal clasificación, a nivel de estructura y organización de contenidos, no hubiera resultado una aportación significativa. Por ello, hemos realizado un esfuerzo para realizar una clasificación por bloques temáticos que permita ofrecer respuesta global y exhaustiva a la aportación que supone, tanto a nivel teórico como práctico, la aprobación de un Reglamento que cuenta con 180 considerandos, 112 artículos y 13 anexos. Por hacernos una idea, frente a las 60 mil palabras del RGPD, el RIA cuenta con unas 108 mil.

El resultado que presentamos en esta obra colectiva incluye un total de 38 capítulos, en el que participan 34 autores, de los cuales 30 son doctores y 29 profesores universitarios. Hemos seleccionado a las personas de referencia en España en cada una de las materias. Ahora parece que todo el mundo es experto en Derecho e IA, pero lo cierto es que no tantas personas ameritan dedicarse décadas al Derecho digital y ya unos cuantos años en concreto a la IA. Y esto sucede con los autores de esta magna obra. Algunos comparten autoría y otros se han encargado de la redacción de dos o más capítulos. Contamos asimismo con la suerte de la participación de máximos referentes internacionales en la materia como Corvarán, Galetta, Mantelero o Ziller.

Los directores agradecemos y mucho las contribuciones de más de treinta personas. Todo sea dicho, no siempre es fácil ni grata la labor de seleccionar a los expertos, distribuir y delimitar sus trabajos, disciplinar a los autores en el cumplimiento de los tiempos y normas de los trabajos, por supuesto, revisar tales trabajos.

Por lo que se refiere a la organización temática, la primera parte está dedicada al estudio del RIA y su contextualización mundial, desde Iberoamérica y en Europa. Contamos en este apartado con las significativas contribuciones de los doctores Alessandro Mantelero, Jacques Ziller, Juan Gustavo Corvalán y María Victoria Carro, todas ellas voces destacadas en la intersección entre el Derecho y la IA. La segunda parte de la obra resuelve las cuestiones terminológicas (qué regula el RIA, las exclusiones y qué se entiende por IA), el ámbito territorial y el alcance del RIA, así como su relación con la protección de datos. La tercera parte incorpora un análisis de las IA prohibidas o inaceptables para el RIA, y a continuación, en la cuarta parte, se analizan los sistemas de IA de alto riesgo, delimitando y analizando los ámbitos más polémicos o que han sido objeto de un mayor debate. Dedicamos un quinto apartado de forma exclusiva al régimen general aplicable a los sistemas de IA de alto riesgo, con la aplicación de las normas armonizadas y el estudio de los modelos de evaluación de la conformidad y los organismos notificados. La sexta parte abre el amplio bloque de obligaciones específicas de los proveedores e implementadores (responsables del despliegue) de sistemas de IA de alto riesgo. La séptima parte está dedicada a la regulación de los sistemas que no han sido clasificados como IA de alto riesgo, a las IA de uso general y a los sistemas del artículo 50 del RIA. Finalmente, la octava parte es la que se ha reservado para el estudio de los mecanismos de gobernanza y vigilancia del cumplimiento, el régimen sancionador, la posibilidad de establecer sandboxes regulatorios y espacios controlados de pruebas, así como el derecho a presentar una reclamación y a obtener una explicación sobre la IA y las posibilidades de acceso a documentación y confidencialidad en el RIA.

En definitiva, la obra que el lector tiene entre sus manos se presenta como una herramienta útil para entender y aplicar el RIA. A través de una estructura temática comprensiva de una técnica armonizadora compleja y fruto de la colaboración de destacados expertos en la materia, se ofrece una visión exhaustiva y detallada de los múltiples aspectos regulatorios, técnicos y éticos que plantea esta normativa. Con ello, no solo se pretende proporcionar una guía práctica y teórica para académicos, profesionales y legisladores, sino también contribuir a un debate más amplio sobre el futuro de la IA en Europa y su impacto global. La profundidad y el rigor con los que se aborda cada tema reflejan el compromiso de los autores con la excelencia académica y la relevancia práctica, haciendo de esta obra una referencia que, esperamos, resulte referencia obligada, en el campo del Derecho de la IA.

Resta por último agradecer y mencionar los apoyos con los que cuenta esta obra. Así, especialmente cabe señalar que la coordinación y publicación del presente tratado estudio es resultado del proyecto MICINN Proyecto «Derechos y garantías públicas frente a las decisiones automatizadas y el sesgo y discriminación algorítmicas» 2023-2025 (PID2022-136439OB-I00) financiado por MCIN/AEI/10.13039/501100011033/, así como del Proyecto «Algorithmic law» (Prometeo/2021/009, 2021-24 Generalitat Valenciana). También esta obra queda en el marco de proyectos como «Algorithmic Decisions and the Law: Opening the Black Box» (TED2021-131472A-I00) y «Transición digital de las Administraciones públicas e inteligencia artificial» (TED2021-132191B-I00) del Plan de Recuperación, Transformación y Resiliencia. Estancia Generalitat Valenciana CIAEST/2022/1. Asimismo, del Convenio de Derechos Digitales-SEDIA Ámbito 5 (2023/C046/00228673) y Ámbito 6. (2023/C046/00229475). Es un orgullo para los coordinadores contar con un reconocimiento y apoyo investigador exhaustivos desde las diversas instituciones hace muchos años.

EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL
Y SU CONTEXTUALIZACIÓN MUNDIAL, DESDE
IBEROAMÉRICA Y EN EUROPA

El Reglamento de inteligencia artificial: la respuesta del legislador europeo a los retos de la inteligencia artificial

ALESSANDRO MANTELERO

Profesor titular de Derecho Civil en el Politécnico de Turín y titular de la Cátedra Jean Monnet de Sociedades Digitales Mediterráneas y Derecho

I. INTRODUCCIÓN

¿Cuál es la visión del legislador europeo en la regulación de la inteligencia artificial? ¿Cuál es la relevancia de adoptar un paradigma centrado en el riesgo? ¿Cómo se cruza este paradigma con la dimensión de los derechos fundamentales? Éstas son las principales preguntas a las que pretende responder un primer examen del RIA, que pone de relieve la necesidad de un enfoque interdisciplinar, que contemple también los escenarios internacionales y de otros países, para comprender plenamente la dinámica que inspiró al legislador europeo y que guiará la aplicación del RIA.

Para contribuir, con una primera breve reflexión sobre el texto finalmente aprobado del RIA, al creciente debate jurídico sobre la IA, en las páginas siguientes se examinará el núcleo de este marco normativo con un enfoque centrado en las opciones de política jurídica.

Debido a la naturaleza y relativo espacio de esta reflexión, no daremos cuenta de las diversas cuestiones que animaron y todavía animan el debate doctrinal respecto a los distintos aspectos de la relación entre IA, derecho y sociedad, tanto en nuestro ordenamiento jurídico como en otros, dejando al lector la oportunidad de profundizar en estos perfiles en la ya extensa bibliografía disponible.

En cuanto al examen del RIA, en la discusión que sigue hemos optado por dar prioridad a la respuesta a tres preguntas de investigación principales: (i) ¿Cuál es la visión del legislador europeo al regular el IA? (ii) ¿Qué relevancia tiene desde el punto de vista normativo la adopción de un paradigma centrado en el riesgo? (iii) ¿Cómo se relaciona el denominado modelo basado en el riesgo con la dimensión de los derechos fundamentales?

II. LA PERSPECTIVA EUROPEA

En 1968, Stanley Kubrick puso en escena 2001: Una odisea del espacio, en la que una inteligencia artificial se preocupaba por el bienestar de los seres humanos, pero

luego se volvía malévolos y daba lugar a un enfrentamiento icónico entre la voluntad del ser humano y la máquina. Desde luego, no era la primera vez que se fantaseaba con autómatas e inteligencia de máquina, pero no fue casualidad que la película se estrenara en los mismos años en los que, junto con la obra seminal de Alan Westin de 1967, se publicaban una serie de libros críticos sobre el papel de los ordenadores en la nueva sociedad digital.¹

Eran los años en los que ya se vislumbraban las posibilidades de las TIC, cuyas bases se estaban sentando, aunque las herramientas fueran aún inadecuadas para desarrollar todo su potencial. Así, ya se hablaba de IA, sistemas expertos y algoritmos de automatización, pero faltaban enormes cantidades de datos digitalizados y ordenadores capaces de procesarlos. Como ocurrió con el vapor, el telégrafo y muchos otros inventos, las ideas estaban ahí, pero su aplicación estaba en pañales.

Sin embargo, la propia visión del potencial de la tecnología de la información, incluso entonces, llevaba a pensar en términos de impacto social con una clara tensión entre la nueva utilidad aportada por las tecnologías y el riesgo relativo. La fe acrítica en el progreso que había caracterizado a los siglos anteriores, en el siglo XX se había roto por la experiencia de las guerras, la incertidumbre de los paradigmas científicos y la percepción de la debilidad del ser humano. Al progreso se unió, pues, la *hybris*, en el reto de generar algo fascinante y terrible (como había sido el caso del átomo), que contrapuso la visión positiva de la contracultura norteamericana a los interrogantes sobre el futuro de un mundo caracterizado por los procesos automatizados.

Podría argumentarse que todo esto se refiere al pasado y que tiene escasa relevancia jurídica en relación con el comentario sobre el RIA objeto de estas breves notas. Sin embargo, convendría partir de nuevo de Westin, para recordar cómo el jurista no puede reflexionar sobre cuáles son las normas prescindiendo de las fuertes instancias que caracterizan a la sociedad y generan el contexto al que las normas jurídicas, meros instrumentos, están llamadas a dar una de las respuestas posibles.

Así, llegando a la actualidad, no se puede entender el RIA y el tenor de sus disposiciones sin tener presente el dominio estadounidense y chino de los mercados de IA, la arriesgada jugada de Open AI (léase Microsoft) al poner en el mercado una tecnología inmadura como ChatGPT, o el uso sistémico por parte de estados totalitarios de herramientas biométricas y de control social. Hacer una lista de las categorías de usos prohibidos de la IA contenidas en el RIA, discutir las normas sobre los modelos de uso general (GPAI) —en particular los grandes modelos generativos—, abordar la cuestión de la evaluación de impacto, serían ejercicios incomprensibles si sólo se pensarán desde la perspectiva de categorías jurídicas abstractas.

Desde esta perspectiva, en primer lugar debemos situar el RIA en su contexto geopolítico pertinente. De hecho, esta legislación no surge de la nada, ni surge únicamente de las necesidades relacionadas con los posibles impactos de la IA, sino que forma parte de un diseño más amplio de la UE para una sociedad digital. Cuando se presentó el plan estratégico de la nueva Comisión Europea en 2019, la regulación digital se colocó como un elemento clave de la legislación de la UE para el período

1. Véase, por ejemplo, Miller, *The Assault on Privacy — Computers, Data Banks, Dossiers*, Ann Arbor, 1971; Brenton, *The Privacy Invaders*. Coward-McCann, New York, 1964; Packard, *The Naked Society*, New York, 1964.

2019-2023. En aquel momento, solo existían unas pocas normativas sobre la sociedad digital, principalmente la Directiva 95/46/CE sobre datos personales, su directiva derivada sobre privacidad electrónica, la Directiva 2000/31/CE sobre comercio electrónico (central sobre todo en el ámbito de la responsabilidad de los proveedores) y la Directiva sobre información del sector público (Directiva 2013/37/UE). Hoy en día, hay docenas de reglamentos adoptados o a punto de adoptarse a nivel europeo.²

Existe, por lo tanto, una estrategia de política reglamentaria que va mucho más allá del RIA y que es necesario comprender para valorar adecuadamente su alcance. Varias son las directrices que han llevado al legislador europeo a un esfuerzo normativo tan intenso, tal vez incluso excesivo, durante la legislatura que finalizará en 2024.

En primer lugar, por supuesto, están los cambios en la estructura de la sociedad digital. Tras la computación distribuida de los años ochenta y la llegada de Internet en los noventa, de donde surgieron las primeras normativas sobre datos y comercio electrónico, la explosión de sensores (léase IoT) y potencia de cálculo (léase computación en la nube) han allanado el camino a la IA, pero también a nuevas amenazas en el frente de la ciberseguridad y del impacto social. Al mismo tiempo, la concentración que ha caracterizado las últimas décadas de la economía digital, junto con la superación de la distinción entre los mundos *on-line* y *off-line*, han dejado solo el recuerdo de un entorno formado por pequeños operadores a los que proteger frente al riesgo legal ante sus inversiones pioneras en el sector digital, y han exigido respuestas más eficaces frente al dominio de las plataformas globales.³

Ya con respecto a estos primeros factores, el RIA se revela como una respuesta necesaria y coherente, tanto porque es precisamente el cambio de paradigma tecnológico (abundancia de datos y potencia de cálculo, omnipresencia de la tecnología y de la recogida de datos, difusión generalizada de los sistemas de interacción humano-máquina) lo que ha permitido la última revolución de la IA, como sobre todo porque esta revolución se basa en fenómenos de concentración de la información y del poder de mercado. De hecho, no es casualidad que las aplicaciones más avanzadas y críticas de la IA, en el ámbito GPAL, sean prerrogativa de un número extremadamente limitado de operadores a escala global, de los que se deriva un fuerte poder de condicionamiento del mercado y del escenario geopolítico, dada su ubicación prevalente en EE.UU. y China.

Es precisamente el escenario geopolítico el que constituye la segunda alma de la ola reguladora de la UE en materia digital. Aquí la atención se centra en la debilidad crónica del sector industrial europeo frente a sus competidores asiáticos y norteamericanos. Desde las materias primas hasta las plataformas, la UE no ha logrado hacerse con el dominio tecnológico en la escena mundial. Además, debido a las agresivas políticas de adquisición de las empresas más innovadoras por parte de los grandes actores, Europa es ahora en gran medida una tierra de colonización para las multinacionales digitales extranjeras. Al mismo tiempo, el gigantismo de estas multinacionales y su peso en el condicionamiento de la sociedad digital las ha llevado en los últimos tiempos a actuar como realidades cuasi-estatales, no

2. Para consultar un mapa de la legislación comunitaria pertinente, véase, por ejemplo https://www.bruegel.org/sites/default/files/2023-11/Bruegel_factsheet.pdf

3. Véase el Digital Markets Act y el Digital Services Act.

sólo definiendo de forma autónoma y autorreferencial políticas en las relaciones sociales mediadas digitalmente (piénsese, por ejemplo, en la dimensión cultural de las políticas de moderación de contenidos), sino también ejerciendo no pocas veces (desde las ciudades inteligentes a la pandemia) funciones propias del Estado.⁴

En este contexto, por lo tanto, es evidente la necesidad de la UE de dotarse de una normativa que regule el sector digital en un amplio espectro. Dado que, de hecho, no puede valerse del llamado *bully pulit*, al que se refería Reidenberg,⁵ para poder condicionar indirectamente a los productores de tecnología,⁶ sólo le queda el recurso exógeno de una normativa vinculante que proteja los intereses europeos.

El establecimiento de normas vinculantes es otro elemento que caracteriza al RIA. Esta postura, en términos de regulación jurídica, se pone de relieve no sólo por el recurso a la intervención legislativa en lugar del uso de *soft law* típico de los países productores de IA, sino también por las disposiciones específicas sobre la eficacia territorial del RIA. De hecho, el RIA prevé su aplicabilidad a los proveedores de sistemas de IA disponibles en la UE independientemente del establecimiento del proveedor, incluido el caso en que los proveedores y los responsables del despliegue de los sistemas de IA (*deployers*) estén situados fuera de la UE, pero la producción generada por los sistemas de IA se utilice en la UE.

Ahora bien, situar los intereses europeos en el centro del planteamiento regulador exige obviamente definir cuáles son, o más bien dar prioridad a los que constituyen la razón de ser de la Unión. A este respecto, si se compara el proceso de elaboración legislativa del RIA con el del RGPD, se observan diferencias significativas e indicativas. El liderazgo de la Dirección General de Justicia, centrado en la libertad, la seguridad y la justicia, que había caracterizado la redacción del RGPD, se sustituye aquí por el de la Dirección General de Mercado Interior, Industria, Emprendimiento y Pymes, y el eje del RIA es claramente el de una normativa de seguridad industrial, tal como la presentó a menudo la Comisión.

De esa manera, es evidente un fuerte hiato entre la forma en que se metabolizaron los riesgos de la IA en el discurso político jurídico, centrado en cuestiones éticas⁷ y los derechos fundamentales, aunque no pocas veces con una lamentable confusión entre ambos planes, y la propia visión de la Comisión dirigida principalmente a la

4. Véase, por ejemplo, sobre el tema Goodman, Powles, *Urbanism Under Google: Lessons from Sidewalk Toronto*, in *Fordham Law Review*, 2019, 88 (2), 457 y ss.

5. Véase Reidenberg, *Lex Informatica: The Formulation of Information Policy Rules Through Technology*, in *Texas Law Review*, 1998, 76 (3), 553, 581 y ss. («Government can use the bully pulpit approach to threaten and cajole industry to develop technical rules [...] The government's bully pulpit resulted in a flexible mechanism that can provide an information policy rule customized by network participants rather than an immutable architectural rule»).

6. Véase en este sentido, The White House, *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, 30 de octubre 2023, <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>

7. Piénsese en las iniciativas del El Supervisor Europeo de Protección de Datos, así como en la contribución más cuestionable del Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial, del que se encuentra rastro en uno de los considerandos de Reglamento.

seguridad industrial, con un amplio enfoque en la gestión de riesgos en términos de evaluación de la conformidad y con un peso significativo otorgado a los estándares.

La propuesta de la Comisión incluía referencias a la protección de los derechos fundamentales en relación con las posibles repercusiones de la IA, pero sin concretarlas. Esto contrastaba con un debate público del que se desprendía que los riesgos de la IA, por ejemplo, tenían menos que ver con el daño que el robot colaborador puede hacer al trabajador y más con la discriminación potencial y la desinformación que los algoritmos están introduciendo en la sociedad. Sólo tras el resultado del debate parlamentario, también gracias al apoyo del mundo académico internacional,⁸ se consiguió que el RIA contenga más detalles sobre el impacto en los derechos fundamentales.

Sin embargo, esta connotación de la propuesta normativa, aquí brevemente esbozada, pone de relieve el objetivo del legislador europeo, que no es primordialmente el de proteger los derechos fundamentales, sino el de fomentar el desarrollo de la IA en un contexto de debilidad industrial del sector en Europa. De ahí el enfoque centrado en la seguridad industrial y, sobre todo, el equilibrio elegido en la gestión de los riesgos, que se analizará con más detalle a continuación. Aquí, en términos generales, es suficiente señalar cómo las opciones de política industrial condujeron a una perspectiva más favorable a la aceptación del riesgo, diferente de la aversión al riesgo más marcada que se observa en el RGPD.⁹

También cabe destacar que la intensa actividad del legislador europeo en materia digital que ha caracterizado estos últimos años plantea una serie de cuestiones sistémicas que también afectan a el RIA. En primer lugar, la fragmentación de las distintas iniciativas en cuanto a promotores lleva a que los textos se redacten más en silos que de forma sistemática. Como ya señalaron organismos reguladores como la EDPB y el Supervisor Europeo, el afán por elaborar un nuevo marco regulador, inducido por las razones antes mencionadas, produjo muchos textos bastante extensos y complejos en un plazo bastante breve, sin que se afinara la coordinación entre ellos.

En segundo lugar, este amplio esfuerzo normativo afectó a los plazos de los distintos procesos de aprobación, con el resultado de que algunos ámbitos sólo están regulados parcialmente: un ejemplo de ello es la no aprobación del pilar complementario del RIA, es decir, la directiva sobre la responsabilidad relacionada con el uso de la IA.

Por otra parte, el enfoque adoptado por el legislador europeo se caracteriza por un considerable pragmatismo, buscando una regulación centrada en remedios *ex ante*, en términos de gestión de riesgos y enfoque *by-design* del producto en lugar de medidas compensatorias *ex post*. Esto lleva cada vez más a que las normas de responsabilidad sean las que cierren una regulación destinada a prevenir los riesgos de los sistemas

8. Brussels Privacy Hub, *More than 150 university professors from all over Europe and beyond are calling on the European institutions to include a fundamental rights impact assessment in the future regulation on artificial intelligence*, 12 de septiembre de 2023, <https://brusselsprivacyhub.com/2023/09/12/brussels-privacy-hub-and-other-academic-institutions-ask-to-approve-a-fundamental-rights-impact-assessment-in-the-eu-artificial-intelligence-act/>

9. Véase el art. 35 RGPD.

complejos. De hecho, el recurso tradicional a la responsabilidad extracontractual se adapta mal a un entorno caracterizado por la complejidad tecnológica, los operadores globales con grandes capacidades financieras, la pulverización de los daños y, con vistas a estimular la confianza en las nuevas tecnologías (la llamada IA fiable), la necesidad de garantizar entornos tecnológicos seguros en lugar de remedios de compensación en caso de consecuencias desastrosas.

Sin embargo, la falta de un sistema de normas de cierre sobre la responsabilidad en la IA —dadas las cuestiones relativas a su reparto tanto con respecto a los diversos componentes de los sistemas de IA como a la interacción humano-máquina— apunta a una falta de coordinación en el enfoque europeo. Otros legisladores, piénsese en las propuestas brasileñas sobre IA, han combinado más adecuadamente la gestión de riesgos, con sanciones por incumplimiento, y la responsabilidad por los daños causados por la IA.

La decisión de mantener separados ambos perfiles no fue, por lo tanto, feliz, como tampoco lo fue la de abordarlos con dos instrumentos legislativos de distinta naturaleza y, además, promover una actualización paralela del marco de responsabilidad por productos defectuosos en general. Desarrollar un modelo de protección *ex ante*, centrado en el análisis de riesgos, sin desarrollar a continuación un marco adecuado para las hipótesis residuales en las que una mala o deficiente gestión del riesgo causa un daño, acaba minando el impacto global de la intervención normativa derivada del RIA, que es así una obra inacabada en una visión amplia de la regulación de la IA. Tampoco cabe argumentar aquí, a diferencia de la protección de datos personales, una relevancia limitada de los perfiles indemnizatorios, dado que la delegación en los sistemas de IA de funciones de gestión de infraestructuras críticas, tanto a nivel funcional como social, no presagia escenarios circunscritos de daños potenciales.

III. LA MODULACIÓN DEL LLAMADO ENFOQUE BASADO EN EL RIESGO EN UNA LEGISLACIÓN DE PRIMERA GENERACIÓN

En segundo lugar, se prefirió un enfoque centrado en la clasificación de los casos de alto riesgo para facilitar que los operadores sepan desde el principio si están o no sujetos a la nueva normativa. Esta elección, encaminada a una aparente simplificación, resultó ser intrínsecamente compleja debido a la dificultad de dar una definición precisa de los sistemas de alto riesgo y a la evolución de los usos de la IA. De ahí las medidas correctoras, como la posibilidad de exenciones¹⁰ y futuras modificaciones del anexo III,¹¹ con un marco global que, en lugar de simplificar, corre

10. Véase art. 6.4, AI Act («A provider who considers that an AI system referred to in Annex III is not high-risk shall document its assessment before that system is placed on the market or put into service. Such provider shall be subject to the registration obligation set out in Article 49 (2). Upon request of national competent authorities, the provider shall provide the documentation of the assessment.»).

11. Véase art. 7, AI Act, que otorga a la Comisión la facultad de «adopt delegated acts in accordance with Article 97 to amend Annex III by adding or modifying use-cases of high-risk AI systems where both of the following conditions are fulfilled: (a) the AI systems are intended to be used in any of the areas listed in Annex III; (b) the AI systems pose a risk of harm to health and safety, or an adverse impact on fundamen-

el riesgo de hacer que el panorama de la explotación industrial de la IA sea tortuoso y susceptible de litigios.

También merece atención el criterio según el cual se tratan los casos de alto riesgo en términos de estrategias de mitigación. Según el modelo de riesgo industrial, se considera que el desarrollo de la IA puede justificarse¹² aunque albergue riesgos elevados. De ahí el criterio de aceptabilidad del riesgo residual, que como tal no tiene por qué ser elevado, sino que sólo se justifica por intereses superiores de otra naturaleza.

Este criterio de aceptabilidad se desarrolla a través de la evaluación de riesgos prevista en el RIA, es decir la evaluación de conformidad prevista en el artículo 43.¹³ No obstante, cabe señalar que un componente de esta evaluación es también la estimación del impacto sobre los derechos y libertades fundamentales, respecto a los cuales parece descartarse una aceptabilidad basada en una comparación indistinta de los intereses en juego.

El nivel de protección proporcionado por el ordenamiento jurídico europeo, así como por los distintos Estados miembros, a los derechos fundamentales excluye la posibilidad de su comprensión por la mera presencia de intereses opuestos o en virtud de la aceptabilidad social, así como la presencia de un riesgo residual aceptable.¹⁴ Sólo en caso de contraposición con intereses considerados iguales o superiores por el ordenamiento jurídico puede justificarse una comprensión necesaria y proporcionada de los derechos fundamentales.

Por último, existen casos en los que el legislador europeo consideró inaceptables determinados usos de la IA precisamente por su fuerte contraste a los derechos fundamentales y los principios del derecho de la Unión. Se trata de las señaladas en el artículo 5 del RIA, entre las que se encuentran las técnicas manipulativas, el denominado *social credit scoring* y determinados usos invasivos de las tecnologías biométricas. A este respecto, tuvo lugar un amplio debate, con participación de la sociedad civil, sobre la identificación de los usos prohibidos y sobre las excepciones (bastante articuladas, especialmente en lo que se refiere al uso de la identificación biométrica) que se añadieron a lo largo del proceso legislativo.

En lugar de basarse en la pirámide de riesgos a menudo mencionada (muchos sistemas de IA no regulados, algunos sujetos a obligaciones limitadas, unos pocos con alto riesgo sujetos a evaluación de cumplimiento, muy pocos prohibidos),¹⁵ el modelo general que se desprende de todo el marco normativo en el planteamiento

tal rights, and that risk is equivalent to or greater than the risk of harm or of adverse impact posed by the high-risk AI systems already referred to in Annex III.».

12. Véase también la referencia expresa a las ventajas de la IA en el art. 7, AI Act («When assessing the condition under paragraph 1, point (b), the Commission shall take into account the following criteria: [...] “the magnitude and likelihood of benefit of the deployment of the AI system for individuals, groups, or society at large, including possible improvements in product safety”»).
13. Véase también los artículos 9 y 17.
14. En algunas circunstancias, los riesgos residuales no pueden excluirse, pero esto implica la adopción de medidas complementarias a posteriori, como las indemnizaciones, y no su aceptabilidad.
15. Este modelo piramidal en realidad aporta poco sobre las opciones de política legislativa y es, sobre todo, funcional a una narrativa que quiere destacar la intervención

del riesgo relacionado con la IA debería reconstruirse más bien en función de la participación de los enfoques adoptados en la evaluación de riesgos.

A este respecto, conviene distinguir entre una evaluación próxima a la evaluación tecnológica, una evaluación de conformidad y una evaluación de impacto sobre los derechos fundamentales. La primera es el ejercicio elaborado en el art. 5 del RIA para definir las categorías prohibidas. Se trata de una evaluación *ex ante* formulada en abstracto sobre los nuevos usos de la tecnología cuya aceptabilidad normativa se valora en función de su impacto en los principios fundacionales del derecho de la UE. Un ejemplo es el uso de tecnologías subliminales destinadas a manipular la voluntad individual de «sistema de IA que se sirva de técnicas subliminales que trasciendan la conciencia de una persona o de técnicas deliberadamente manipuladoras o engañosas con el objetivo o el efecto de alterar de manera sustancial el comportamiento de una persona o un grupo de personas, mermando de manera apreciable su capacidad para tomar una decisión informada y haciendo que una persona tome una decisión que de otro modo no habría tomado, de un modo que provoque, o sea probable que provoque, perjuicios considerables a esa persona, a otra persona o a un grupo de personas». Dentro del mismo tipo de evaluación se encuentra también la lista de usos de alto riesgo que figura en el Anexo III, donde, también en este caso, los sistemas de IA se consideran en términos de categorías de uso, independientemente de su configuración específica y de su uso contextual.¹⁶

A este respecto, mientras que para la posible variación de las categorías prohibidas, en razón de la evolución tecnológica y del contexto sociotécnico, se prevé recurrir a modificaciones posteriores del artículo 5 del RIA, la evaluación de los sistemas de alto riesgo se deja para el futuro a los trabajos de la Comisión europea. Esta última opción, aunque permite circunscribir la actuación de la Comisión dentro de los límites definidos por el artículo 7 del RIA, implica sin embargo atribuirle la posibilidad de modificar el objeto de la legislación, lo que parece peculiar, habida cuenta de la naturaleza institucional de la Comisión y de la legitimidad del proceso legislativo de la Unión.

La evaluación de la conformidad, en cambio, es de otra naturaleza. Tanto si se basa en los procedimientos del Anexo VII (Conformidad fundamentada en la evaluación del sistema de gestión de la calidad y la evaluación de la documentación técnica) como en los del Anexo VI (Procedimiento de evaluación de la conformidad fundamentado en un control interno), según se trate o no de la utilización de tecnologías biométricas de alto riesgo definidas en el Anexo III,¹⁷ exige siempre la aplicación de un sistema de gestión de la calidad conforme al Artículo 17 del RIA, del que el sistema de gestión de riesgos conforme al Artículo 9 es un componente central, y que incluya también la evaluación del impacto sobre los derechos fundamentales.

minimalista, limitada a los casos más graves, por parte del legislador europeo, subrayando la orientación del Reglamento de AI favorable a la innovación.

16. Se hace referencia, por ejemplo, en el ámbito educativo a «AI systems intended to be used to determine access or admission or to assign natural persons to educational and vocational training institutions at all levels», donde existen diversas posibilidades de configuración de dichos sistemas en función de los parámetros utilizados y de los umbrales adoptados, así como diferentes implicaciones de aplicación en función del contexto sociocultural específico de utilización.
17. Véase art. 43, apartados 1 y 2, AI Act.

La evaluación de la conformidad, a diferencia de la evaluación tecnológica, es una evaluación centrada en un uso concreto de la IA, caracterizado por funcionalidades específicas y diferenciadoras, aunque susceptible de ser empleada en escenarios diferentes.¹⁸ Esta evaluación se centra en el riesgo industrial en términos tradicionales de daño a la integridad física y a la seguridad del producto/servicio, pero también incluye riesgos en términos de daño a los derechos fundamentales.¹⁹ A este respecto, el planteamiento del legislador europeo es dejar esta evaluación de la conformidad a la adopción de estándares.²⁰

Cabe señalar cómo el recurso al proceso de estandarización para la evaluación de la conformidad es coherente con la práctica de la gestión de los riesgos industriales y de los productos, en términos de seguridad (incluida la seguridad física de los seres humanos que interactúan con las máquinas, aquí la IA), pero parece inadecuado en lo que respecta al componente de evaluación del impacto sobre los derechos fundamentales. Por lo que respecta a estos últimos, no sólo la opacidad de los sistemas de estandarización, sino también la falta de participación de expertos en derechos fundamentales constituyen una primera cuestión crítica, estigmatizada incluso en el proyecto de solicitud de normalización presentado por la Comisión al CEN-CENELEC, cuya falta de experiencia en materia de derechos fundamentales se admite explícitamente.²¹

Más allá de los problemas estructurales del sistema de estandarización, existe una objeción metodológica más importante sobre la dificultad de utilizar estándares para evaluar el impacto sobre los derechos fundamentales. En efecto, los estándares, por su propia naturaleza, son utilizables en presencia de procesos caracterizados por una dinámica constante y repetitiva, por lo que es posible definir un estándar en la construcción de ferrocarriles, ya que las variables de velocidad, peso, pendiente, etc. se mueven dentro de rangos constantes con respecto a una actividad de circulación de trenes que resulta tener características uniformes independientemente de los distintos trazados.

Esta uniformidad no puede verse en el contexto del impacto de la IA sobre los derechos fundamentales, donde una misma aplicación de IA puede tener impactos significativamente diferentes debido a las características de las tecnologías utilizadas,

-
18. Siguiendo con la hipótesis planteada anteriormente, véase la nota 15, la evaluación de la conformidad se referirá a una determinada aplicación de IA que, a partir de parámetros relativos a la calificación del alumno en un plazo determinado, el rendimiento en las distintas asignaturas, la edad, las series temporales utilizadas en la fase de formación y otros muchos parámetros, podrá evaluar la admisión en una determinada carrera. Por tanto, no se tratará de un tipo de aplicación de IA, sino de un producto específico con un diseño y unas opciones de formación propias, aunque será susceptible de ser aplicado en diferentes contextos en cuanto a variables demográficas, tipo de carrera, etc.
 19. Véase art. 9.2.a, AI Act («identification and analysis of the known and the reasonably foreseeable risks that the high-risk AI system can pose to the health, safety or fundamental rights when the high-risk AI system is used in accordance with its intended purpose»).
 20. Véase art. 40, AI Act (Harmonised standards and standardisation deliverables).
 21. Véase European Commission, *Draft standardisation request to the European Standardisation Organisations in support of safe and trustworthy artificial intelligence*, 5 de diciembre de 2022, <https://ec.europa.eu/docsroom/documents/52376?locale=en>

el contexto de uso y los actores implicados. Si consideramos, por ejemplo, los sistemas de videovigilancia basados en IA, en términos de su impacto sobre los derechos fundamentales, existen diferentes escenarios dependiendo de si se utilizan en espacios públicos o privados, de si en estos últimos hay menores u otras personas vulnerables, de si se implementan o no funcionalidades de seguimiento en tiempo real, de si se utilizan en contextos caracterizados por altos niveles de delincuencia con el objetivo de luchar contra el crimen, y dependiendo de muchos otros factores que podrían añadirse debido a la variedad de escenarios posibles.

Por lo tanto, es evidente cómo la variabilidad de la dimensión contextual de la evaluación del impacto sobre los derechos fundamentales no puede conciliarse con una idea de estandarización, si por tal entendemos la posibilidad de definir un procedimiento preciso y uniforme, compuesto por etapas específicas de conformación de la tecnología según patrones predefinidos. Por otra parte, la conclusión puede ser diferente si los estándares se entienden como reglas metodológicas, es decir, no la definición de un proceso específico, sino más bien como un marco metodológico general para la gestión de riesgos en el caso de los derechos fundamentales, por ejemplo en lo que respecta a la cuestión central de los criterios de evaluación de impacto necesarios para comparar distintas opciones de diseño en el desarrollo de la IA.²²

Por último, a raíz del debate en el Parlamento Europeo, se introdujo en el RIA una obligación específica de evaluación del impacto en los derechos fundamentales (FRIA)²³ por parte de los responsables del despliegue de los sistemas de IA. Esta evaluación puede ser desarrollada en parte por el proveedor de IA sobre la base de posibles escenarios de uso, como ocurre en la evaluación de la conformidad, pero también debe tener en cuenta la aplicación concreta de la IA en el caso específico. Esto está en consonancia con los procesos de evaluación del impacto sobre los derechos humanos y de evaluación de impacto sobre la protección de datos, que se basan en evaluaciones contextuales del daño potencial a los derechos y libertades en juego.

Según la teoría general, con referencia a la distribución de los riesgos y las responsabilidades correspondientes, la evaluación del impacto sobre los derechos fundamentales se combina así con la evaluación de la conformidad, trasladando a los responsables del despliegue de los sistemas de IA parte de la carga de la gestión de las posibles consecuencias negativas de la IA vinculadas al contexto operativo específico de uso, con respecto al cual los responsables disponen de mayores márgenes de control o, al menos, de evaluación del riesgo real.

Por último, a diferencia de la evaluación de la conformidad, no se prevén procesos de estandarización para la evaluación del impacto en los derechos fundamentales. Esto está en línea con las experiencias anteriores en materia de evaluación del impacto sobre los derechos fundamentales y evaluación de impacto sobre la protección de datos, ámbitos en los que han surgido modelos de evaluación de riesgos basados en las mejores prácticas.

22. Para un ejemplo de este enfoque, véase Mantelero, *Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI*, The Hague, 2022, capítulo 2, <https://doi.org/10.1007/978-94-6265-531-7> (acceso abierto).

23. Véase Art. 27, AI Act.

Observando esta tripartición del proceso de evaluación a través del cual el modelo basado en el riesgo toma forma en el RIA y en su aplicación, cabe señalar, no obstante, una falta de uniformidad en el enfoque, que se pone de manifiesto en la falta de parámetros comunes de evaluación del riesgo y de metodologías comunes para evaluar el impacto sobre los derechos fundamentales en el contexto de la IA. Así, si por una parte el riesgo se define en términos generales en el Artículo 3 como una combinación de probabilidad y gravedad del evento perjudicial, en el Artículo 7 se enumeran una serie de parámetros adicionales en relación con la evaluación de la tecnología. En cambio, faltan indicaciones específicas para la evaluación de la conformidad y, en un cuestionable intento de simplificación durante los trilogos, se suprimieron las referencias a los parámetros que deben tenerse en cuenta en la evaluación incluso en el texto final del Artículo 27 sobre el impacto en los derechos fundamentales, frente a la más precisa propuesta parlamentaria. De ahí la necesidad de una reflexión metodológica sobre cómo llevar a cabo la evaluación de impacto, especialmente en relación con los derechos fundamentales, un punto crucial en la aplicación del RIA.

Con respecto a esta estructura básica del RIA, deben considerarse a continuación dos bloques de disposiciones, relativos respectivamente a las obligaciones de transparencia previstas para los sistemas que no son de alto riesgo y a las disposiciones añadidas en la fase final de la redacción del Reglamento para responder a las preocupaciones suscitadas por los sistemas de IA basados en modelos de uso general (GPAI), que se dieron a conocer al público en general especialmente tras la publicación de ChatGPT.

En lo que respecta al primer conjunto de normas, el RIA adopta lo que el Consejo de Europa ya indicó en sus directrices sobre IA y protección de datos personales acerca de la obligación de hacer consciente al usuario final del hecho de que está interactuando con un sistema de IA. Esta es una obligación justificada por la capacidad de la IA de emular diversos comportamientos humanos en la interacción humano-máquina. También se imponen obligaciones específicas adicionales de transparencia tanto a los productores, como a los responsables del despliegue de los sistemas de IA en relación con la capacidad de la IA de generar contenidos sintéticos, sobre todo teniendo en cuenta las criticidades que ello puede conllevar a la hora de alterar la realidad con importantes repercusiones sociales (por ejemplo, las noticias falsas).²⁴

Más complejo es el discurso en relación con la IA basada en modelos de uso general (GPAI), donde la premura debida a la aparición del problema en la fase final del proceso legislativo y las posiciones contrarias de algunos gobiernos a una regulación incisiva de este importante aspecto han llevado a perfilar una regulación que podría definirse como minimalista.

Sobre este punto, la reflexión debería ir mucho más allá de las limitadas consideraciones consignadas en estas páginas, planteando cuestiones que están en la raíz del problema de la regulación de la tecnología y que tienen que ver con el conocido dilema de Collingridge, donde la GPAI es una tecnología aún en fase incipiente, no por casualidad aquejada de varios problemas operativos no resueltos

24. Véase art. 50, AI Act.

y carente también de un verdadero modelo de negocio que justifique sus elevados costes de explotación y su impacto medioambiental.

La indiferencia ante el planteamiento centrado en la innovación responsable por parte de los operadores estadounidenses, la decisión de poner en el mercado soluciones aún inestables y fuente de múltiples riesgos, así como generadas en violación de las normas de protección del tratamiento de datos personales²⁵ y de protección de los derechos de propiedad intelectual,²⁶ han llevado al legislador europeo a una reacción reguladora encaminada a encontrar un equilibrio entre la protección y la fascinación por las posibilidades económicas (apoyada especialmente en la fase de trilogía por Francia en relación con el caso *Mistral AI*), cuando quizá hubiera tenido más razón de ser un enfoque basado en el principio de precaución.

El resultado de estas tensiones de política industrial ha sido la elaboración de una serie de normas que distinguen entre los modelos de GPAI y los sistemas que utilizan estos modelos. Lo que preocupa es el llamado riesgo sistémico, que se presume esencialmente sobre la base del tamaño de estos modelos y con una presunción relativa, con un registro público para los modelos GPAI caracterizados por el riesgo sistémico. El punto central es precisamente este riesgo, del que los proveedores de modelos tendrán que demostrar que han llevado a cabo un análisis y una gestión adecuados, haciendo un seguimiento de sus acciones según el modelo de rendición de cuentas ahora dominante en la normativa europea sobre la sociedad digital.

Dado que estos modelos están destinados a ser incluidos en sistemas de IA también de operadores distintos de los desarrolladores de los modelos, también existen obligaciones de transparencia para sus creadores en favor de estos operadores, así como de información sobre las fuentes utilizadas para el entrenamiento de los propios modelos (el conocimiento de las fuentes puede ser útil, por ejemplo, para identificar posibles sesgos).

Por último, también con vistas a equilibrar la gestión de los riesgos potenciales de la IA y los beneficios deseados, deben leerse las diversas disposiciones específicas en favor de la innovación, empezando por la amplia excepción prevista para las

-
25. Véase *Garante per la protezione dei dati personali*, Registro dei provvedimenti n. 112, 30.03.2023, doc. web n. 9870832, <https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870832>; *Garante per la protezione dei dati personali*, Registro dei provvedimenti n. 114, 11 de marzo de 2023, doc. web n. 9874702, <https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9874702>; *Garante per la protezione dei dati personali*, *ChatGPT: Garante privacy, notificato a OpenAI l'atto di contestazione per le violazioni alla normativa privacy*, comunicado de prensa del 29 de enero de 24, <https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9978020>
26. Véase United States District Court, Southern District of New York, *The New York Time Company v. Microsoft Corporation, OpenAI, Inc., OpenAI LP, OpenAI GP, LLC, OpenAI, LLC; OpenAI OPCO LLC, OpenAI Global LLC, OAI Corporation, LLC, and OpenAI Holdings, LLC*, 27 de diciembre de 2023, https://nytco-assets.nytimes.com/2023/12/NYT_Complaint_Dec2023.pdf

actividades de investigación,²⁷ hasta las normas *ad hoc* sobre las *sandboxes*,²⁸ es decir, áreas de experimentación controlada (ya adoptadas en el contexto de la aplicación del RGPD en varios países), y las disposiciones destinadas a permitir que los productos de IA se prueben en el mundo real con consecuencias implícitas en cuanto a la experimentación social, que, por esta razón, también implican un proceso de evaluación ética.²⁹

IV. CONCLUSIONES

Varias iniciativas, en todo el mundo y a distintos niveles, se centran en regular la IA. Los legisladores intentan dar una primera respuesta a los retos que plantea la revolución de la IA.

Las soluciones propuestas representan un compromiso entre la protección de los derechos fundamentales y los beneficios esperados de la IA. Esto ha llevado a los legisladores a atender sólo parcialmente a la demanda de protección de derechos y libertades de los individuos y de la sociedad, para no frenar el desarrollo de la IA, aún más así en aquellos contextos en los que no existe una industria fuerte de IA.

A la luz de este compromiso, es crucial un análisis interpretativo exhaustivo del reglamento y proporcionar directrices para su aplicación. El papel crucial del enfoque basado en el riesgo requiere tanto un planteamiento armonizado coherente con la teoría de la gestión del riesgo como el desarrollo de una metodología específica para el impacto sobre los derechos fundamentales. Esta última debe basarse en criterios y variables clave coherentes con en la normativa sobre IA y el marco normativo europeo, empezando por la Carta de los Derechos Fundamentales de la Unión Europea, y debe ser aplicada adecuadamente por las autoridades competentes y no puede delegarse en organismos de estandarización.

A este respecto, en relación con varios comentarios críticos sobre el RIA, conviene señalar que al redactar la ley es bueno ser ambicioso, pero en la evaluación *a posteriori* tenemos que ser realistas. Deberíamos situar el RIA en su contexto y recordar a los años en que Europa y muchos académicos defendían un enfoque puramente ético para la reglamentación de la IA. También deberíamos recordar al marco original de esta normativa, concebida principalmente como un instrumento de seguridad industrial, con la protección de los derechos fundamentales como un mero elemento de una evaluación más amplia de la conformidad.

También es importante tener en cuenta el contexto global con poderosos actores gubernamentales y empresariales que abogan por directrices y soluciones distintas de

27. Véase art. 2.6 («This Regulation does not apply to AI systems or AI models, including their output, specifically developed and put into service for the sole purpose of scientific research and development»).

28. Véase art. 57 y ss., AI Act. Per *sanbox* regulatoria, l'AI Act intende, ai sensi dell'art. 3.55, «a controlled framework set up by a competent authority which offers providers or prospective providers of AI systems the possibility to develop, train, validate and test, where appropriate in real-world conditions, an innovative AI system, pursuant to a sandbox plan for a limited time under regulatory supervision».

29. Véase art. 60.3, AI Act («The testing of high-risk AI systems in real world conditions under this Article shall be without prejudice to any ethical review that is required by Union or national law»).

las obligaciones legales y (como en el caso del RGPD) Cassandras que proporcionan una larga lista de desgracias para la UE a consecuencia del RIA.

Sobre esta delgada línea se construyó el RIA, dentro de un proceso legislativo que no facilita la interacción con voces ajenas a la industria, margina al mundo académico (con la excepción de las voces favorables a la industria) y compromete a la sociedad civil de forma difusa.

El RIA no es la mejor ley posible, pero es una ley de primera generación. Incluso las primeras leyes de protección de datos estaban muy lejos del RGPD. Esto es normal en la regulación tecnológica, el cruce entre interés económico, innovación y protección de los derechos exige compromisos. Llegarán interpretaciones de este RIA para clarificar y mitigar sus límites, seguirán herramientas de implementación más puntuales —especialmente en relación a los modelos de evaluación de impacto— y, a lo largo de los años, llegarán también nuevas generaciones de leyes sobre IA y un mayor nivel de protección.

El Convenio del Consejo de Europa de inteligencia artificial frente al Reglamento de la Unión Europea: dos instrumentos jurídicos muy diversos

JACQUES ZILLER

Catedrático de derecho público y de la Unión europea, Universidades Paris-1 Panthéon Sorbonne y Pavia

Mientras las instituciones de la Unión Europea trabajaban en la regulación de la inteligencia artificial, lo que dio lugar al RIA, el Consejo de Europa (adelante CdE), que reúne a todos los Estados europeos a excepción de Bielorrusia y Rusia¹, también se encargaba de esta problemática. Sería erróneo decir que las dos organizaciones han trabajado en paralelo: como muestra el propio texto del proyecto de *Convenio marco sobre la inteligencia artificial, los derechos humanos, la democracia y el Estado de derecho* (*Projet de Convention-cadre sur l'intelligence artificielle, les droits de l'homme, la démocratie et l'État de droit / Draft Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law*² en adelante CMLA), ha habido y sigue habiendo una

1. El Estado de la Ciudad del Vaticano tiene personalidad de organismo soberano de Derecho internacional público, distinto de la Santa Sede (es decir, la cabeza de la Iglesia católica de rito romano), y goza de reconocimiento universal, pero tiene una naturaleza especial que explica que, además de las Organizaciones en las que la Santa Sede participa como observador permanente, como el CdE, el Estado de la Ciudad del Vaticano sólo es miembro de algunas OI, como la Unión Postal Universal (UPU), la Unión Internacional de Telecomunicaciones (UIT), el Organismo Internacional de Energía Atómica (OIEA) y la Organización Mundial del Turismo (OMT). V. <https://www.vaticanstate.va/it/stato-governo/note-generali/origini-natura.html>
2. <https://rm.coe.int/cai-2023-28-fr-projet-de-convention-cadre/1680ae19a1>; <https://rm.coe.int/cai-2023-28-draft-framework-convention/1680ade043> El texto aprobado en esta reunión no se ha hecho público hasta mediados de abril y circula por las redes sociales desde el 19 de marzo <https://rm.coe.int/1680afae3d>; <https://rm.coe.int/1680afae3c> A continuación se presentará al Comité de Ministros, que tiene la última palabra, por lo que no se descarta que haya más cambios en esta fase. También se ha publicado una relación explicativa <https://rm.coe.int/1680afae68>; <https://rm.coe.int/1680afae67>

gran interacción entre las instituciones de las dos organizaciones europeas³. Es lo menos que podemos hacer, dado que los 27 Estados miembros de la Unión Europea son también miembros del CdE, junto con otros 19 Estados. Por ello, el Consejo de la UE adoptó el 21 de noviembre de 2022 una decisión por la que se autoriza la apertura de negociaciones Unión Europea para un Convenio del Consejo de Europa sobre Inteligencia Artificial, Derechos Humanos, Democracia y Estado de Derecho⁴. Merece la pena citar algunos considerandos de la decisión.

«(4) La Unión ha adoptado normas comunes que se verán afectadas por los elementos que se prevé incluir en el convenio. Dichos elementos incluyen, en particular, un conjunto completo de normas sobre el mercado único aplicables a los productos y servicios para los que pueden utilizarse sistemas de IA, así como normas de Derecho derivado de la Unión por las que se aplica la Carta de los Derechos Fundamentales de la Unión Europea (UE), teniendo en cuenta que es probable que esos derechos se vean negativamente afectados en determinadas circunstancias por el desarrollo y la utilización de determinados sistemas de IA». Veremos más adelante las particularidades derivadas de los poderes de atribución del CdE y de la UE.

En el considerando (5) se dice que el ámbito de aplicación previsto para el convenio y el de la propuesta de RIA «se superponen en gran medida, ya que ambos instrumentos tienen por objeto establecer normas aplicables a la concepción, al desarrollo y a la aplicación de sistemas de IA proporcionados y utilizados por entidades públicas o privadas». Entonces en el considerando (6): «Por consiguiente, la celebración del convenio puede afectar a normas comunes de la Unión, vigentes y futuras, o alterar su alcance en el sentido del artículo 3, apdo. 2, del (TFUE)». Llama la atención que esté considerando se refiera al apdo. del art. 3 sobre los valores de la Unión, según el cual «La Unión ofrecerá a sus ciudadanos un espacio de libertad, seguridad y justicia sin fronteras interiores, en el que esté garantizada la libre circulación de personas conjuntamente con medidas adecuadas en materia de control de las fronteras exteriores, asilo, inmigración y de prevención y lucha contra la delincuencia». Como veremos, la propuesta de la Comisión se refiere únicamente a las bases jurídicas relativas al mercado interior y no a las relativas a los controles en las fronteras interiores y exteriores de la Unión (artículo 77, apdo. 2, del TFUE).

Cabe señalar asimismo que el Supervisor Europeo de Protección de Datos (SEPD) emitió un informe sobre el proyecto del CdE, al que se hace referencia en una nota de la Decisión del Consejo, con una serie de recomendaciones que se han

3. V. por ejemplo la noticia del CdE de 12 octubre 2023 «La secretaria general, Marija Pejčinović Burić, se ha reunido con el comisario europeo de Justicia, Didier Reynnders. La reunión se ha centrado en la cooperación entre el Consejo de Europa y la Unión Europea y en los preparativos en marcha para el Convenio sobre Inteligencia Artificial». <https://www.coe.int/es/web/portal/-/secretary-general-meets-european-commissioner-for-justice>
4. Decisión (UE) 2022/2349 del Consejo de 21 de noviembre de 2022 por la que se autoriza la apertura de negociaciones en nombre de la Unión Europea con vistas a un convenio del Consejo de Europa sobre inteligencia artificial, derechos humanos, democracia y Estado de Derecho <https://eur-lex.europa.eu/legal-content/ES/TX/?uri=CELEX:32022D2349>

repetido a menudo en las sucesivas versiones del proyecto⁵. En sus observaciones generales, el SEPD observa que el «planteamiento centrado en el mercado está en consonancia con uno de los principales objetivos de la propuesta de RIA, la dimensión de mercado único de la regulación de los sistemas de IA. [...] Al mismo tiempo, el ámbito de competencias del Consejo de Europa es mucho más amplio [...]. En este contexto, el SEPD considera que la Convención representa una importante oportunidad para complementar la propuesta de Ley sobre la IA reforzando la protección de los derechos fundamentales de todas las personas afectadas por los sistemas de IA. En consecuencia [...] el SEPD considera que la salvaguardia de los derechos de las personas y grupos de personas sujetos al uso de sistemas de IA debería ocupar un lugar más destacado entre los objetivos generales de la negociación del convenio»⁶. Como también veremos, el CMIA, desde la versión de diciembre de 2023, incluye disposiciones específicas para la Unión Europea, que son claramente el resultado de la participación de la Comisión en las negociaciones.

El objetivo de esta contribución es poner de relieve las ventajas e inconvenientes de un tratado del CdE, como el CMIA, frente a un reglamento de la Unión Europea, como el RIA. Entonces se presentará solo brevemente el contenido del CMIA, texto del que Lorenzo Cotino Hueso dice con razón que «el Convenio pone *la lírica a la prosa* que supone el RIA. El RIA establece las bases y estructuras de un ecosistema de IA seguro y confiable, el Convenio centra en su impacto en las personas y la sociedad democrática. El RIA es metódico, detallado y preciso, trazando un camino claro a través de la complejidad técnica y jurídica, estableciendo estándares firmes y obligaciones concretas para los proveedores y usuarios o implantadores de sistemas de IA. En contraste, por cuanto a la lírica, el Convenio se eleva para integrar normativamente los valores fundamentales, los principios éticos y los derechos humanos que deben guiar la evolución de la IA»⁷.

I. EL CONTENIDO DEL PROYECTO DE CONVENIO MARCO DEL CONSEJO DE EUROPA

El capítulo II del CMIA está dedicado a las «Obligaciones generales». Según el art. 4 «Cada Parte adoptará o mantendrá las medidas necesarias para garantizar que las actividades realizadas dentro del ciclo de vida de los sistemas de inteligencia artificial sean compatibles con sus obligaciones de protección de los derechos humanos, tal y como se recogen en el Derecho internacional aplicable y en su Derecho interno». Ello implica no sólo la adopción de la normativa y la legislación necesarias (que, para los Estados miembros de la UE, está cubierta en parte por el RIA, que es un instrumento

5. Dictamen 20/2022 del Supervisor Europeo de Protección de Datos sobre la Recomendación de Decisión del Consejo por la que se autoriza la apertura de negociaciones en nombre de la Unión Europea con vistas a un convenio del Consejo de Europa sobre inteligencia artificial, derechos humanos, democracia y Estado de Derecho (solo en inglés) https://www.edps.europa.eu/system/files/2022-10/22-10-13_edps-opinion-ai-human-rights-democracy-rule-of-law_en.pdf
6. Puntos 10 y 11, p. 7.
7. Cotino Hueso, L. «El Convenio sobre inteligencia artificial, derechos humanos, democracia y Estado de Derecho del Consejo de Europa», *Revista Administración & Ciudadanía*, EGAP, 2024, Vol. 19.

directamente aplicable), sino también los recursos humanos y presupuestarios y las medidas de formación e información necesarias.

El artículo 5 especifica que se trata de «medidas para garantizar que los sistemas de inteligencia artificial no se utilicen para socavar la integridad, independencia y eficacia de las instituciones y procesos democráticos, incluido el principio de separación de poderes, el respeto de la independencia del poder judicial y el acceso a la justicia» y que «cada Parte adoptará o mantendrá medidas para proteger sus procesos democráticos en las actividades del ciclo de vida de los sistemas de inteligencia artificial, incluido el acceso equitativo de las personas al debate público y su participación en el mismo, así como su capacidad para formarse libremente una opinión».

El CMIA establece la obligación de adoptar medidas respecto de la «Integridad de los procesos democráticos y respeto del Estado de Derecho» (art. 5) así como «para respetar la dignidad humana y la autonomía individual» (art. 7). Como veremos más adelante, el CMIA forma parte por lo demás de la misión principal del CdE, que es proteger, principalmente mediante instrumentos jurídicos vinculantes, los derechos humanos y el Estado de Derecho en una sociedad democrática, tal como se establece en el Estatuto del Consejo de Europa y en el CEDH⁸.

El capítulo III está dedicado a los «Principios relativos a las actividades realizadas como parte del ciclo de vida de los sistemas de inteligencia artificial», que «establece los principios generales comunes que cada Parte deberá aplicar en relación con los sistemas de inteligencia artificial de manera adecuada a su ordenamiento jurídico interno y a las demás obligaciones del presente Convenio» (art. 6). Se trata de «la dignidad humana y la autonomía individual» (art. 7), «la transparencia y el control» (art. 8), «la obligación de rendir cuentas y la responsabilidad» (art. 9), «la igualdad y la no discriminación» (art. 10), «el respeto de la vida privada y la protección de los datos de carácter personal» (art. 11), «la fiabilidad», es decir «medidas para fomentar la fiabilidad de los sistemas de inteligencia artificial y la confianza en sus resultados, que podrían incluir requisitos relacionados con una calidad y seguridad adecuadas durante todo el ciclo de vida de los sistemas de inteligencia artificial (art.12), y “Innovación segura [...] se pide a cada Parte que permita, según proceda, el establecimiento de entornos controlados para desarrollar, experimentar y probar sistemas de inteligencia artificial bajo la supervisión de sus autoridades competentes” (art. 13). Como dice justamente Cotino “El Convenio no sólo tiene un valor simbólico y metajurídico, sino que es un instrumento normativo, con capacidad de integración casi constitucional en los ordenamientos jurídicos de los Estados parte y cuenta con gran potencial interpretativo. Es por ello que el Convenio IA supera a decenas de instrumentos declarativos y de soft law que ya resultaban superfluos, inocuos e incluso tediosos”».

El Capítulo IV está dedicado a los «recursos» y a las «garantías procesales». Se trata de obligaciones de los Estados Partes, no de un sistema de recursos a nivel del CdE, como veremos más adelante. El Capítulo V trata de la «evaluación y mitigación de riesgos e impactos negativos».

8. V. Ziller, J. *L'État de droit, une perspective de droit comparé* — Conseil de l'Europe, Bruselas, European Parliament Research Service PE 745.673 — 2023. [https://www.europarl.europa.eu/RegData/etudes/STUD/2023/745676/EPRS_STU\(2023\)745676_FR.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2023/745676/EPRS_STU(2023)745676_FR.pdf)

El capítulo VI está dedicado a la «aplicación del Convenio», con disposiciones recurrentes en instrumentos recientes del CdE, relativas a la no discriminación (art. 17), los derechos de las personas discapacitadas y los niños (art. 18), la consulta pública (art. 19), la salvaguardia de los derechos humanos existentes (art. 21), la relación con otros instrumentos jurídicos y una protección más amplia (art. 22 y 23). El art. 20 «Alfabetización y competencias digitales» es más específico de la IA: «Cada Parte fomentará y promoverá la alfabetización digital y las competencias digitales adecuadas para todos los sectores de la población, incluidos competencias especializadas específicas para los responsables de la identificación, evaluación, prevención y mitigación de los riesgos planteados por los sistemas de inteligencia artificial».

El Capítulo VII establece un «mecanismo de seguimiento y cooperación». En cuanto al Capítulo VIII sobre las cláusulas finales es significativo que se requiera únicamente la ratificación de cinco Estados, de los cuales al menos tres deben ser miembros del CdE, evidenciando la intención de activar el Convenio IA lo antes posible.

Como dice acertadamente Cotino «Aunque en general el Convenio IA no se caracteriza por establecer nítidas obligaciones y derechos específicos hay varios motivos para tenerlo normativamente en cuenta. [...] Considero relevante la regulación en el Convenio IA de “principios” generales aplicables a todos los sistemas de IA. A este respecto cabe recordar que, desde hace años, entre decenas de declaraciones y documentos, se han ido visibilizando y destilando unos principios éticos esenciales de la IA. Desde Harvard se analizaron más de treinta de las principales declaraciones internacionales y corporativas de ética de la IA y se sintetizaron en privacidad, rendición de cuentas, seguridad, transparencia y explicabilidad, equidad y no discriminación, control humano, responsabilidad profesional, valores humanos y, sostenibilidad. El futuro Convenio es positivo por cuanto va más allá de las declaraciones en el ámbito del *soft law* y regula estos principios, si se me permite, pasa de las musas de la ética al teatro del Derecho. [...] Sin embargo, hay algunos elementos concretos del Convenio que pueden ir algo más allá del RIA y el Derecho de la UE».

Por último, es necesario presentar las disposiciones específicas resultantes de la participación de la Comisión de la UE en las negociaciones, que comentaremos en la sección 5, sobre las particularidades de la firma y ratificación de los tratados del CdE frente a la adopción de un reglamento o directiva de la UE. Se trata de dos artículos. Secundo el Art. 27 — Efectos del Convenio «1. Si dos o más Partes han celebrado ya un acuerdo o tratado sobre las materias objeto del presente Convenio o han establecido de otro modo sus relaciones con respecto a dichas materias, también tendrán derecho a aplicar ese acuerdo o tratado o a regular esas relaciones en consecuencia, siempre que lo hagan de manera que no sea incompatible con el objeto y la finalidad del presente Convenio. 2. Las Partes que sean miembros de la Unión Europea aplicarán, en sus relaciones mutuas, las disposiciones de la Unión Europea que regulen las materias comprendidas en el ámbito de aplicación del presente Convenio, sin perjuicio del objeto y la finalidad del presente Convenio y sin perjuicio de su plena aplicación con las demás Partes. Lo mismo se aplicará a las demás Partes en la medida en que estén vinculadas por dichas normas.».

En la versión propuesta antes de la última reunión de la CAI, en marzo de 2014, también figuraba una disposición específica para la UE en el Artículo 29 — Solución de controversias «Si surge una controversia entre las Partes relativa a la interpretación o aplicación del presente Convenio que no pueda ser resuelta por la Conferencia de las Partes según lo dispuesto en el párrafo 1 e del artículo 24, las Partes tratarán de solucionar la controversia mediante negociación o cualquier otro medio pacífico de su elección. La Unión Europea y sus Estados miembros no se acogerán, en sus relaciones mutuas, al artículo 29 del Convenio. Los Estados miembros de la Unión Europea tampoco podrán invocar este artículo del Convenio en cualquier litigio entre ellos relativo a la interpretación o aplicación del Derecho de la Unión Europea». Las dos últimas frases no se incluyeron en la versión adoptada el 14 de marzo; de hecho, eran un recordatorio de principios bien conocidos de la legislación de la UE. Como veremos, estas disposiciones deben entenderse a la luz de una posible disposición relativa a las reservas al Convenio.

II. LAS RAZONES DE UN TRATADO DEL CONSEJO DE EUROPA SOBRE INTELIGENCIA ARTIFICIAL

Recuerde que el CdE fue fundado por el Tratado de Londres de 5 de mayo de 1949, firmado por diez Estados europeos⁹ y entró en vigor el 3 de agosto de 1949. Es la más antigua de las organizaciones creadas tras la Segunda Guerra Mundial con el objetivo de reunir a los países europeos que comparten los valores de la democracia liberal. Según el artículo 1 de su Estatuto, el objetivo del CdE es lograr «una mayor unidad entre sus miembros, a fin de salvaguardar y realizar los ideales y principios que constituyen su patrimonio común», entre los que se incluye la «primacía del derecho» (*prééminence du droit / rule of law*), y «facilitar su progreso económico y social»¹⁰. Uno de los objetivos primordiales del CdE es la protección de los derechos humanos, lo que llevó a sus órganos a preparar el Convenio para la Protección de los Derechos Humanos y de las Libertades Fundamentales (CEDH), que se firmó el 4 de noviembre de 1950 y entró en vigor el 3 de noviembre de 1953 tras ser ratificado por ocho Estados miembros; para España fue el 24 noviembre 1977. La Federación Rusa dejó de ser miembro del CdE el 16 de marzo de 2022, después de la agresión militar de Rusia contra Ucrania. Bielorrusia no es miembro de pleno derecho del CdE, ya que no ha firmado el Convenio Europeo de Derechos Humanos. Su participación en grupos de trabajo del CdE también ha sido suspendida, según el CdE. El 17 de marzo de 2022, el CdE suspendió las relaciones con Bielorrusia debido a la «participación activa» del país en la invasión rusa de Ucrania. Ni Rusia ni Bielorrusia han estado representadas en el Comité *Ad hoc* sobre Inteligencia Artificial (CAHAI) instaurado el 11 de septiembre de 2019¹¹

9. Bélgica, Dinamarca, Francia, Irlanda, Italia, Luxemburgo, Noruega, Países Bajos, Reino Unido y Suecia.

10. Traducción oficial en el Instrumento de Ratificación del Convenio para la Protección de los Derechos Humanos y de las Libertades Fundamentales, hecho en Roma el 4 de noviembre de 1950, y enmendado por los Protocolos adicionales números 3 y 5, de 6 de mayo de 1963 y 20 de enero de 1966, respectivamente, <https://www.boe.es/buscar/doc.php?id=BOE-A-1979-24010>

11. Decisión del Comité de Ministros del Consejo de Europa CM/Del/Dec(2019)1353/1.5, 11 de septiembre de 2019.

y, por supuesto, menos aún en su sucesor desde enero de 2022, el Comité sobre Inteligencia Artificial (CAI)¹².

El mandato del CAI se otorgó desde el primero de enero de 2022 hasta el 31 diciembre de 2024 con el «Mandato del CAI»¹³ en el marco del programa «Cumplimiento efectivo del CEDH»¹⁴, el que explica el enfoque del CMIA en Estado de Derecho y derechos humanos. El mandato fue adoptado bajo la autoridad del Comité de Ministros, en el que los Estados miembros del CdE están representados en general por su Representante Permanente en Estrasburgo, excepcionalmente por sus Ministros de Asuntos Exteriores. El Comité de Ministros puede adoptar resoluciones y, en particular, recomendaciones a los Gobiernos de los Estados miembros, en particular sobre el curso que debe darse a las sentencias del Tribunal Europeo de Derechos Humanos, que son vinculantes para los Estados (CEDH art. 46). El Estatuto del CdE especifica (art. 20) los procedimientos de votación. Van desde la mayoría de los representantes con unanimidad de los votos emitidos para las cuestiones más importantes, hasta la mayoría de los representantes con una mayoría de dos tercios de los votos emitidos para la mayoría de las resoluciones, pasando por la mayoría simple para las cuestiones relativas al Reglamento interno o al Reglamento financiero y administrativo.

El Comité se ha «encomendado a la CAI que tenga en cuenta las principales conclusiones y los retos pertinentes expuestos en el informe 2023 del Secretario General sobre el estado de la democracia, los derechos humanos y el Estado de Derecho, titulado “Invitación a un nuevo compromiso con los valores y normas del CdE””. Se trataba de “establecer un proceso de negociación internacional y llevar a cabo trabajos para ultimar un marco jurídico adecuado sobre el desarrollo, el diseño, la utilización y el desmantelamiento de la inteligencia artificial, que se base en las normas del CdE sobre derechos humanos, democracia y Estado de Derecho, así como en otras normas internacionales pertinentes, y que favorezca la innovación, que podrá consistir en un instrumento jurídico vinculante de carácter transversal que incluya, entre otras cosas, principios generales comunes, así como instrumentos adicionales vinculantes o no vinculantes para abordar los retos relacionados con la aplicación de la inteligencia artificial en sectores específicos, de conformidad con las decisiones pertinentes del Comité de Ministros”. Se trataba también de “mantener un enfoque transversal coordinando también su trabajo con otros comités y entidades intergubernamentales del CdE que también se ocupan de las implicaciones de la inteligencia artificial en sus respectivos ámbitos de actividad, proporcionando orientación a estos comités y entidades en consonancia con el marco jurídico en desarrollo y ayudándoles en la resolución de problemas”, así como “basando el trabajo en pruebas sólidas y en un proceso de consulta inclusivo, incluso con socios internacionales y supranacionales para garantizar una visión holística del tema”. Se trataba por último, de “contribuir” a la consecución de la Agenda 2030 de las

12. <https://www.coe.int/fr/web/artificial-intelligence/cai> / <https://www.coe.int/en/web/artificial-intelligence/cai>

13. <https://rm.coe.int/mandat-du-comite-sur-l-intelligence-artificielle-cai-/1680addf7e> / <https://rm.coe.int/terms-of-reference-of-the-committee-on-artificial-intelligence-for-2022/1680a74d2f>

14. <https://www.coe.int/fr/web/civil-society/effective-echr-implementation> / <https://www.coe.int/en/web/civil-society/effective-echr-implementation>

Naciones Unidas para el Desarrollo Sostenible y examinar los progresos realizados al respecto, en particular en relación con el Objetivo 5: Igualdad de género, Objetivo 16: Paz, justicia e instituciones eficaces».

Como sintetizado de Cotino «este mandato tenía una clara intención de trascender fronteras, buscando crear un “instrumento atractivo no sólo para los Estados de Europa sino para el mayor número posible de Estados de todas las regiones del mundo”, involucrando a “Observadores” como Israel, Canadá, Estados Unidos, Japón, la Asociación Mundial sobre Inteligencia Artificial (GPAI), empresas de Internet, y organizaciones de la sociedad civil».

Conforme al artículo 30 apdo. 1 del CMIA, «El presente Convenio estará abierto a la firma de los Estados miembros del CdE, de los Estados no miembros que hayan participado en su elaboración y de la Unión Europea». Recuerden que la Unión Europea es parte en varios tratados del CdE, como el Convenio de Estambul sobre prevención y lucha contra la violencia contra las mujeres y la violencia doméstica ¹⁵ a partir del 1 de enero de 2023, y que la adhesión de la UE al CEDH está prevista del artículo 17 del Protocolo n.º 14 al CEDH, por el que se modifica el sistema de control del Convenio y del artículo 16 TFUE.

Pues, conforme al Artículo 31 del CMIA — Afiliación: «1. Tras la entrada en vigor del presente Convenio, el Comité de Ministros del [CdE] podrá, previa consulta a las Partes en el presente Convenio y obtenido su consentimiento unánime, invitar a cualquier Estado no miembro del [CdE] que no haya participado en la elaboración del Convenio a adherirse al mismo, mediante decisión adoptada por la mayoría prevista en el artículo 20.d del Estatuto del [CdE] y por unanimidad de los representantes de las Partes con derecho a formar parte del Comité de Ministros». Hay varios tratados del CdE a los que se han adherido Estados no europeos, por ejemplo. Canadá, Chile, Costa Rica, Estados Unidos, Japón, México y la Santa Sede suelen ser invitados. En cuanto al Convenio del CdE relativo al blanqueo, seguimiento, embargo y decomiso de los productos del delito y a la financiación del terrorismo (STCE n.º 198), Marruecos, que también ha sido invitado, es el único Estado no miembro que lo ha ratificado. Aunque dejó de ser miembro del CdE en 2022, la Federación Rusa sigue siendo parte de varios convenios, que no ha denunciado, a diferencia del CEDH.

En nuestra opinión, las dos principales razones para redactar un tratado del CdE sobre inteligencia artificial eran disponer de un texto común a todos los Estados europeos, incluido el Reino Unido tras el Brexit, y participar en la carrera mundial por ser los primeros en adoptar una normativa al respecto, con la esperanza de servir de modelo al menos para las democracias pluralistas. Por ejemplo, se puede leer en de la página de noticias-Inteligencia Artificial del CdE: «Los días 5 y 6 de marzo de 2024, la Unidad de Inteligencia Artificial del Consejo de Europa participó en el Diálogo OCDE-Unión Africana (UA) sobre Inteligencia Artificial (IA), patrocinado por el gobierno del Reino Unido y celebrado en la sede de la OCDE en París, para presentar el trabajo del Comité de Inteligencia Artificial (AIC). El acto reunió a miembros de la Comisión de la UA (Argelia, Camerún, República del Congo, Yibuti, Egipto, Etiopía, Kenia), el Grupo de Trabajo de la UA sobre IA y expertos invitados, incluidas otras organizaciones internacionales con mandatos complementarios sobre IA, para debatir

15. Convenio del Consejo de Europa sobre prevención y lucha contra la violencia contra las mujeres y la violencia doméstica (STCE n.º 210), <https://rm.coe.int/1680462543>

la Estrategia Continental de la UA sobre Inteligencia Artificial, la gobernanza de la IA, el fomento de la colaboración y la respuesta a retos comunes. La Sra. Louise Riondel, Cosecretaria del CAI, participó en la sesión titulada “La perspectiva internacional: de las iniciativas globales a la gobernanza global”, durante la cual presentó las actividades del Consejo de Europa en el ámbito de la IA, y más concretamente los trabajos del CAI sobre la Convención Marco sobre IA y la metodología para evaluar los riesgos e impactos de los sistemas de IA (HUDERIA)¹⁶. Esta es solamente una de las muchas actividades del Consejo de Europa en el campo de la inteligencia artificial desde 2019, que puede encontrar fácilmente en el sitio web www.coe.int/ai.

Una razón adicional era, obviamente, intentar proponer un texto que pudiera ser adoptado por un gran número de otros Estados, incluidos los Estados Unidos de América. La participación de estos últimos en las negociaciones, en particular para la última reunión de la CAI del 11 al 14 de marzo de 2024, tuvo sin embargo el efecto de reducir su ámbito de aplicación, en particular porque finalmente se decidió que el Convenio Marco no se aplicaría al sector privado. Por supuesto, esto no impide en modo alguno que un Estado parte del futuro Convenio adopte una legislación más inclusiva, como será el caso de los Estados miembros de la UE a través del RIA.

III. EL INSTRUMENTO DEL CONVENIO MARCO DE FRENTE AL INSTRUMENTO DEL REGLAMENTO

Como ya hemos mencionado, el instrumento utilizado por el CdE para regular la inteligencia artificial es un convenio marco, es decir, un tratado internacional conocido como *Convenio marco sobre la inteligencia artificial, los derechos humanos, la democracia y el Estado de derecho*, mientras que la UE utiliza una normativa conocida como RIA¹⁷.

Hay que subrayar que el término «Ley de inteligencia artificial», utilizado entre paréntesis en el título del RIA en la versión española del borrador publicado por la Comisión Europea el 21 de abril de 2021 —y en la versión adoptada por el Parlamento europeo el 13 marzo 2024— es incorrecto desde el punto de vista jurídico. La razón es que «Ley europea» no existe como instrumento del Derecho de la Unión, al no haber incorporado en el Tratado de Lisboa la nueva categorización de los actos de la Unión contenida en el Tratado Constitucional de 24 de octubre de 2004, que, como es sabido, no entró en vigor por no haber sido ratificado por todos los Estados miembros. El hecho de que las versiones alemana, italiana y neerlandesa (por ejemplo) del texto utilizaban también la palabra «ley» (*Gesetz, legge, wet*) no justificaba la versión española establecida por la Comisión europea. La versión portuguesa decía simplemente *Regulamento inteligência artificial*; la versión francesa *législation sur*

16. <https://www.coe.int/fr/web/artificial-intelligence/-/presentation-of-the-council-of-europe-s-activities-on-artificial-intelligence-ai-during-the-oecd-african-union-ai-dialogue>

17. Reglamento (UE) 2024/... del Parlamento Europeo y del Consejo de ... por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial) Diario Oficial de la Unión Europea,

l'intelligence artificielle es también más correcta porque se trata de un acto legislativo (es decir adoptado por un procedimiento legislativo); la versión inglesa *Artificial Intelligence Act* también es correcta, ya que el reglamento es un acto jurídico de la Unión en el sentido del artículo 288 del TFUE; asimismo la versión danesa utiliza la palabra *Retsakten*, que significa «acto jurídico».

En los últimos años, parece que los servicios de la Comisión tienden a no ser quisquillosos con los títulos del Derecho derivado y adoptan más una actitud de marketing, utilizando términos que se dirigen a un público no jurista; es también verdad que el uso del inglés como idioma principal en la *praxis* institucional —aunque las 24 lenguas oficiales y de trabajo mencionadas en el artículo 55 del TUE y en el Reglamento 1/58¹⁸ tengan el mismo valor jurídico¹⁹— permite mantener cierta ambigüedad, dado que en el Reino Unido una ley se denomina *Act of Parliament*. En realidad, la palabra inglesa que mejor corresponde al español ley es *statute*. Cabe señalar que, a diferencia de la propuesta de RIA o de la propuesta de «Ley Europea de Libertad de los Medios de Comunicación»²⁰ (*European Media Freedom Act*) en los llamados *Digital Markets Act*²¹ *Digital Services Act*²², se menciona entre paréntesis como Reglamento de mercados digitales / servicios digitales en la mayoría de los idiomas, excepto en alemán, donde se utiliza el término *Gesetz*.

Menos mal que el texto final adoptado por el Consejo el 14 de mayo de 2024 se ha corregido y ahora dice «Reglamento de Inteligencia Artificial» en lugar de «Ley».

El derecho del CdE es más sencillo desde este punto de vista formal: no hay diferencia jurídica entre un Convenio, un Convenio Marco, un Acuerdo, un Protocolo, un Arreglo o incluso una Carta, como la Carta Social Europea, hasta el Estatuto del CdE²³; son todos tratados de derecho internacional público. De un total de 226

18. Reglamento n.º 1 por el que se fija el régimen lingüístico de la Comunidad Económica Europea, <https://eur-lex.europa.eu/legal-content/es/ALL/?uri=CELEX%3A31958R0001>

19. V. entre otros Ziller, J. «Le multilinguisme, caractère fondamental du droit de l'Union européenne», Condinanzi, Canizzarro, Adam et al. (eds.), *Liber amicorum Antonio Tizzano. De la Cour CECA à la Cour de l'Union: le long parcours de la justice européenne*, Torino, Giappichelli, 2018, pp. 1067-1082.

20. Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establece un marco común para los servicios de medios de comunicación en el mercado interior (Ley Europea de Libertad de los Medios de Comunicación) y se modifica la Directiva 2010/13/UE, <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A52022PC0457>

21. Reglamento (UE) 2022/1925 del Parlamento Europeo y del Consejo de 14 de septiembre de 2022 sobre mercados disputables y equitativos en el sector digital y por el que se modifican las Directivas (UE) 2019/1937 y (UE) 2020/1828 (Reglamento de Mercados Digitales) <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex%3A32022R1925>

22. Reglamento (UE) 2022/2065 del Parlamento Europeo y del Consejo de 19 de octubre de 2022 relativo a un mercado único de servicios digitales y por el que se modifica la Directiva 2000/31/CE (Reglamento de Servicios Digitales) <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex%3A32022R2065>

23. Salvo indicación contraria, las traducciones de textos del Consejo de Europa sido realizadas por el autor a partir de la versión francesa.

acuerdos firmados a principios de marzo de 2024²⁴ hay tres convenios marco: el Convenio marco europeo sobre cooperación transfronteriza entre comunidades o autoridades territoriales de 21/05/1980 (STE n° 106 *Convention-cadre européenne sur la coopération transfrontalière des collectivités ou autorités territoriales / European Outline Convention on Transfrontier Co-operation between Territorial Communities or Authorities*), el Convenio marco para la Protección de las Minorías Nacionales de 01/02/1995 (STE n° 157 *Convention-cadre pour la protection des minorités nationales / Framework Convention for the Protection of National Minorities*) y el Convenio marco sobre el valor del patrimonio cultural para la sociedad de 27/10/2005 (STCE n° 199 *Convention-cadre sur la valeur du patrimoine culturel pour la société / Framework Convention on the Value of Cultural Heritage for Society*). Según la Dirección General de Democracia y Dignidad Humana del CdE «la única diferencia entre “convenios” y “acuerdos” es la forma en que el Estado puede expresar su consentimiento en obligarse. Los acuerdos pueden firmarse con o sin reservas en cuanto a la ratificación, mientras que los convenios, en principio, siempre deben ser ratificados por el Estado»²⁵. En realidad, esta indicación también no es exacta, ya que es el derecho interno de cada Estado el que determina si es necesaria la ratificación o si es posible una simple firma para vincular al Estado en cuestión. Según el Informe Explicativo del *Convenio marco sobre el valor del patrimonio cultural para la sociedad*, por ejemplo «se trata de una Convención marco que establece principios y amplios campos de acción acordados por los Estados Partes». Sin embargo, hay otros tratados del CdE que afirman lo mismo. Del mismo modo, el hecho de que los Estados que no son miembros del CdE puedan adherirse no es una característica específica de los convenios marco. De hecho, hay que remitirse al considerando n° 11 del CMIA, según el cual «el Convenio se concibe como un marco y puede complementarse con otros instrumentos destinados a abordar cuestiones específicas relacionadas con el diseño, desarrollo, utilización y desmantelamiento de los sistemas de inteligencia artificial».

Como sabemos, las diferencias entre un Tratado del CdE y un Reglamento de la UE se deben esencialmente a que el primero solo es vinculante para los Estados que lo han firmado y, en su caso, ratificado, mientras que el segundo es vinculante para todos los Estados miembros de la UE — salvo que exista una exención excepcional, generalmente temporal, o basada en los protocolos relativos a Irlanda y Dinamarca (y al Reino Unido antes del Brexit—. Además, existen grandes diferencias debido a que las competencias de la UE están definidas de forma mucho más precisa y, por tanto, son más limitadas que las del CdE.

IV. LOS LÍMITES DERIVADOS DE LAS RESPECTIVAS COMPETENCIAS DEL CONSEJO DE EUROPA Y DE LA UNIÓN EUROPEA

Para evitar errores en la comparación entre los textos del CdE y de la UE, los juristas son los más indicados para explicar el origen de ciertas formulaciones. Existe una primera diferencia esencial entre la acción del CdE y la de la UE, a saber, la forma en que se formulan y enmarcan las competencias de ambas organizaciones.

24. <https://www.coe.int/fr/web/conventions/full-list> / <https://www.coe.int/en/web/conventions/full-list>

25. <https://www.coe.int/fr/web/democracy-and-human-dignity/treaties> / <https://www.coe.int/en/web/democracy-and-human-dignity/treaties>

Las organizaciones internacionales no poseen un poder general para actuar; a diferencia de un Estado soberano —cuyas competencias sólo están limitadas por sus obligaciones en virtud de acuerdos internacionales—, las competencias de una OI se limitan a las conferidas por sus Estados miembros, de conformidad con el principio de atribución que se aplica a las organizaciones intergubernamentales — y que se menciona en modo explícito desde el Tratado de Lisboa en los tratados de la UE—.

De acuerdo con el art. 1 del Estatuto del CdE «a) La finalidad del Consejo de Europa consiste en realizar una unión más estrecha entre sus miembros para salvaguardar y promover los ideales y los principios que constituyen su patrimonio común y favorecer su progreso económico y social. b) Esta finalidad se perseguirá a través de los órganos del Consejo, mediante el examen de los asuntos de interés común, la conclusión de acuerdos y la adopción de una acción conjunta en los campos económicos, social, cultural, científico, jurídico y administrativo, así como la salvaguardia y la mayor efectividad de los derechos humanos y las libertades fundamentales. c) La participación de los Miembros en los trabajos del Consejo de Europa no debe alterar su contribución a la obra de las Naciones Unidas y de las restantes organizaciones o uniones internacionales de las que formen parte. d) Los asuntos relativos a la defensa nacional no son de la competencia del Consejo de Europa.» En breve, el único límite material a las competencias del CdE es la exclusión de los asuntos relativos a la defensa nacional.

Por lo que respecta a la Unión Europea, hay que tener en cuenta una serie de disposiciones: el Art. 5 apdo. 2 TUE que afirma que «En virtud del principio de atribución, la Unión actúa dentro de los límites de las competencias que le atribuyen los Estados miembros en los Tratados para lograr los objetivos que éstos determinan. Toda competencia no atribuida a la Unión en los Tratados corresponde a los Estados miembros.» Debe completarse con el art. 2 par 6. TFUE que afirma que «El alcance y las condiciones de ejercicio de las competencias de la Unión se determinarán en las disposiciones de los Tratados relativas a cada ámbito». Esta última redacción, introducida por el Tratado de Lisboa, se limita a poner negro sobre blanco lo que ya estaba claro en el Tratado constitutivo de la Comunidad Europea del Carbón y del Acero de 1951 y en los Tratados de Roma de 1957 por la precisión de sus disposiciones.

Cuando se prevé una iniciativa de acción de la UE, la primera tarea de los juristas de la Comisión, el Consejo y el Parlamento Europeo es, por tanto, comprobar si existe una base jurídica para dicha acción en los tratados. De no ser así, existe un alto riesgo de que los actos adoptados sean impugnados y, tarde o temprano, anulados por el Tribunal de Justicia. Una base jurídica consiste en una o varias disposiciones de los tratados que reúnen los siguientes elementos.

En primer lugar, la acción prevista debe corresponder a un ámbito cuya competencia se haya atribuido a la Unión. Por ejemplo, el mercado interior (artículos 26 y 27 del TFUE, así como 114 y 115, entre otros), la política monetaria (artículos 127 y siguientes del TFUE), la política medioambiental (artículos 191 y siguientes del TFUE), etc. En algunos casos, la competencia se confiere implícitamente y puede deducirse combinando diferentes elementos del «sistema de tratados» según la expresión utilizada a menudo por el Tribunal de Justicia.

En segundo lugar, sólo se puede actuar para alcanzar los objetivos de la Unión. Éstos se mencionan a veces específicamente junto con la disposición que se refiere al

ámbito de actuación (por ejemplo, el artículo 191 del TFUE para la política monetaria); de lo contrario, se derivan de los objetivos más generales del artículo 3 del TUE. Usualmente, los objetivos se enuncian en una redacción cuidadosamente elegida que establece límites a las opciones políticas que pueden adoptarse en el ejercicio de las competencias atribuidas por los Estados miembros a la Unión. Al controlar la legalidad de los actos de Derecho derivado, el Tribunal de Justicia comprueba si sus disposiciones son coherentes con los objetivos establecidos en los Tratados y, en caso contrario, anula el acto en cuestión.

En tercer lugar, sólo se puede actuar utilizando el tipo de acto especificado en la disposición correspondiente. Los artículos de los tratados especifican a menudo si se trata de utilizar directivas, o reglamentos, o decisiones, o dejan la elección entre distintos actos; alternativamente, en muchos casos dejan un margen de elección más amplio con el uso de la palabra «medidas» (véanse, por ejemplo, por el mercado interior, los artículos 114 —medidas— y 115 TFUE —directivas—). En cualquier caso, incluso cuando se utiliza la palabra «medidas», éstas sólo pueden adoptar la forma de los actos previstos en los Tratados, como se desprende del art. 288 TFUE.

En cuarto lugar, para constituir una base jurídica, las disposiciones pertinentes deben precisar el procedimiento que deben seguir las instituciones. Para la adopción de actos legislativos, se hace referencia al procedimiento legislativo ordinario, cuyos detalles se especifican en el art. 294 TFUE, o se indica explícitamente un procedimiento legislativo especial (véanse, por ejemplo, los arts. 114 y 115 TFUE). Para los actos no legislativos, el procedimiento que debe seguirse se especifica en cada caso en la disposición pertinente del Tratado (véanse, por ejemplo, los artículos 108 y 109 del TFUE para el control de las ayudas estatales). Si las bases jurídicas pertinentes para la acción prevista no prevén un tipo de acto que las instituciones desearían utilizar —por ejemplo, un reglamento en lugar de una directiva—, el artículo 352 del TFUE permite adoptar tal acto mediante un procedimiento específico que requiere una decisión unánime del Consejo y la aprobación del PE; en cambio, el artículo 352 no puede utilizarse para actuar en un ámbito no atribuido a la UE. Además, el Derecho derivado establece la base jurídica de los actos de ejecución ulteriores que deben adoptar las instituciones, órganos y organismos de la Unión y, en su caso, las autoridades de los Estados miembros. Estos actos de ejecución deben ajustarse a las disposiciones del Derecho derivado pertinente y, en primer lugar, a los objetivos establecidos en el cuerpo del acto de la Unión o en sus considerandos introductorios.

Es esencial tener en cuenta lo anterior para comprender el marco del Derecho de la UE aplicable a la inteligencia artificial. En efecto, dado el gran número de propuestas de actos y comunicaciones de la Comisión relativos a la digitalización y a la IA publicados en los últimos años, se corre el riesgo de olvidar los límites que el principio de atribución impone a las instituciones de la UE.

Un ejemplo típico son las «Directrices éticas para el uso de la inteligencia artificial (IA) y los datos en la educación y formación para los educadores» publicadas por la Comisión el 25 de octubre de 2022²⁶. Leyendo este documento, así como la descripción del llamado «Espacio Europeo de la Educación»²⁷, parece como si la

26. <https://op.europa.eu/es/publication-detail/-/publication/d81a0d54-5348-11ed-92ed-01aa75ed71a1>

27. <https://education.ec.europa.eu/es/about-eea/the-eea-explained>

Comisión Europea actuara en cierto modo como un Ministerio Europeo de Educación y Universidades. Entre otras cosas, se explica que «La idea de crear un Espacio Europeo de Educación fue respaldada por primera vez por los líderes europeos en la Cumbre Social de 2017, celebrada en Gotemburgo (Suecia). Los primeros paquetes de medidas se adoptaron en 2018 y 2019. [...] En septiembre de 2020, la Comisión expuso en una Comunicación su visión renovada del Espacio Europeo de Educación y las medidas concretas para alcanzarlo. El Consejo de la UE respondió con la Resolución, de febrero de 2021, relativa a un marco estratégico para la cooperación europea en el ámbito de la educación y la formación para el período 2021-2030». Quienes no estén familiarizados con la legislación de la UE podrían esperar una legislación europea relativa, precisamente, a la educación.

Ahora bien, el artículo 165 del TFUE, única base jurídica posible para tal acción específica que se utilizará el procedimiento legislativo ordinario para la adopción de «medidas de fomento, con exclusión de toda armonización de las disposiciones legales y reglamentarias de los Estados miembros», lo que reduce drásticamente las competencias de la Unión en este ámbito. Es cierto que los Estados miembros siguen siendo libres de dar cierto alcance jurídico a los denominados documentos *soft law* adoptados por las instituciones. Sin embargo, tal referencia no significa que se aplica un instrumento de Derecho de la UE.

El texto del Estatuto del CdE es muy sencillo en su aplicación, comparado con las acrobacias que supone encontrar una base jurídica adecuada en la legislación de la UE y garantizar que el texto no vaya más allá de lo que permite el principio de atribución.

Como subrayado en la exposición de motivos de la propuesta de la Comisión «la base jurídica de la propuesta es, en primer lugar, el artículo 114 del Tratado de Funcionamiento de la Unión Europea (TFUE), que trata de la adopción de medidas para garantizar el establecimiento y el funcionamiento del mercado interior. Esta propuesta constituye una parte fundamental de la Estrategia para el Mercado Único Digital de la UE. Su objetivo primordial es garantizar el correcto funcionamiento del mercado interior mediante el establecimiento de normas armonizadas, en particular en lo que respecta al desarrollo, la introducción en el mercado de la Unión y el uso de productos y servicios que empleen tecnologías de IA o se suministren como sistemas de IA independientes. Algunos Estados miembros ya están estudiando normas nacionales destinadas a garantizar que la IA sea segura y se desarrolle y utilice de conformidad con las obligaciones asociadas a los derechos fundamentales. Es probable que esto ocasione dos problemas fundamentales: i) la fragmentación del mercado interno en lo que respecta a elementos esenciales, en particular los requisitos aplicables a los productos y servicios de IA, su comercialización, su utilización, y la responsabilidad y supervisión de las autoridades públicas; y ii) la disminución considerable de la seguridad jurídica de los proveedores y usuarios de sistemas de IA en lo tocante a cómo se aplicarán a dichos sistemas las normas vigentes y nuevas en la Unión. Habida cuenta de la amplia circulación transfronteriza de productos y servicios, la mejor manera de solucionar estos dos problemas es mediante legislación de armonización de la UE».

Para ser más precisos, se trata del artículo apdo. 2 del art. 114 del TFUE «El Parlamento Europeo y el Consejo, con arreglo al procedimiento legislativo ordinario y previa consulta al Comité Económico y Social, adoptarán las medidas relativas a la

aproximación de las disposiciones legales, reglamentarias y administrativas de los Estados miembros que tengan por objeto el establecimiento y el funcionamiento del mercado interior», entonces de la consecución de los objetivos enunciados en el art. 26 apdo. 2 TFUE sobre el mercado interior: «El mercado interior implicará un espacio sin fronteras interiores, en el que la libre circulación de mercancías, personas, servicios y capitales estará garantizada de acuerdo con las disposiciones de los Tratados».

La exposición de motivos añade que «además, dado que la presente propuesta contiene determinadas normas específicas para la protección de las personas en relación con el tratamiento de los datos personales, fundamentalmente restricciones del uso de sistemas de IA para la identificación biométrica remota “en tiempo real” en espacios de acceso público con fines de aplicación de la ley, resulta adecuado basar este Reglamento, en lo que atañe a dichas normas específicas, en el artículo 16 del TFUE». Secundo el art. 16 TFUE «1. Toda persona tiene derecho a la protección de los datos de carácter personal que le conciernan. 2. El Parlamento Europeo y el Consejo establecerán, con arreglo al procedimiento legislativo ordinario, las normas sobre protección de las personas físicas respecto del tratamiento de datos de carácter personal por las instituciones, órganos y organismos de la Unión, así como por los Estados miembros en el ejercicio de las actividades comprendidas en el ámbito de aplicación del Derecho de la Unión, y sobre la libre circulación de estos datos. El respeto de dichas normas estará sometido al control de autoridades independientes».

Llama la atención que entre los documentos citados en el apéndice de la exposición de motivos hay un Anexo IX *Legislación de la Unión en materia de sistemas informáticos de gran magnitud en el espacio de libertad, seguridad y justicia*. Por ejemplo, se refiere al Reglamento de 30 de noviembre de 2017, por el que se establece un Sistema de Entradas y Salidas (SES)²⁸. Las bases jurídicas de este Reglamento son «el Tratado de Funcionamiento de la Unión Europea, y en particular su artículo 77, apartado 2, letras b) y d), relativo a “los controles a los cuales se someterá a las personas que crucen las fronteras exteriores” y “cualquier medida necesaria para el establecimiento progresivo de un sistema integrado de gestión de las fronteras exteriores”; y su artículo 87, apartado 2, letra a), relativo a “la recogida, almacenamiento, tratamiento, análisis e intercambio de información pertinente” en materia de cooperación policial. Dado que estas bases jurídicas prevén también el recurso al procedimiento legislativo ordinario, cabe preguntarse por qué estas disposiciones no se citan también en el texto propuesto por la Comisión. Por otra parte, el artículo 87, apartado 3, del TFUE establece que para la cooperación en operaciones entre la cooperación operativa entre “los servicios de policía, los servicios de aduanas y otros servicios con funciones coercitivas especializados en la prevención y en la detección e investigación de infracciones penales”. El Consejo se pronunciará por unanimidad, previa consulta al Parlamento Europeo. Y el art. 77 apdo. 3 prevede que “el Consejo podrá establecer,

28. Reglamento (UE) 2017/2226 del Parlamento Europeo y del Consejo, de 30 de noviembre de 2017, por el que se establece un Sistema de Entradas y Salidas (SES) para registrar los datos de entrada y salida y de denegación de entrada relativos a nacionales de terceros países que crucen las fronteras exteriores de los Estados miembros, se determinan las condiciones de acceso al SES con fines policiales y se modifican el Convenio de aplicación del Acuerdo de Schengen y los Reglamentos (CE) n.º 767/2008 y (UE) n.º 1077/2011. <https://eur-lex.europa.eu/legal-content/ES/TX-T/?uri=CELEX%3A32017R2226>

con arreglo a un procedimiento legislativo especial, disposiciones relativas a los pasaportes, documentos de identidad, permisos de residencia o cualquier otro documento asimilado. El Consejo se pronunciará por unanimidad, previa consulta al Parlamento Europeo”».

Quizá por eso la Comisión ha evitado citar estos dos artículos, que se aplican, entre otras cosas, al sistema de gobernanza de las disposiciones de RIA.

A todos los que critican el proyecto de la Comisión por no ser suficientemente amplio en materia de IA y por dar demasiado peso a la protección de datos, basta recordarles este punto. En particular, el artículo 114 exige encontrar un vínculo con el mercado interior, es decir, las cuatro libertades de circulación, y no permite adoptar un texto vinculante para las instituciones de la Unión, permite solo adoptar un texto vinculante para los Estados miembros. En cambio, el art. 16 sí lo permite. Esto explica por qué, a diferencia del RGDP²⁹ que es basado sobre el art. 16 TFUE, el Reglamento de 2001 relativo al acceso del público a los documentos³⁰ se aplica solo a las instituciones, órganos y organismos de la Unión y no a los Estados miembros. Este último se basa en el art. 15 TFUE (antiguo art. 255 TCE) en el que se afirma que, entre otros que «El Parlamento Europeo y Consejo, con arreglo al procedimiento legislativo ordinario, determinarán mediante reglamentos los principios generales y los límites, por motivos de interés público o privado, que regulan el ejercicio de este derecho de acceso a los documentos».

Por lo tanto, no es sorprendente que los considerandos y disposiciones del proyecto que afectan directamente a las autoridades públicas sean muy complejos.

V. LA NECESIDAD DE RATIFICAR EL TRATADO DEL CONSEJO DE EUROPA DE FRENTE A LA APLICABILIDAD DIRECTA DEL REGLAMENTO DE LA UNIÓN EUROPEA

A diferencia de las directivas, reglamentos y decisiones de carácter general de la UE, que en principio se aplican directamente a todos los Estados miembros, los tratados del CdE sólo se aplican a los Estados que los han firmado y ratificado si su constitución así lo exige.

El CEDH es el instrumento más importante del CdE en términos generales, empezando por el Estado de Derecho, los derechos y libertades fundamentales y la democracia. A diferencia de sus otros convenios y acuerdos —incluidos los protocolos adicionales al CEDH—, la adhesión al CEDH es obligatoria para todos los Estados miembros del CdE, por lo que es vinculante para los cuarenta y seis Estados del CdE y, por tanto, para todos los Estados miembros de la UE. Por otra parte, el CMIA, como suele ocurrir con los tratados del CdE, sólo será vinculante para los Estados

29. Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos) <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:32016R0679>

30. Reglamento (CE) n° 1049/2001 del Parlamento Europeo y del Consejo, de 30 de mayo de 2001, relativo al acceso del público a los documentos del Parlamento Europeo, del Consejo y de la Comisión <https://eur-lex.europa.eu/legal-content/ES/ALL/?uri=celex%3A32001R1049>

que lo hayan ratificado. Por otra parte, el CMIA, como ocurre normalmente con los tratados del CdE, sólo será vinculante para los Estados que lo hayan ratificado, como se desprende del art. 30, apdo. 3. que se requiera únicamente la ratificación de cinco Estados, de los cuales al menos tres deben ser miembros del CdE.

Salvo que se especifique lo contrario l Estados Partes pueden formular reservas o declaraciones a los Convenios de la CDN en el momento de la firma o al depositar el acta de ratificación. El objeto y efecto de una reserva o declaración puede ser especificar cómo debe aplicarse un tratado en relación con un Estado parte. Las reservas de carácter general no están permitidas con respecto al CEDH; sólo puede formularse una reserva con respecto a una disposición concreta del Convenio «en la medida en que una ley vigente en ese momento en su territorio no se ajuste a dicha disposición». El proyecto de CMIA prevé en su art. 34 — Reservas «No podrán formularse reservas con respecto a ninguna disposición del presente Convenio» con una única excepción prevista en el art. 33 relativa a los Estados federales, que podría ser necesaria para los Estados que no son miembros del CdE, como los Estados Unidos de América o Canadá en particular. Cabe indicar que las disposiciones del art. 27 — Efectos del Convenio sobre Estados miembros de la UE permiten evitar reservas por parte de ésta o de la Unión Europea.

El CEDH es directamente aplicable en la mayoría de los Estados Partes. La aplicación directa significa que el Convenio puede invocarse ante todos los tribunales nacionales. Esto no significa que las instituciones del Estado en cuestión —legislatura, administración y poder judicial— no estén obligadas a respetar el Convenio, sino que puede resultar más complicado para un particular hacer valer los derechos garantizados por el Convenio. El caso de la CMIA es más delicado. Por un lado, habrá que determinar hasta qué punto sus disposiciones son lo suficientemente precisas como para considerarlas *self executive*, lo que variará de un Estado parte a otro. En concreto, hay Estados en los que los propios tribunales se pronuncian sobre la interpretación de un tratado, y otros en los que solicitan una interpretación al Ministerio de Asuntos Exteriores. Además, hay Estados en los que la Constitución considera explícitamente que los tratados son superiores a la ley, y otros en los que la cuestión no está resuelta. Todo ello conducirá a una aplicación nada uniforme de las disposiciones del CMIA, máxime cuando, a diferencia del CEDH o de la Carta de Derechos Sociales, no prevé un órgano judicial como el Tribunal Europeo de Derechos Humanos o un órgano cuasi judicial como el Comité Europeo de Derechos Sociales³¹ para resolver los litigios derivados de su aplicación.

Dicho esto, tanto el Tribunal como el Comité se remiten en su jurisprudencia a todos los instrumentos pertinentes del CdE, así como a otros instrumentos de Derecho internacional como contexto útil para el ejercicio de su jurisdicción.

VI. A MODO DE CONCLUSIÓN

Como conclusión sobre el posible impacto del Acto vinculante del CdE sobre la IA hay que añadir que la jurisprudencia del Tribunal Europeo de Derechos Humanos

31. Salcedo Beltrán, V. C. «La Carta Social Europea y el procedimiento de reclamaciones colectivas: un nuevo y excepcional escenario en el marco legislativo laboral», *Trabajo y Derecho* 91-92/2022, pp. 1-36.

está llamada a evolucionar en el tema de la inteligencia artificial. Un primer signo de ello es el voto particular discrepante del juez Darian Pavli sobre la sentencia de 4 de junio de 2019 en el asunto 39757/15 *Sigurður Einarsson y Otros c. Islandia*³² relativa a la investigación de posibles actos delictivos vinculados a la crisis financiera y, en su caso, al enjuiciamiento de las personas afectadas que son miembros de los órganos de dirección de uno de los mayores bancos islandeses, Kaupþing banki.

En opinión de la mayoría de la Sala, a pesar de las frecuentes quejas ante el fiscal por la falta de acceso a los documentos, en ningún momento los demandantes parecen haber solicitado formalmente a un tribunal el acceso a la «recopilación completa de datos» o que se lleven a cabo nuevas investigaciones, o haber sugerido nuevas medidas de investigación — por ejemplo, una nueva búsqueda utilizando palabras clave sugeridas por ellos—. A este respecto, el Tribunal supremo de Irlanda toma nota de la alegación del Gobierno según la cual las pruebas presentadas ante el Tribunal de Primera Instancia incluían una descripción general de los objetos incautados y de su contenido aproximado. En estas circunstancias, y teniendo en cuenta que los demandantes no proporcionaron ningún detalle sobre el tipo de material que buscaban, el Tribunal Europeo está convencido de que la falta de acceso a los datos en cuestión no era tal como para privar a los interesados de un juicio justo en general.

El juez Pavli, en el párrafo 21 de su dictamen, dijo: «Con el debido respeto a mis colegas, este argumento, en mi opinión, subestima considerablemente la complejidad de analizar grandes cantidades interconectadas de datos de investigación, tanto si se dispone de “mera” inteligencia humana como si se cuenta con la asistencia de inteligencia artificial». Una golondrina no trae un verano, pero es muy probable es muy probable que el Tribunal de Estrasburgo tenga que pronunciarse cada vez más sobre cuestiones relativas a la utilización de sistemas de inteligencia artificial, como ocurrió en materia de protección de datos, donde se basó en particular en el art. 8 del Derecho al respeto a la vida privada y familiar, y que tenga debidamente en cuenta el CMIA, así como el RIA y las leyes y reglamentos nacionales, para construir su jurisprudencia.

32. <https://hudoc.echr.coe.int/fre?i=001-193738>

El Reglamento de inteligencia artificial desde fuera de la Unión Europea: impulsos reguladores desde otras partes del mundo y una visión desde Iberoamérica

JUAN GUSTAVO CORVARÁN — MARÍA VICTORIA CARRO

Director del Laboratorio de Innovación e Inteligencia Artificial de la Facultad de Derecho de la Universidad de Buenos Aires — Doctoranda, Universidad de Génova. Directora de investigación, UBA IALAB

I. INTRODUCCIÓN

El 2023 ha sido testigo de cambios trascendentales en el campo de la inteligencia artificial (en adelante, IA), en el cual todos los actores relevantes han hecho su parte para acelerar una transformación que ha tenido impacto en casi todos los ámbitos del conocimiento y nuestra vida cotidiana.

Por un lado, tras el éxito de ChatGPT, los gigantes tecnológicos se han lanzado en una carrera no sólo por la IA generativa multimodal, sino también por sistemas que sean capaces de cautivar a una mayor cantidad de usuarios y ganarse su preferencia. Luego, los grandes pensadores e investigadores del campo que se han encargado de encender las alarmas y difundir los grandes riesgos de este tipo de tecnologías, han llegado incluso a plasmar sus preocupaciones en una carta abierta que pretendió detener su desarrollo por un período de tiempo que ya expiró¹. Por último, los usuarios, quienes aprenden a aprovechar las nuevas herramientas, ya sea como consumidores finales o reutilizándolas de manera creativa para perfeccionar la precisión en tareas específicas, contribuyendo así al florecimiento del ecosistema.

Pero eso no es todo. El 2023 ha sido también el año en que las regulaciones en materia de IA se han convertido, de una mera voluntad futura, a una prioridad en agenda. A nivel global y frente al escenario descrito en el párrafo anterior, los Estados han comenzado a pensar en medidas concretas y buscar un enfoque activo,

1. Sobre la carta abierta ampliar en: «La carta en la que más de 1.000 expertos piden frenar la inteligencia artificial por ser una “amenaza para la humanidad”» *BBC News Mundo* (2023), disponible en: <https://www.bbc.com/mundo/noticias-65117146> (consultado el 24 de enero de 2024).

incluso por parte de aquellos que tradicionalmente decidían adoptar una postura de no intervención².

En primer lugar, tras un intenso trabajo técnico y político, la Unión Europea consiguió elaborar el texto definitivo de la primera Ley de Inteligencia Artificial en el mundo. Se espera que, en lo que queda del primer semestre de 2024 este Reglamento entre en vigor³.

Por su parte, Estados Unidos no quedó fuera de la ola de regulación y la IA por fin entró con mayor fuerza en el debate político. Pero no fueron solo palabras. La discusión culminó con la Orden Ejecutiva del presidente Biden sobre IA a finales de octubre 2023, cuyo objetivo es mejorar la seguridad de la IA. Sin embargo, este tipo de instrumentos pueden ser revocados en cualquier momento por otro presidente. A diferencia del enérgico trabajo que vienen realizando los representantes de la UE para llegar a consensos, el instrumento estadounidense carece de la legitimidad del Congreso, que de hecho se encuentra tan polarizado que vuelve improbable producir ninguna legislación significativa sobre IA en un corto plazo⁴.

China también emergió como uno de los protagonistas en este panorama normativo. De hecho, ha sido uno de los estados que más ha avanzado en el último tiempo, promulgando leyes individuales sobre distintos aspectos y riesgos que cobran importancia relacionados con estas tecnologías. Su último instrumento ha sido en relación a la IA generativa, que trata cuestiones como la privacidad de los datos y la propiedad intelectual. Sin embargo, en junio de 2023 el Consejo de Estado de China anunció un cambio de enfoque: una ley de inteligencia artificial comprensiva e integral similar a la de la UE estaría en camino⁵.

Finalmente, desde América Latina adelantamos que nuestro continente poco a poco comienza a participar de manera progresiva en los esfuerzos para regular la IA. En términos generales, los países cuentan con planes estratégicos por un lado y recomendaciones éticas por el otro. Asimismo, en todos ellos existe una ley de protección de datos más o menos actualizada. Lo novedoso, sin embargo, son ciertos proyectos de ley generales como los de México y Brasil que examinaremos con mayor profundidad en las próximas líneas.

2. Por ejemplo, Estados Unidos ha sido un país criticado por su pasividad frente a este tipo de tecnologías. Ver Knight Will, «Lluvia de críticas para los países que están ignorando la IA» *MIT Technology Review*, (2019), disponible en: <https://www.technologyreview.es/s/10939/lluvia-de-criticas-para-los-paises-que-estan-ignorando-la-ia> (consultado el 24 de enero de 2024).
3. «El Parlamento Europeo avanza con la legislación para regular la Inteligencia Artificial» *El Observador*, (2024), disponible en: <https://www.elobservador.com.uy/nota/el-parlamento-europeo-avanza-con-la-legislacion-para-regular-la-inteligencia-artificial-2024213104726> (consultado el 22 de febrero de 2024).
4. Ryan-Mosley Tate, «EE UU vs Europa: Biden toma la delantera en la carrera por regular la IA» *MIT Technology Review*, (2023), disponible en: <https://www.technologyreview.es/s/15896/ee-uu-vs-europa-biden-toma-la-delantera-en-la-carrera-por-regular-la-ia> (consultado el 22 de febrero de 2024).
5. Ryan-Mosley Tate, «Vuelta al mundo por las regulaciones de la IA en 2024», *MIT Technology Review*, (2024), disponible en: <https://www.technologyreview.es/s/16069/vuelta-al-mundo-por-las-regulaciones-de-la-ia-en-2024> (consultado el 22 de febrero de 2024).

Desde UBA IALAB, durante el último tiempo hemos realizado un esfuerzo significativo para segmentar y analizar estos documentos con el objetivo de tener una perspectiva detallada de lo que ocurre a nivel regional y global⁶. Estas investigaciones nos han servido como insumo para ser presentadas en las «Jornadas sobre regulación y legislación de la inteligencia artificial: IA generativa y tendencias internacionales» celebradas el 5 de junio de 2023 en la Cámara de Diputados de la Nación Argentina donde se discutieron y elaboraron una serie de recomendaciones como hoja de ruta para abordar una posible regulación de uso e implementación de Inteligencia Artificial en la Argentina⁷.

A lo largo de este trabajo, comentaremos algunos de los aspectos más relevantes y problemáticos de estas y otras iniciativas de regulación, incluyendo un desarrollo especial de las tendencias en algunos países latinoamericanos. En este sentido, se buscará ampliar nuestra comprensión de los efectos, tanto positivos como negativos, de estas regulaciones en diversos y variados sectores y contextos socioeconómicos. En resumen, el propósito será poner de manifiesto la dirección hacia la cual se orientan los esfuerzos legislativos a nivel global.

II. SALIR DE LA OLLA A TIEMPO Y OTROS DESAFÍOS DE LA REGULACIÓN

¿Cómo regular algo que no deja de cambiar? ¿cómo controlar efectos masivos, macros y muchas veces imperceptibles? Europa puso fin a estas incógnitas en un intento de salir a tiempo de la olla de acuerdo a la famosa parábola de la rana hervida⁸, y muchos otros estados acompañaron esta tendencia para no terminar siendo presos de agua hervida. En realidad, fueron muchos los factores coyunturales que contribuyeron a que, a nivel global, la regulación de la IA pase de ser una mera voluntad futura, a una prioridad en agenda.

Antes de 2023, los Estados y las organizaciones internacionales se limitaron a emitir «derecho blando» (*soft law*) en forma de sugerencias o recomendaciones éticas

-
6. Ver Corvalán Juan G. (dirección), Sánchez Caparrós Mariana, Rabán Melisa (coordinación), Stringhini Antonella, Papini Carina Mariel, Heleg Giselle, Bonato Valentín, «Propuestas de regulación y recomendaciones de inteligencia artificial en el mundo. Síntesis de principales aspectos» IALAB UBA, (2023), disponible en: <https://ialab.com.ar/wp-content/uploads/2023/08/Propuestas-de-regulacion-y-recomendaciones-de-IA-en-el-mundo-1.pdf> (consultado el 9 de marzo de 2024).
 7. Sobre los consensos y las recomendaciones producto de las Jornadas ver: «Puntos de partida para la regulación de la inteligencia artificial en Argentina» en Corvalán Juan G. (director), «Tratado de Inteligencia Artificial y Derecho» *Thompson Reuters La Ley*, (2023), 2da edición.
 8. La famosa parábola de la rana hervida nos enseña que, si ponemos uno de estos anfibios en una olla de agua hirviente, inmediatamente intenta salir. En cambio, si lo colocamos en agua a temperatura ambiente, y no lo asustamos, se queda tranquilo. A medida que los grados se elevan, permanece allí sin hacer nada, pero cada vez estará más aturdido hasta que el agua hierva y ya no sea capaz de escapar. Si bien su aparato interno para detectar amenazas a la supervivencia está preparado para cambios repentinos en el medio ambiente, no es capaz de detectar efectos lentos y graduales.

destinadas a los actores relevantes en el campo de la IA⁹. Se trata de principios orientativos de alcance general, ya que la imposición de requisitos obligatorios con frecuencia fue considerada excesiva y hasta precipitada en un campo de constante evolución. Esto, por dos razones. Primero, porque las regulaciones «preventivas» podrían obstaculizar la innovación y como consecuencia sus beneficios¹⁰. Segundo, porque algunas industrias han logrado regularse a sí mismas de manera exitosa guiadas por la presión cultural e institucional¹¹.

Sin embargo, en este supuesto las diferentes presiones tardaron en llegar. De manera similar a como ocurre con los datos personales, los usuarios suelen restar importancia a las afectaciones a sus derechos que no pueden ver o percibir directamente. Hacia finales de 2022, era probable que cualquier persona promedio ajena a la industria tecnológica concibiera a la IA como algo que estaba por venir, que razonablemente podíamos esperar en el futuro. Ello sin tener presente que su cuenta de Netflix ya empleaba IA para ayudarlo a elegir la próxima serie o que mediante esta tecnología los filtros de Instagram reconocen su cara.

Sin embargo, durante el último año, el auge de los grandes modelos de lenguaje (LLMs, por sus siglas en inglés) —o en atención a otras de sus características, conocidos como *Foundation Models* (modelos base)—, lo cambió todo. En particular, la llegada de GPT-4 de la empresa OpenAI llevó a algunos expertos a pensar que nos encontramos frente a una especie de antesala de la superinteligencia¹², y junto con ello, su respectivo espectro de riesgos.

9. Algunos ejemplos de estos documentos éticos son: el Libro Blanco sobre Inteligencia Artificial elaborado por la Comisión Europea en 2020, el primer conjunto de Directrices de Políticas Intergubernamentales sobre IA adoptado en 2019 por los 36 países socios de la OCDE, las Directrices Éticas para una IA Fiable creadas por el Grupo de expertos de alto nivel sobre IA constituido por la Comisión Europea en 2019, y la Recomendación sobre la ética de la Inteligencia Artificial de la UNESCO adoptada por los Estados miembros en 2021.
10. O'sullivan Andrea, «Si los gobiernos controlan demasiado la inteligencia artificial perderemos sus beneficios» *MIT Technology Review*, (2017), disponible en: <https://www.technologyreview.es/s/9688/si-los-gobiernos-controlan-demasiado-la-inteligencia-artificial-perderemos-sus-beneficios> (consultado el 28/2/2024).
11. La autoregulación de la IA ha sido una propuesta para evitar normativas excesivamente restrictivas. Ver Páez Giménez Efrén, «ExCEO de Google propone autorregulación en Inteligencia Artificial» *DPL News*, (2023), disponible en: <https://dplnews.com/exceo-de-google-propone-autoregulacion-en-inteligencia-artificial/> (consultado el 29 de febrero de 2024).

Asimismo, el cofundador de DeepMind, Mustafa Suleyman ha dicho que si bien hace falta regulación de arriba hacia abajo, existen ejemplos de industrias que han logrado autorregularse exitosamente. Ver Douglas Heaven Will, «DeepMind's cofounder: Generative AI is just a phase. What's next is interactive AI» *MIT Technology Review*, (2023), disponible en: https://www.technologyreview.com/2023/09/15/1079624/deepmind-inflection-generative-ai-whats-next-mustafa-suleyman/?utm_source=LinkedIn&utm_medium=tr_social&utm_campaign=site_visitor.unpaid.engagement (consultado el 29 de febrero de 2024).

12. Romero Sarah, «Microsoft afirma que GPT-4 puede razonar como un humano» *Muy Interesante*, (2023), disponible en: <https://www.muyinteresante.com/actualidad/60456.html> (consultado el 3 de marzo de 2024). Asimismo ver: Bubeck et al., «Sparks of Artificial General Intelligence: Early experiments with GPT-4», *arXiv:2303.12712*, (2023), disponible en: <https://arxiv.org/abs/2303.12712> (consultado el 29 de febrero de 2024).

Además, los usuarios comenzaron a evidenciar cómo algunos de los sistemas inteligentes más sofisticados se colaban en varias tareas de sus vidas cotidianas. Desde UBA IALAB hemos documentado el impacto en la productividad en múltiples tareas, junto a otros estudios que también demuestran cómo esta tecnología está cambiando la forma en la que trabajamos a gran escala¹³.

Estos condimentos y otros más, por fin llevaron a la creación de una conciencia generalizada sobre la importancia de controlar la IA, lo que incluyó poner cada vez más en agenda cuestiones vinculada a los riesgos existenciales. De esta forma, las incógnitas sobre regulación de la IA que antes parecían obstáculos casi imposibles de superar se volvieron necesarias de abordar y resolver de alguna forma. De debates periféricos que de vez en cuando eran merecedores de algún artículo de opinión, se han transformado en el centro de la discusión pública y política.

Hasta aquí, ya es posible identificar la primera tendencia. Los principios de IA adoptados por la OCDE en 2019, sirvieron durante todo este tiempo como una referencia global para guiar el resto de las recomendaciones éticas tanto de las organizaciones internacionales como de los gobiernos y ahora, se constituyen también como un punto de partida que ayuda a estos mismos actores a dar forma a una regulación centrada en el ser humano y los valores democráticos para una IA confiable.¹⁴

En lo sucesivo, presentaremos algunas orientaciones más que se tuvieron en cuenta y también, cuales otras se dejaron de lado. Para ello, es importante tener en cuenta que cuando se pregunta por el enfoque de la regulación, no existe una taxonomía generalmente aceptada, sino que la respuesta depende de distintos aspectos que aquí se presentan en forma de dicotomías, por ejemplo, unidad vs. fragmentación, o «derecho duro» vs. «derecho blando». En el medio de cada una de ellas, existe una escala de matices que se materializan en las decisiones efectivamente tomadas por los distintos Estados para dirigir su regulación.

1. UNIDAD VS. FRAGMENTACIÓN. ¿ENFOQUE «HORIZONTAL» O «VERTICAL»?

Uno de los grandes debates a la hora de crear un marco de normas de IA gira en torno a cómo regular algo que es tan heterogéneo. Desde sistemas de reconocimiento facial, pasando por autos autónomos hasta sistemas predictivos y generativos, el espectro de herramientas, funcionalidades, técnicas y niveles de autonomía es tan

el 3 de marzo de 2024). También consultar Aguera y Arcas Blaise, Norving Peter, «Artificial General Intelligence Is Already Here» *NOEMA*, (2023), disponible en: <https://www.noemamag.com/artificial-general-intelligence-is-already-here/> (consultado el 18 de marzo de 2024).

13. Ver Corvalán Juan G., Díaz Dávila Laura Cecilia, Guilera Soledad, Le Fevre Enzo (dirección), «La revolución de la productividad. Cómo impacta la IAGen y ChatGPT en la reducción de tiempos y en la optimización de las tareas» *UBA IALAB*, (2024), disponible en: <https://ialab.com.ar/wp-content/uploads/2024/02/Resumen-Ejecutivo.pdf> (consultado el 18 de marzo de 2024).
14. Morini Bianzino et. al, «The Artificial Intelligence (AI) global regulatory landscape. Policy trends and considerations to build confidence in AI» *EY*, (2024).

variado que con frecuencia los expertos se plantean si una regulación general e integral es el enfoque adecuado.

En primer lugar, esta dicotomía se relaciona a su vez con la postura de quienes nos recuerdan que ya existen normas que regulen la IA. Sostener que dictar reglas que sean aplicables sobre sistemas inteligentes es necesario, implica ignorar grandes tramos de ley existente, porque, de hecho, estas regulaciones ya existen, aunque sean imperfectas¹⁵. Marcos de responsabilidad civil, de contratos, de propiedad intelectual son aplicables incluso en casos que involucren tecnologías autónomas. Sin embargo, esto no significa que los arreglos legales existentes sean óptimos o que no surjan dificultades de interpretación y aplicación frente situaciones cada vez más complejas.

Otra forma de plantear esta dualidad es entre los enfoques «horizontal» o «vertical». En una perspectiva horizontal, los reguladores crean una regulación integral que cubre los muchos impactos que la IA puede tener. En una estrategia vertical, los responsables políticos adoptan un enfoque a medida, creando diferentes regulaciones para dirigirse a diferentes aplicaciones o tipos de IA¹⁶. En este sentido se ha dicho que la Ley de IA de la UE se inclina horizontalmente y las regulaciones de China se inclinan verticalmente¹⁷.

En la introducción de este trabajo hemos dicho que China ha adoptado un enfoque distintivo al regular la IA, promulgando leyes individuales sobre distintos aspectos y riesgos particulares, como las *deepfakes*. Este enfoque práctico ha convertido al gigante asiático al mejor candidato para reaccionar rápidamente y dar respuesta a los cambios originados por las nuevas tecnologías. Tal es así que China ha sido probablemente el primer país del mundo en introducir legislación sobre IA generativa pocos meses después de la gran erupción de ChatGPT¹⁸. Sin embargo, en junio de 2023 el Consejo de Estado ha anunciado un cambio de rumbo: una ley de inteligencia artificial comprensiva e integral estaría en camino. Como es de esperar, no parece que el texto llegará tan rápido como las mencionadas normas específicas nos tenían acostumbrados.

Esta misma dualidad es trasladable a los organismos encargados de la supervisión y aplicación de la ley. Ahora, muchos estados manifiestan a través de sus iniciativas regulatorias, su voluntad es crear un organismo centralizado y especializado para

-
15. Cuellar Mariano-Florentino, «Reconciling Law, Ethics, and Artificial Intelligence: The Difficult Work Ahead» *Stanford University Human-Centered Artificial Intelligence*, (2019), disponible en: <https://hai.stanford.edu/news/reconciling-law-ethics-and-artificial-intelligence-difficult-work-ahead> (consultado el 29 de febrero de 2024).
 16. O'Shaughnessy Matt, Sheehan Matt, «Lessons From the World's Two Experiments in AI Governance», *Carnegie Endowment for International Peace*, (2023), disponible en: <https://carnegieendowment.org/2023/02/14/lessons-from-world-s-two-experiments-in-ai-governance-pub-89035> (consultado el 3 de marzo de 2024).
 17. O'Shaughnessy Matt, Sheehan Matt, «Lessons From the World's Two Experiments in AI Governance», *Carnegie Endowment for International Peace*, (2023), disponible en: <https://carnegieendowment.org/2023/02/14/lessons-from-world-s-two-experiments-in-ai-governance-pub-89035> (consultado el 3 de marzo de 2024).
 18. Yang Zeyi, «Four things to know about China's new AI rules in 2024» *MIT Technology Review*, (2024), disponible en: <https://www.technologyreview.com/2024/01/17/1086704/china-ai-regulation-changes-2024/> (consultado el 1 de marzo de 2024).

asegurar la aplicación efectiva de las normas de IA. Hace poco, la UE ha inaugurado su Oficina de Inteligencia Artificial para fomentar el uso de IA fiable¹⁹. Por su parte, en su proyecto de ley, México planea establecer una entidad conocida como el Consejo Mexicano de Ética para la Inteligencia Artificial y Robótica (CMETIAR) encargado de proponer leyes y vigilar su cumplimiento²⁰.

Incluso Estados Unidos, que siempre se ha enorgullecido de tener un mosaico de autoridades federales y estatales que examinan las partes que les tocan de estas tecnologías, ha considerado la idea de concentrar estas facultades en un único organismo. En una comparecencia ante el Congreso, senadores de ambos partidos y Sam Altman, CEO de OpenAI, afirmaron que era necesaria una nueva agencia federal para proteger a los ciudadanos de la IA perjudicial²¹.

El problema de tal iniciativa en este punto es el solapamiento con el trabajo vigente de otros entes públicos, tanto a nivel estatal como federal. Por ejemplo, la Administración Nacional de Seguridad del Tráfico en las Carreteras se encarga de los autos autónomos y el Departamento de Seguridad Nacional ha publicado informes sobre posibles amenazas a infraestructuras críticas a manos de estas tecnologías. Asimismo, la Comisión Federal de Comercio y la Administración de Alimentos y Medicamentos, reglamentan la manera en que las empresas utilizan la IA. Además, ya ha habido antecedentes de proyectos de ley respecto de los cuales ciertos legisladores se abstienen de votar porque su promulgación anularía la legislación estatal en la materia²². Como si ello fuera poco, se ha dicho que un enfoque descentralizado o fragmentado evita obstaculizar la industria²³.

En realidad, si bien algunos Estados tienden a acercarse más a uno u otro extremo, la realidad es que termina adoptándose un enfoque dual, es decir, tanto intersectorial como específico para cada sector²⁴. El primer enfoque, intersectorial, proporciona un marco de referencia de salvaguardias fundamentales, independientemente del

-
19. «La UE inaugura su Oficina de Inteligencia Artificial para fomentar su uso fiable», *El Tiempo*, (2024), disponible en: <https://www.eltiempo.com/tecnosfera/novedades-tecnologia/la-union-europea-inaugura-su-oficina-de-inteligencia-artificial-857260> (consultado el 1 de marzo de 2024).
 20. González Fernanda, «Presentan propuesta de ley para regular la IA en México» *The Wired*, (2023), disponible en: <https://es.wired.com/articulos/diputado-presenta-propuesta-de-ley-para-regula-la-ia-en-mexico> (consultado el 1 de marzo de 2024).
 21. Johnson Kari, «Asustados por ChatGPT, legisladores de EE UU quieren crear un organismo regulador de la IA», *The Wired*, (2023), disponible en: <https://es.wired.com/articulos/legisladores-de-ee-uu-quieren-crear-organismo-regulador-de-inteligencia-artificial-y-chatgpt> (consultado el 1 de marzo de 2024).
 22. Esto es lo que ha sucedido con los legisladores de California al tenes que decidir sobre la regulación federal de privacidad. Ver: Johnson Kari, «Asustados por ChatGPT, legisladores de EE UU quieren crear un organismo regulador de la IA», *The Wired*, (2023), disponible en: <https://es.wired.com/articulos/legisladores-de-ee-uu-quieren-crear-organismo-regulador-de-inteligencia-artificial-y-chatgpt> (consultado el 1 de marzo de 2024).
 23. O'sullivan Andrea, «Si los gobiernos controlan demasiado la inteligencia artificial perderemos sus beneficios» *MIT Technology Review*, (2017), disponible en: <https://www.technologyreview.es/s/9688/si-los-gobiernos-controlan-demasiado-la-inteligencia-artificial-perderemos-sus-beneficios> (consultado el 28/2/2024).
 24. Morini Bianzino et. al, «The Artificial Intelligence (AI) global regulatory landscape. Policy trends and considerations to build confidence in AI» *EY*, (2024).

sector en el que se desarrolle o utilice la IA. El segundo enfoque, específico para el sector, establece directrices u obligaciones adicionales para el uso de la IA para abordar riesgos y vulnerabilidades dentro de ámbitos específicos²⁵. Mientras el primer marco tiende a ser corto, poco detallado y aspira a perdurar en el tiempo, la regulación exhaustiva viene de la mano de los expertos que se encuentran cerca del campo de aplicación.

El Marco Modelo de Gobernanza de la IA de Singapur, por ejemplo, proporciona orientación independiente del sector a las organizaciones privadas para alinearse con los principios rectores sobre el uso ético de la IA. De forma complementaria, la Autoridad Monetaria de Singapur (MAS) emitió una guía sectorial específica para el sector financiero sobre equidad, ética, responsabilidad y transparencia en el uso de la IA y el análisis de datos²⁶.

2. SEGUNDO. OBLIGATORIEDAD VS. VOLUNTARIEDAD Y EL FORTALECIMIENTO DE LA COLABORACIÓN

Para evitar que la regulación decepcione, es necesario que los políticos y legisladores que promueven y participan en la elaboración de normas que controlan la IA, entiendan en profundidad esta tecnología²⁷. Con frecuencia, los riesgos de la IA son exagerados y se propagan mitos, lo que conduce a una sobrestimación de la urgencia y rigurosidad necesarias en las regulaciones. A su vez, ello genera respuestas desproporcionadas y mal dirigidas cuando no se cuenta con una evaluación crítica y experta que tome en consideración, entre otros aspectos, los intereses subyacentes en juego.

En realidad, no solo hace falta un diálogo multidisciplinario, sino también multisectorial. El verdadero desafío que plantea la regulación de la IA es cómo lograr conciliación y colaboración entre dos mundos tan diferentes: por un lado, los cuerpos políticos de los Estados y sus respectivos asesores, con una perspectiva centrada en el impacto y las necesidades sociales, que insumen tiempo en negociar y buscar consensos con el objetivo de asegurar una IA segura a través de estándares como la transparencia. Por el otro, las empresas más o menos grandes de la industria tecnológica, inmersas en una carrera por la innovación intentando proteger sus intereses económicos, más conscientes que cualquiera sobre las limitaciones y potencialidades de la tecnología.

Algunas tensiones muy evidentes y ya teorizadas surgen en relación a lo anterior como las dificultades de los legisladores para comprender algunos conceptos técnicos básicos, y el ritmo rápido en que las empresas avanzan y que, a su vez, los

25. Morini Bianzino et. al, «The Artificial Intelligence (AI) global regulatory landscape. Policy trends and considerations to build confidence in AI» *EY*, (2024).

26. Morini Bianzino et. al, «The Artificial Intelligence (AI) global regulatory landscape. Policy trends and considerations to build confidence in AI» *EY*, (2024).

27. Knight Will, «Los políticos necesitan entender cómo funciona la Inteligencia Artificial (urgentemente)» *The Wired*, (2023), disponible en: <https://es.wired.com/articulos/inteligencia-artificial-los-politicos-deben-aprender-rapido-sobre-ia> (consultado el 1 de marzo de 2024).

gobiernos y sus leyes no son capaces de seguir²⁸. No obstante, aparte de eso, una cuestión más desafiante es el hecho de que, el cumplimiento de requisitos que las leyes imponen de acuerdo al estado del arte, en muchos supuestos podría depender, pura y exclusivamente de la voluntad de los gigantes tecnológicos. Veamos.

Mientras la actuación estatal en materia del control de IA se limitaba —y se limita en muchos casos—, a documentos éticos, las empresas privadas tienen pocos incentivos para preocuparse por la ética de la IA. Primero, porque todos los requisitos y recaudos que proponen los organismos internacionales, como producir y publicar información, implican invertir dinero. Si luego de ello, deben hacer transparentes sus proyectos tal vez esta inversión estaría justificada. Pero si por el contrario, no se les exige dar información ni siquiera de los sistemas inteligentes creados para el sector público, entonces, desde el punto de vista del negocio, seguir los recaudos éticos o no, simplemente da lo mismo. A lo sumo, el hecho de hacerlo podría servir para una buena campaña de marketing.

No puede pasarse por alto que nunca antes los valores y la ética empresarial ante la IA había sido un tema tan recurrente dentro del ámbito económico y social como lo es hoy en día. Mientras pequeñas *startups* pueden pasar desapercibidas, las grandes consultoras y gigantes tecnológicas lideran el camino por estar constantemente bajo escrutinio, o por lo menos eso parece. Elon Musk fue uno de los primeros en solicitar regulación. La carta abierta también se hizo eco de esta cuestión²⁹. En mayo de 2023, Sam Altman pidió al Congreso de Estados Unidos que regule la IA³⁰. Sundar Pichai, CEO de Google no se quedó atrás y reiteró el pedido³¹.

También es cierto que, estos llamados a la acción estatal vienen seguidos y justificados por advertencias sobre ciertos riesgos sino existenciales, muy graves, que la tecnología podría causar en nuestra sociedad. Resulta curioso que, las personas que piden límites y encienden alarmas, sean las mismas que crean estas herramientas. Probablemente, presentarse como el creador de la IA que más se

28. Jonhson Bobbie, «El problema legal de la IA: cómo regular algo que no deja de cambiar» *MIT Technology Review*, (2019), disponible en: <https://www.technologyreview.es/s/11060/el-problema-legal-de-la-ia-como-regular-algo-que-no-deja-de-cambiar> (consultado el 1 de marzo de 2024).

29. En la carta abierta para pausar el desarrollo de la IA se estableció que la pausa debe ser pública y verificable, e incluir a todos los actores clave. Si tal pausa no se puede implementar rápidamente, los gobiernos deberían intervenir e instituir una suspensión. Ver «La carta en la que más de 1.000 expertos piden frenar la inteligencia artificial por ser una amenaza para la humanidad» *BBC News Mundo*, (2023), disponible en: <https://www.bbc.com/mundo/noticias-65117146> (consultado el 18 de marzo de 2024).

30. Kang Cecilia, «OpenAI's Sam Altman Urges A.I. Regulation in Senate Hearing» *The New York Times*, (2023), disponible en: <https://www.nytimes.com/2023/05/16/technology/openai-altman-artificial-intelligence-regulation.html> (consultado el 1 de marzo de 2024).

31. «¿Por qué el CEO de Google pide que se regule mejor la inteligencia artificial?» *Mix* (2023), disponible en: https://gestion.pe/mix/gente/inteligencia-artificial-google-ceo-pide-que-se-regule-la-inteligencia-artificial-openai-chatgpt-bard-noticia/#-google_vignette (consultado el 4 de marzo de 2024).

acerca a la superinteligencia hasta el momento, pueda convencer a los usuarios de que el servicio de uno es más potente y mejor que el de los competidores³².

El punto es que si bien los esfuerzos éticos son valorables, nunca estarán por delante de los intereses lucrativos. Aún queda un gran trayecto pendiente para un sector corporativo que comienza a preocuparse cada vez más. Las empresas privadas no están acostumbradas a responder preguntas ni dar a conocer sus procesos. La cultura empresarial lleva a *priorizar* el desarrollo y lanzamiento de productos sin una atención suficiente a las implicaciones éticas. Sumado a ello, el sector tecnológico se vuelve cada vez más competitivo y mantener la información de productos y servicios protegidos por secreto comercial, garantiza a las compañías y a sus ventajas competitivas mantenerse lejos de la imitación directa.

Ahora bien, cuando estas recomendaciones éticas se vuelven normas obligatorias, nos encontramos frente a un panorama distinto. La transparencia, ahora se materializa en requisitos concretos, como el etiquetado de contenidos generados artificialmente y la publicación de más información sobre los datos con los que se ha entrenado un modelo base³³, por ejemplo. El problema sigue estando, sin embargo, cuando el cumplimiento de los mismos y su alcance depende del estado del arte, y este último básicamente, es definido por el mismo pequeño grupo de empresas a las cuales se le aplican las normas.

Veamos algunos ejemplos concretos. La Ley de IA de la UE establece una serie de requisitos para todos los modelos fundacionales, base o generales. Sin embargo, añade algunos otros cuando se trata de sistemas de este tipo más potentes en función de la potencia informática necesaria para entrenarlos. Si bien no conoce si el límite incluiría modelos como GPT-4 o Gemini, solo las propias empresas creadoras saben cuánta potencia de cálculo utilizaron para entrenar sus modelos. Tal como reconoció un funcionario de la Comisión Europea, a medida que se desarrolle la tecnología, se debería cambiar la forma de medir y reconocer esta potencia³⁴, para volverla, asimismo, más transparente.

Algo similar puede llegar a ocurrir con la vigilancia humana. La norma europea dedica el artículo 14 a la supervisión humana eficaz, que tendrá por objeto prevenir o reducir los riesgos de los sistemas de IA categorizados como de alto riesgo. Si bien se aclara en el apartado 3 que ello se garantizará mediante medidas que sean técnicamente viables, existe el riesgo de que en algunos supuestos esta posibilidad se vuelva obsoleta, especialmente en entornos dinámicos y de alta complejidad.

Para ejemplificar, incluso en el ámbito del lenguaje natural, ya pueden identificarse tareas que son significativamente difíciles de supervisar por parte

32. Richards Blake, Aguera y Arcas Blaise, Lajoie Guillaume y Sridhar Dhanya, «The Illusion Of AI's Existential Risk» *Noema*, (2023), disponible en: <https://www.noemamag.com/the-illusion-of-ais-existential-risk/> (consultado el 1 de marzo de 2024).

33. Heikkilä Melissa «Las cinco claves sobre la Ley de la inteligencia artificial de la UE» *MIT Technology Review*, (2023), disponible en: <https://www.technologyreview.es/s/15997/las-cinco-claves-sobre-la-ley-de-la-inteligencia-artificial-de-la-ue> (consultado el 1 de marzo de 2024).

34. Heikkilä Melissa «Las cinco claves sobre la Ley de la inteligencia artificial de la UE» *MIT Technology Review*, (2023), disponible en: <https://www.technologyreview.es/s/15997/las-cinco-claves-sobre-la-ley-de-la-inteligencia-artificial-de-la-ue> (consultado el 1 de marzo de 2024).

de un ser humano. Imagine la capacidad de resumir texto, en la cual el evaluador debe conocer con profundidad tanto el texto que se resume como el texto resumido. Esto le requiere al humano insumir una gran cantidad de tiempo prestando mucha atención, lo que facilita la comisión de errores. Como si fuera poco, deberá repetir el ejemplo un número considerable de veces, ya que un solo escrito no es suficiente para valorar una habilidad.

Frente a este enfoque poco práctico, las empresas tecnológicas e investigadores han desarrollado métodos automatizados de evaluación. Esto es, un sistema de IA examinando a otro. Pero si aun así la intención es que la supervisión sea humana, para lograr un enfoque correcto de alineación y otros propósitos valiosos que lo justifican, se continúan pensando enfoques para lograr que ello sea eficaz, conveniente y posible, al menos en alguna medida.

Por ejemplo, OpenAI ha creado una prueba de debate, en que dos agentes artificiales discuten un tema entre sí y el ser humano juzga el intercambio. Incluso si estos sistemas tienen una comprensión más avanzada del problema que el juzgador, el humano puede ser capaz de juzgar qué agente tiene el mejor argumento (similar a los testigos expertos que argumentan para convencer a un jurado)³⁵. Otra posibilidad que se ha postulado consiste en descomponer tareas sumamente intrincadas que el humano no podría ni juzgar ni realizar, como diseñar un sistema de tránsito complicado o administrar cada detalle de la seguridad de una gran red de computadoras, en subtareas o componentes más pequeños que pueden ser valorados³⁶.

3. TERCERO. UN ENFOQUE BASADO EN RIESGOS

Si hay un aspecto de la norma europea que ya está causando el famoso «efecto Bruselas» es el enfoque basado en la identificación de riesgos. Países como Australia y Canadá también trazaron la línea que separa los sistemas de alto riesgo para imponer requisitos y obligaciones más estrictos. A grandes rasgos, se trata de crear categorías de sistemas de acuerdo a su riesgo, y asignar obligaciones de cumplimiento para cada una de ellas. Su beneficio es que permite una intervención regulatoria temprana centrada en la prevención del daño al tiempo que impone costos proporcionales a los posibles impactos negativos.

Esta tendencia incluso es promovida desde los grupos de cooperación internacional. Los países miembros del G7 (Canadá, Francia, Alemania, Italia, Japón, Reino Unido, Estados Unidos y la UE) expresaron una visión unificada sobre la IA y pidieron que las políticas en la materia estén basadas en el riesgo³⁷. Asimismo, alcanzaron un acuerdo sobre los Principios Rectores Internacionales sobre la IA y sobre un Código de conducta para desarrolladores.

35. Amodei Dario, Irving Geoffrey, «AI safety via debate», *OpenAI Blog*, (2018), disponible en: <https://openai.com/research/debate> (consultado el 2 de marzo de 2024).

36. Christiano Paul, Amodei Dario, «Learning complex goals with iterated amplification» *OpenAI*, (2018), disponible en: <https://openai.com/research/learning-complex-goals-with-iterated-amplification> (consultado el 2 de marzo de 2024).

37. Morini Bianzino et. al, «The Artificial Intelligence (AI) global regulatory landscape. Policy trends and considerations to build confidence in AI» *EY*, (2024).

En primer lugar, es preciso aclarar que la legislación de la UE sobre IA propuesta contiene obligaciones *ex ante* para garantizar la seguridad, la ciberseguridad y la protección de los derechos fundamentales, así como normas de responsabilidad *ex post* para compensar los daños cuando se materialice un riesgo de IA³⁸. El enfoque basado en el riesgo se ubica en el primer grupo de medidas preventivas.

En particular, la Ley de IA clasifica las herramientas de IA en distintos grupos: sistemas de riesgo inaceptable como ciertos tipos de biometría, los cuales prohíbe; sistemas de alto riesgo que podrían tener un impacto adverso en la seguridad o los derechos fundamentales como sistemas de contratación, los cuales deberán cumplir una serie de requisitos específicos; sistemas de riesgo limitado, como chatbots que estarán sujetos a normas de transparencia y por último, sistemas de riesgo mínimo para los cuales se establecerán medidas voluntarias.

Por otro lado, Estados Unidos también se ha adherido al enfoque regulatorio basado en los riesgos. La Orden Ejecutiva del presidente Biden³⁹ menciona el abordaje de los riesgos de la IA en diferentes oportunidades. Sin embargo, no define categorías ni criterios para la clasificación.

En este sentido, puede tomarse como referencia el Marco de Gestión de Riesgos de Inteligencia Artificial del Instituto Nacional de Estándares y Tecnología (NIST)⁴⁰, desarrollado en colaboración de los sectores públicos y privado y publicado en enero de 2023. Este marco es aplicable como guía voluntaria tanto para los formuladores de políticas que desarrollan regulaciones de IA basadas en riesgos como para las empresas que consideran cómo organizar su gobernanza interna de la IA.

Este Marco prevé un conjunto de funciones centrales para el manejo de los riesgos: gobernar, mapear, medir y gestionar. Cada una de estas funciones de alto nivel se divide en categorías y subcategorías que, a su vez, contienen en acciones y resultados específicos. Se espera que el proceso sea llevado a cabo por un equipo diverso de forma interactiva y no lineal. Ello, con el objetivo de crear oportunidades para sacar a la luz problemas e identificar riesgos existentes y emergentes.

38. Kretschmer Martin, Kretschmer Tobias, Peukert Alexander, Peukert Christian, «The risks of risk-based AI regulation: taking liability seriously», *ArXiv:2311.14684v1* (2023).

39. La Orden Ejecutiva se encuentra disponible en: <https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/> (consultado el 3 de marzo de 2024).

40. El Marco se encuentra disponible en: <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf> (consultado el 3 de marzo de 2024).



Fuente: AI Risk Management Framework (AI RMF 1.0), NIST, 2023

III. IMPULSOS REGULADORES DESDE OTRAS PARTES DEL MUNDO

Hasta aquí, hemos evidenciado las principales tendencias de regulación proporcionando ejemplos concretos. A continuación nos centraremos en algunos Estados en particular, destacando cuestiones relevantes de su marco regulatorio. Para complementar estas explicaciones, hemos desarrollado un cuadro comparativo que resume la forma en que los diferentes Estados han resuelto las dicotomías presentadas en el apartado anterior.

1. Australia

En los últimos meses, el gobierno de Australia ha sido criticado por su falta de acción en torno al aprovechamiento de oportunidades y la respuesta frente a los riesgos que plantea la IA⁴¹. Si bien es cierto que su reacción ha sido más lenta que la de otros países, ya ha dado sus primeros pasos en torno a la regulación. Tras una consulta pública sobre IA segura y responsable que ha demostrado el deseo de protecciones fuertes por parte de la sociedad australiana a través de más de 500 contestaciones, el gobierno ha emitido una respuesta provisional a comienzos del 2024⁴².

41. Taylor Josh, «Australia “at the back of the pack” in regulating AI, experts warn» *The Guardian*, (2023), disponible en: <https://www.theguardian.com/australia-news/2023/nov/07/australia-ai-artificial-intelligence-regulations-back-of-pack> (consultado el 3 de marzo de 2024).

42. El documento se encuentra disponible en la página oficial del Departamento de Industria, Ciencia y Recursos del Gobierno Australiano: <https://consult.industry.gov>.

El texto informa que se encuentran desarrollando requisitos obligatorios en entornos de alto riesgo únicamente, porque allí los daños ocasionados serán imposibles de revertir. Para estos, las pruebas previas con el objetivo de garantizar la seguridad, la transparencia y la rendición de cuentas serán ejes centrales. Asimismo, se trabaja con la industria para desarrollar un estándar de seguridad y opciones de etiquetado para los contenidos generados por IA, ambos voluntarios.

En este sentido, el enfoque consiste en combinar obligaciones específicas sobre la IA de alto riesgo con una «ley blanda» voluntaria de tacto más ligero para usos menos arriesgados⁴³. El propósito es lograr un equilibrio que le demuestre a los ciudadanos una respuesta activa frente a sus preocupaciones y la intención de proteger a los consumidores, alineada con los desarrollos internacionales, pero que a su vez logre fomentar la adopción de la IA y la innovación a través de la colaboración estrecha con la industria.

Por otro lado, el impacto práctico de la propuesta sigue sin quedar claro dada la incertidumbre del alcance de la definición de los usos de la IA de alto riesgo. En el documento de consulta se dieron dos ejemplos: robots utilizados en cirugía y vehículos autónomos. Esta definición se considerará durante una nueva consulta junto con las obligaciones que podrán imponerse, por lo que las empresas que se consideren afectadas deben considerar una activa participación.

Dejando de lado los planes futuros y concentrándonos en la legislación existente, Australia ha mostrado un fuerte compromiso con algunas regulaciones sectoriales. Por ejemplo, en cuanto a la IA generativa, ha emitido una Declaración de posición de tendencias tecnológicas sobre Inteligencia Artificial Generativa, un documento que examina el panorama de esta tecnología, identificando ejemplos de mal uso y potenciales riesgos. Asimismo, repasa desafíos y enfoques de la regulación, estableciendo que el comisionado eSafety —oficina emisora del documento— utiliza un enfoque multifacético que implica prevención, protección y cambio proactivo y sistémico⁴⁴.

Asimismo, en relación a este último tipo de IA, a nivel federal existe un Marco Australiano para Inteligencia Artificial Generativa en las Escuelas⁴⁵ y una Guía Interina para el Uso del Gobierno de Herramientas Públicas de Inteligencia Artificial

au/supporting-responsible-ai (consultado el 3 de marzo de 2024). Asimismo para un resumen ver: «Action to help ensure AI is safe and responsible» *Minister of Industry and Science*, (2024), disponible en: <https://www.minister.industry.gov.au/ministers/husic/media-releases/action-help-ensure-ai-safe-and-responsible> (consultado el 3 de marzo de 2024).

43. Lincoln Julian, Wilkinson Susannah, Lundie Alex, «Australian Government announces mandatory regulation for high-risk AI» *Herbert Smith Freehills*, (2024), disponible en: <https://www.herbertsmithfreehills.com/insights/2024-01/australian-government-announces-mandatory-regulation-for-high-risk-AI> (consultado el 3 de marzo de 2024).

44. El documento se encuentra disponible en: <https://www.esafety.gov.au/sites/default/files/2023-08/Generative%20AI%20-%20Position%20Statement%20-%20August%202023%20.pdf> (consultado el 3 de marzo de 2024).

45. El Marco Australiano para IA generativa en las escuelas se encuentra disponible en: <https://www.education.gov.au/schooling/resources/australian-framework-generative-artificial-intelligence-ai-schools> (consultado el 3 de marzo de 2024).

Generativa⁴⁶. Por otro lado, a nivel estatal puede mencionarse la Guía Básica de Inteligencia Artificial Generativa del Gobierno de Nueva Gales del Sur⁴⁷ y la Guía de Uso de Inteligencia Artificial Generativa del Gobierno de Queensland⁴⁸.

Otro ejemplo sectorial relacionado con la IA en cuya regulación se viene trabajando intensamente es el caso de los vehículos autónomos y los distintos documentos y análisis de la Comisión Nacional de Transporte de Australia⁴⁹. Uno de los más recientes es el Marco Regulatorio de los Vehículos Autónomos emitido en 2022⁵⁰ que presenta propuestas finales de modificaciones de la legislación actual para dar cabida al uso y despliegue comercial de los vehículos sin conductor.

2. CANADÁ

La parte tercera del proyecto de Ley de la Carta Digital C-27, implementaría la Ley de Inteligencia Artificial y Datos (AIDA, por sus siglas en inglés) para regular el desarrollo responsable de la IA en el mercado canadiense⁵¹. Esta norma, sigue la tendencia global adoptando un enfoque basado en el riesgo. En realidad, se regulan los sistemas de «alto impacto» cuya precisión y requisitos específicos se desarrollarán tras realizar una consulta con las partes interesadas, un proceso similar al propuesto por Australia.

En realidad, se planea que la regulación defina los criterios para identificar sistemas de IA de alto impacto, para que las actualizaciones puedan producirse de forma más ágil a medida que avanza la tecnología. Eso se debe a que los beneficios y riesgos de la IA todavía están surgiendo, y ni siquiera los expertos en tecnología pueden predecir hacia dónde se dirigirá el mercado de la IA.

46. La Guía Interina para el Uso del Gobierno de Herramientas Públicas de IA Generativa se encuentra disponible en: <https://architecture.digital.gov.au/guidance-generative-ai> (consultado el 3 de marzo de 2024).

47. La Guía Básica de Inteligencia Artificial Generativa del Gobierno de Nueva Gales del Sur se encuentra disponible en: <https://www.digital.nsw.gov.au/policy/artificial-intelligence/generative-ai-basic-guidance> (consultado el 3 de marzo de 2024).

48. La Guía de Uso de Inteligencia Artificial Generativa del Gobierno de Queensland se encuentra disponible en: <https://www.forgov.qld.gov.au/information-and-communication-technology/qgea-policies-standards-and-guidelines/use-of-generative-ai-in-queensland-government> (consultado el 3 de marzo de 2024).

49. Otros documentos que pueden mencionarse son: «Directrices nacionales de aplicación de la ley para vehículos automatizados» de 2017, disponible en: https://www.ntc.gov.au/sites/default/files/assets/files/AV_enforcement_guidelines.pdf (consultado el 3 de marzo de 2024) y el «Documento de discusión sobre seguro de lesiones por accidentes de motor y vehículos automatizados» de 2018, disponible en: <https://www.ntc.gov.au/sites/default/files/assets/files/NTC%20Discussion%20Paper%20-%20Motor%20Accident%20Injury%20Insurance%20and%20Automated%20Vehicles.pdf> (consultado el 3 de marzo de 2024).

50. El Marco Regulatorio de los Vehículos Autónomos emitido en 2022 se encuentra disponible en: <https://www.ntc.gov.au/sites/default/files/assets/files/NTC%20Policy%20Paper%20-%20regulatory%20framework%20for%20automated%20vehicles%20in%20Australia.pdf> (consultado el 3 de marzo de 2024).

51. Los detalles de la norma se encuentran disponibles en: <https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document#s10> (consultado el 3 de marzo de 2024).

El Gobierno actualmente considera que son factores clave: 1. Riesgos de daño a la salud, la seguridad o los derechos humanos, basado tanto en el propósito previsto como en las posibles consecuencias no deseadas; 2. La gravedad de los posibles daños; 3. La escala de uso; 4. La naturaleza de los daños o impactos adversos que ya han ocurrido; 5. La medida en que, por razones prácticas o legales, no es razonablemente posible optar por no participar en ese sistema; 6. Desequilibrios de las circunstancias económicas o sociales, o la edad de las personas afectadas; y; 7. El grado en que los riesgos están adecuadamente regulados por otra ley.

Asimismo, se adelanta que las obligaciones de los sistemas de IA de alto impacto se guiarían por los principios de: supervisión y monitoreo humano, transparencia, equidad, seguridad, responsabilidad, validez y robustez.

Asimismo, AIDA prevé varios tipos de sanciones por incumplimiento normativo. Primero, sanciones monetarias administrativas que podrían ser impuestas por el regulador frente a cualquier violación con el fin de fomentar el cumplimiento. Segundo, enjuiciamiento de infracciones reglamentarias, las cuales están previstas para casos más graves en los cuales la culpabilidad debe demostrarse más allá de toda duda razonable. Finalmente, un mecanismo separado para los delitos penales cuando se viole la prohibición de un comportamiento consciente o intencional y se cause un daño grave.

Mientras se continúa trabajando en este proyecto, el Ministro de Innovación, Ciencia e Industria anunció el «Código de Conducta Voluntario sobre el Desarrollo y la Gestión Responsables de los Sistemas Avanzados de Inteligencia Artificial Generativa»⁵², que proporciona temporalmente a las empresas canadienses estándares comunes y les permite demostrar, voluntariamente, que están desarrollando y utilizando este tipo de sistemas de manera responsable hasta que la regulación formal esté en vigor. Todo ello, con el fin de fortalecer la confianza de los ciudadanos en estos sistemas.

IV. UNA PERSPECTIVA DESDE LATINOAMÉRICA

En cuanto a los países latinoamericanos, como hemos adelantado, en los últimos años los esfuerzos de regulación de la IA se han concentrado en recomendaciones éticas, planes estratégicos y actualizaciones de los ecosistemas de protección de datos personales. Sin embargo, hubo proyectos de ley que consistieron en intentos de regulación integrales, en los países de México, Chile y Brasil. Los describiremos a continuación.

1. MÉXICO

El diputado Ignacio Loyola presentó un primer proyecto de ley para establecer un marco legal alrededor del uso y desarrollo de la IA en el país. Se llamaría, «Ley para la Regulación Ética de la Inteligencia Artificial y la Robótica» y tendría el objetivo

52. El Código de Conducta Voluntario sobre el Desarrollo y la Gestión Responsables de los Sistemas Avanzados de Inteligencia Artificial Generativa se encuentra disponible en: <https://ised-isde.canada.ca/site/ised/en/voluntary-code-conduct-responsible-development-and-management-advanced-generative-ai-systems> (consultado el 3 de marzo de 2024).

de regular el uso de esta tecnología con fines gubernamentales y económicos para que este sea siempre basado en la ética y el derecho. La propuesta consiste en crear un organismo burocrático llamado «Consejo Mexicano de Ética para la Inteligencia Artificial y Robótica (CMETIAR)» que servirá como una plataforma en la que confluirán profesionales de distintos sectores para desarrollar y proponer nuevas normas⁵³.

Esta nueva entidad, además estaría conformada por un representante del Poder Ejecutivo designado por el presidente de México, así como por miembros del Consejo Nacional de Humanidades, Ciencias y Tecnologías, de la Comisión Nacional de Derechos Humanos (CNDH) y del Congreso de la República. También por civiles y algunos jugadores de la iniciativa privada⁵⁴. Su función, consistirá en revisar protocolos éticos, vigilar el cumplimiento de las normas y entregar informes de este monitoreo.

Por último, como prácticas prohibidas se establece el uso de la IA y la Robótica con fines de manipulación social, discriminación o violación al estado de derecho.

Sin embargo, en los últimos días de febrero 2024, el senador Ricardo Monreal presentó en la Gaceta del Senado la Iniciativa con proyecto de decreto por el que se expide Ley Federal que Regula la Inteligencia Artificial⁵⁵. El mismo cuenta con 25 artículos y adopta un enfoque de riesgo similar al de la Unión Europea. Los principales aspectos de su enfoque se han analizado en el cuadro comparativo que acompaña este artículo.

Además de estos proyectos, México cuenta con Recomendaciones para el tratamiento de datos personales derivado del uso de la Inteligencia Artificial (2022)⁵⁶ y con una Agenda Nacional de IA (2020)⁵⁷. Este último posee un capítulo de ética que trata las siguientes cuestiones: libertad de expresión, privacidad, igualdad y no discriminación, derechos humanos y democracia. Cabe destacar que se dedica un análisis específico a grupos minoritarios como los pueblos indígenas y las mujeres en problemáticas tales como la brecha digital. Finalmente se elaboran recomendaciones dirigidas a los diferentes actores, como el mantenimiento de canales de diálogo con la ciudadanía, la creación de sandbox regulatorios y la medición de los riesgos.

53. González Fernanda, «Presentan propuesta de ley para regular la IA en México», *The Wired*, (2023), disponible en: <https://es.wired.com/articulos/diputado-presenta-propuesta-de-ley-para-regula-la-ia-en-mexico> (consultado el 2 de marzo de 2024).

54. González Fernanda, «Presentan propuesta de ley para regular la IA en México», *The Wired*, (2023), disponible en: <https://es.wired.com/articulos/diputado-presenta-propuesta-de-ley-para-regula-la-ia-en-mexico> (consultado el 2 de marzo de 2024).

55. Riquelme Rodrigo, «Ricardo Monreal presenta iniciativa para regular la inteligencia artificial» *El Economista*, (2024), disponible en: <https://www.economista.com.mx/tecnologia/Ricardo-Monreal-presenta-iniciativa-para-regular-la-inteligencia-artificial-20240227-0055.html> (consultado el 18 de marzo de 2024).

56. El documento de Recomendaciones para el tratamiento de datos personales derivado del uso de la Inteligencia Artificial se encuentra disponible en: <https://home.inai.org.mx/wp-content/documentos/DocumentosSectorPublico/RecomendacionesP-DP-IA.pdf> (consultado el 9 de marzo de 2024).

57. El documento de Agenda Nacional de IA de México se encuentra disponible en: https://36dc704c-0d61-4da0-87fa-917581cbce16.filesusr.com/ugd/7be025_6f45f669e2fa4910b32671a001074987.pdf (consultado el 9 de marzo de 2024).

2. CHILE

En abril de 2023, se presentó en el Congreso de Chile un proyecto de ley sobre robótica, inteligencia artificial y tecnologías conexas que busca regular los sistemas de IA. Esta iniciativa legislativa realiza una clasificación de los sistemas de IA basada en el riesgo, dividiéndolos en sistemas de riesgo inaceptable y sistemas de alto riesgo, de manera similar al enfoque europeo.

Por otro lado, Chile cuenta con una Política Nacional de Inteligencia Artificial⁵⁸, lanzada por el Ministerio de Ciencia, Tecnología, Conocimiento e Innovación que contiene un capítulo de ética, aspectos legales y regulatorios e impactos socioeconómicos. Dentro de él se establecen los siguientes objetivos:

1. Impulsar la construcción de certezas regulatorias sobre los sistemas de IA que permitan su desarrollo, respetando los derechos fundamentales de acuerdo con la Constitución y las leyes.

2. Impulsar la transparencia algorítmica.

3. Realizar análisis prospectivos para detectar activamente las ocupaciones más vulnerables, anticipar la creación de nuevos empleos por IA y apoyar a los trabajadores en la transición a nuevas ocupaciones, minimizando sus costos personales y familiares.

4. Proveer apoyo a los trabajadores frente a la automatización.

5. Fomentar un uso de IA en el comercio digital transparente, no discriminatorio y respetuoso de las normas de protección de datos personales.

6. Promover un sistema de Propiedad Intelectual actualizado, capaz de fomentar y fortalecer la creatividad y la innovación basada en IA, recompensando a los creadores e innovadores de manera de incentivarlos a hacer pública su creación e innovación y que así la sociedad toda pueda beneficiarse de ella.

7. Posicionar la IA como un componente relevante en el ámbito de la ciberseguridad y ciberdefensa, promoviendo sistemas tecnológicos seguros.

8. Fomentar la participación de mujeres en áreas de investigación y desarrollo relacionadas a la IA hasta alcanzar un nivel igual o mayor a la OCDE.

9. Fomentar la participación de mujeres en áreas de IA en la industria hasta alcanzar, al menos, un valor igual o superior al promedio OCDE y velar porque el impacto de automatización no perjudique por género y que la creación de empleo sea equitativa.

10. Fomentar la equidad de género en la implementación de sistemas de IA.

58. La Política Nacional de Inteligencia Artificial se encuentra disponible en: https://www.minciencia.gob.cl/uploads/filer_public/bc/38/bc389daf-4514-4306-867c-760ae7686e2c/documento_politica_ia_digital_.pdf (consultado el 9 de marzo de 2024).

3. URUGUAY

En 2020 Uruguay⁵⁹ aprobó la Estrategia de Inteligencia Artificial para el Gobierno Digital⁶⁰, con el objetivo de promover y fortalecer el uso responsable de IA en su Administración Pública.

En aquel documento se enumeran nueve principios generales que deben guiar el diseño, desarrollo y despliegue de sistemas inteligentes:

1. Finalidad: los sistemas inteligentes deben potenciar las capacidades humanas, complementarlas y mejorar la calidad de vida de las personas.

2. Interés general: los sistemas inteligentes impulsados desde el Estado deben tender al interés general, garantizar la inclusión y la equidad, reducir sesgos no deseados en datos y modelos y no incurrir en prácticas discriminatorias.

3. Respeto de los Derechos Humanos: los sistemas inteligentes deben respetar los Derechos Humanos, las libertades individuales y la diversidad.

4. Transparencia: los sistemas inteligentes que se utilicen en el sector público deben ser transparentes y cumplir con la normativa vigente, para lo que se debe poner a disposición los algoritmos y datos utilizados para su entrenamiento y puesta en práctica, así como las pruebas y validaciones realizadas, y visibilizar explícitamente todos aquellos procesos que utilicen IA, sea como apoyo o para tomar decisiones.

5. Responsabilidad: los sistemas inteligentes deben tener un responsable claramente identificable que responda por las consecuencias derivadas del accionar de la solución.

6. Ética: cuando los sistemas inteligentes presenten dilemas éticos, éstos deben ser abordados y resueltos por seres humanos.

7. Valor agregado: los sistemas inteligentes sólo deben usarse cuando agreguen valor a un proceso. La IA no debe ser un fin en sí mismo.

8. Privacidad por diseño: los sistemas inteligentes deben contemplar la privacidad de las personas desde su diseño.

9. Seguridad: los sistemas inteligentes deben cumplir con los principios básicos de seguridad de la información desde su diseño.

59. El análisis de los principios éticos que recogen los documentos emitidos por el Gobierno de Uruguay es parte de una investigación que se ha realizado desde UBA IALAB centrada en los principios éticos de IA que se han elaborado, tanto por distintos gobiernos nacionales como por los principales actores de la industria y por organismos internacionales. Ver: Sánchez Caparrós Mariana, «Principios éticos para una inteligencia artificial antropocéntrica: consensos actuales desde una perspectiva global y regional» en Corvalán Juan G. (director) «Tratado de Inteligencia Artificial y Derecho» *Thompson Reuters La Ley*, (2023), 2da edición.

60. La Estrategia de Inteligencia Artificial para el Gobierno Digital se encuentra disponible en: <https://www.gub.uy/agencia-gobierno-electronico-sociedad-informacion-conocimiento/comunicacion/publicaciones/estrategia-inteligencia-artificial-para-gobierno-digital/estrategia> (consultado el 9 de marzo de 2024).

4. BRASIL

En Brasil existe tanto un proyecto de ley para regular la IA (Proyecto de Ley 2338/2023)⁶¹ como una estrategia nacional. En cuanto al primero, los principales aspectos de su enfoque se han analizado en el cuadro comparativo que acompaña este artículo. Sin embargo, la peculiaridad de la norma que escapa a la segmentación y por lo tanto añadimos aquí es la disposición sobre responsabilidad civil.

Esta última establece que el proveedor u operador del sistema de inteligencia artificial que cause daño material, moral, individual o colectivo está obligado a repararlo íntegramente, independientemente del grado de autonomía del sistema: a) Cuando se trata de un sistema de inteligencia artificial de alto riesgo o riesgo excesivo, el proveedor u operador responde objetivamente por el daño causado, en la medida de su participación en el daño; b) Cuando no se trate de un sistema de inteligencia artificial de alto riesgo o riesgo excesivo, se presumirá la culpa del causante del daño, aplicándose la inversión de la carga de la prueba a favor de la víctima.

Asimismo, otro aspecto interesante es la previsión de sandboxes regulatorios que deben cumplir los siguientes requisitos: a) innovación en el uso de la tecnología o en el uso alternativo de tecnologías existentes; b) mejoras hacia ganancias de eficiencia, reducción de costos, mayor seguridad, reducción de riesgos, beneficios para la sociedad y consumidores, entre otros; c) plan de discontinuidad, con previsión de medidas a tomar para asegurar la viabilidad operativa del proyecto una vez que el período de autorización regulatorio del sandbox haya finalizado.

Por otro lado, en línea con las directrices de la OCDE, la Estrategia de Brasil para la Inteligencia Artificial⁶², publicada en el año 2021, se basa en los cinco principios para una IA responsable definidos por dicha organización, que deben ser seguidos en todas las etapas de desarrollo y uso de la IA pudiendo, inclusive, ser elevados a requisitos normativos para todas las iniciativas gubernamentales en la materia.

Los principios a los que refiere el documento son:

1. Crecimiento y desarrollo inclusivo, desarrollo sostenible y bienestar: los sistemas inteligentes deben beneficiar a las personas y al planeta, impulsar el crecimiento inclusivo y el desarrollo sustentable, así como el bienestar.

61. Los principales aspectos de este proyecto han sido analizados en un trabajo reciente publicado por UBA IALAB. Ver Corvalán Juan G. (dirección), Sánchez Caparrós Mariana, Rabán Melisa (coordinación), Stringhini Antonella, Papini Carina Mariel, Heleg Giselle, Bonato Valentín, «Propuestas de regulación y recomendaciones de inteligencia artificial en el mundo. Síntesis de principales aspectos» *IALAB UBA*, (2023), disponible en: <https://ialab.com.ar/wp-content/uploads/2023/08/Propuestas-de-regulacion-y-recomendaciones-de-IA-en-el-mundo-1.pdf> (consultado el 9 de marzo de 2024).

62. El análisis de los principios éticos que recogen los documentos emitidos por el Gobierno de Brasil es parte de una investigación que se ha realizado desde UBA IALAB centrada en los principios éticos de IA que se han elaborado, tanto por distintos gobiernos nacionales como por los principales actores de la industria y por organismos internacionales. Ver: Sánchez Caparrós Mariana, «Principios éticos para una inteligencia artificial antropocéntrica: consensos actuales desde una perspectiva global y regional» en Corvalán Juan G. (director) «Tratado de Inteligencia Artificial y Derecho» *Thompson Reuters La Ley*, (2023), 2da edición.

2. Valores centrados en el ser humanos y la equidad: la IA debe respetar el Estado de Derecho, los derechos humanos, los valores democráticos y la diversidad, e incluir los resguardos adecuados para garantizar una sociedad más justa.

3. Transparencia y explicabilidad: se debe garantizar la transparencia y la divulgación responsable en relación con los sistemas inteligentes, de acuerdo con las reglas del arte, que permitan promover una comprensión general sobre estos sistemas, que las personas sean conscientes de cuando interactúan con IA, que los afectados puedan comprender cómo se ha producido el resultado y que los afectados adversamente puedan cuestionarlo.

4. Robustez, seguridad y protección: los sistemas deben ser robustos y seguros a lo largo de su ciclo de vida, y los riesgos potenciales deben ser gerenciados continuamente.

5. Rendición de cuentas y responsabilidad: dependiendo de la aplicación de IA y los riesgos asociados, se deben establecer estructuras de gobernanza que aseguren la adopción de los principios para una IA confiable y los mecanismos para su observancia. La idea de rendición de cuentas debe guiarse por el principio de precaución.

6. Transparencia y explicación: los sistemas deben poder brindar información significativa y comprensible —sin comprometer la confidencialidad del modelo— sobre su diseño, funcionamiento e impacto, tanto para los desarrolladores como para los usuarios e individuos que puedan verse afectados por sus decisiones y resultados.

7. Privacidad: los sistemas deben respetar la intimidad de las personas, evitar utilizar información que no hayan autorizado y el perfilamiento.

8. Control humano de las decisiones: cuando se trate de sistemas con relativa autonomía en la toma de decisiones el ser humano siempre debe estar en control, especialmente en la etapa de implementación. El control humano debe ser proporcional al nivel de riesgo de los sistemas.

9. Seguridad: los sistemas no deben vulnerar la integridad y la salud física y mental de las personas. La seguridad y confidencialidad de los datos personales y en especial de los datos sensibles es fundamental para evitar afectaciones a la seguridad física y mental de los individuos.

10. Responsabilidad: se debe partir de la solidaridad de los diversos actores que intervienen en el ciclo de vida de los sistemas por los daños que su uso pueda provocar a las personas.

11. No discriminación: los sistemas no pueden tener resultados o respuestas que afecten los derechos de grupos específicos o poblaciones históricamente marginadas; deben adoptar un enfoque de neutralidad de género y garantizar que ese parámetro no sea empleado como factor de discriminación; no pueden estar limitados a un grupo específico por motivos de raza, sexo, religión, edad, discapacidad u orientación sexual.

12. Inclusión: se debe dar participación a los grupos históricamente marginados en el ciclo de vida de los sistemas, así como en su evaluación.

13. Prevalencia de los derechos de niños, niñas y adolescentes: los sistemas de inteligencia artificial deben reconocer, respetar y privilegiar los derechos de niños niñas y adolescentes; siempre deben respetar su interés superior; se los debe empoderar y educar para que puedan tomar parte efectiva.

14. Beneficio social: los sistemas deben permitir o estar directamente relacionados con una actividad que genere un beneficio social claro y determinable. Aquellos que persigan otro tipo de fines no deben ser implementados en el sector público y se debe desincentivar su uso en otros sectores.

5. COLOMBIA

Colombia⁶³ es uno de los estados latinoamericanos que más documentos ha emitido para tratar distintos aspectos de la IA demostrando su compromiso con la creación de un ecosistema responsable. Por un lado, existen estrategias nacionales, agendas y planes: Plan estratégico para la transferencia de conocimiento en Inteligencia Artificial⁶⁴, Plan de Seguimiento a la Implementación en Colombia de Principios y Estándares Internacionales en Inteligencia Artificial⁶⁵ y la Política Nacional para la Transformación Digital e Inteligencia Artificial⁶⁶, entre otros. Por otro, también se han emitido criterios para sandbox regulatorios como Modelo Conceptual para el Diseño de *Regulatory Sandboxes & Beaches* en Inteligencia Artificial (documento borrador para discusión)⁶⁷ y

63. El análisis de los principios éticos que recogen los documentos emitidos por el Gobierno de Colombia es parte de una investigación que se ha realizado desde UBA IALAB centrada en los principios éticos de IA que se han elaborado, tanto por distintos gobiernos nacionales como por los principales actores de la industria y por organismos internacionales. Ver: Sánchez Caparrós Mariana, «Principios éticos para una inteligencia artificial antropocéntrica: consensos actuales desde una perspectiva global y regional» en Corvalán Juan G. (director) «Tratado de Inteligencia Artificial y Derecho» *Thompson Reuters La Ley*, (2023), 2da edición.
64. El documento Plan estratégico para la transferencia de conocimiento en Inteligencia Artificial se encuentra disponible en: https://dapre.presidencia.gov.co/AtencionCiudadana/DocumentosConsulta/consulta-Plan-estrategico-transferencia-conocimiento-inteligencia-artificial-210708.pdf?TSPD_101_R0=08394a-21d4ab2000ad47a40d2942398a3afd43b1cf6ddc68ee01a62e6b7ddb4ba90e5fef6630a4608e2a81956143000a21c50a82d22231f3752d884d7f114087af3c80c0db6ca300c0a7476cfb73e4532ed193a19e700d58d63817dba9c2eae (consultado el 9 de marzo de 2024).
65. El documento Plan de Seguimiento a la Implementación en Colombia de Principios y Estándares Internacionales en Inteligencia Artificial se encuentra disponible en: https://dapre.presidencia.gov.co/TD/plan-seguimiento-implementacion-colombia-estandares-internacionales-inteligencia-artificial-ocde.pdf?TSPD_101_R0=08394a21d4ab20003ce781987b45f801b436fefee21570395b2f0af80498840c752d7f9356e396f508f3d002e214500049b04c4c1af8bc686cdc6b0aedc6392a3f57fcc1b8445a48cb55659b6841af5a10357db7c1294aa242aefd7aa3202b95e19da334e85bbf489163be0308e025c7655769e9ae6b38f2593551645e60ed63 (consultado el 9 de marzo de 2024).
66. La Política Nacional para la Transformación Digital e Inteligencia Artificial se encuentra disponible en: <https://colaboracion.dnp.gov.co/CDT/Conpes/Economicos/3975.pdf> (consultado el 9 de marzo de 2024).
67. El documento se encuentra disponible en: https://dapre.presidencia.gov.co/AtencionCiudadana/DocumentosConsulta/consulta-200820-MODELO-CONCEPTUAL-DISENO-REGULATORY-SANDBOXES-BEACHES-IA.pdf?TSPD_101_R0=08394a21d4ab20003a42e18525bea92cf2a46d01179eb2f6f8cce49d0b07777314d10d54b571ead00826b165171430002696a9a61e05251cfe99fbf7caef2d41ace97aad23d8712bd8f78dabd992658305ff1241c103fa8683ef189469120601c (consultado el 9 de marzo de 2024).

Sandbox sobre privacidad desde el diseño y por defecto en proyectos de inteligencia artificial⁶⁸.

En octubre de 2021 Colombia aprobó su Marco Ético para la Inteligencia Artificial, con el objetivo de proteger, reforzar y garantizar los derechos humanos en el desarrollo, uso y gobernanza de la IA. En este documento se reconocen los siguientes principios éticos.

6. ARGENTINA

La Subsecretaría de Tecnologías de la Información aprobó en junio de 2023 las «Recomendaciones para una Inteligencia Artificial Fiable»⁶⁹. Con esta iniciativa, el país busca garantizar el desarrollo responsable y beneficioso de la IA fortaleciendo el ecosistema científico y tecnológico.

El documento, que consta de unas 30 páginas, está dirigido principalmente a quienes formen parte del sector público y elabora una serie de recomendaciones éticas específicas a tener en cuenta en cada etapa del ciclo de vida de estas tecnologías. Por ejemplo, antes de comenzar es apropiado conformar un equipo diverso y multidisciplinario y explorar otros tipos de tecnologías menos costosas que puedan resolver el problema. Luego, en cuanto a los datos se abordan cuestiones como la privacidad, la calidad y la validación, entre otras.

Argentina se suma así a los esfuerzos internacionales en materia de ética de la IA, tomando en cuenta antecedentes como la Recomendación sobre la Ética de la Inteligencia Artificial de la UNESCO, la Conferencia de Asilomar, las reuniones del Consejo de Ministros de la OCDE y la reunión ministerial sobre Comercio y Economía Digital del G20. De este modo recoge y analiza todos los principios éticos contenidos en estos documentos o elaborados por estos grupos.

El documento identifica el enfoque de IA centrada en el ser humano con la exigencia de que los respectivos actores respeten el Estado de Derecho, los derechos humanos y los valores democráticos durante todo el ciclo de vida del sistema de IA. Estos valores incluyen la libertad, la dignidad y la autonomía, la privacidad y la protección de datos, la no discriminación y la igualdad, la diversidad, la equidad, la justicia social y los derechos laborales reconocidos internacionalmente.

Asimismo, se mencionan dos tipos de modelos sobre los que se puede optar para adoptar IA. Uno de ellos es la automatización, en el cual la intervención humana se limita al control del sistema. Este paradigma es el que se encuentra más asociado a la idea de reemplazo de las capacidades de las personas por las nuevas funciones de los sistemas inteligentes. El segundo es el de *human-in-the-loop*, que implica la colaboración humano-máquina para resolver problemas. En este, se incluye de manera selectiva la participación de las personas, para aprovechar los beneficios o los

68. El documento se encuentra disponible en: https://www.sic.gov.co/sites/default/files/normatividad/112020/031120_Sandbox-sobre-privacidad-desde-el-dise-no-y-por-defecto.pdf (consultado el 9 de marzo de 2024).

69. Las Recomendaciones para una Inteligencia Artificial Fiable se encuentran disponibles en: <https://www.boletinoficial.gob.ar/detalleAviso/primera/287679/20230602> (consultado el 9 de marzo de 2024).

aspectos más eficientes de ambos componentes que desemboquen en una solución de inteligencia aumentada.

Por último, cabe mencionar algunos aspectos de las recomendaciones elaboradas en las «Jornadas sobre regulación y legislación de la inteligencia artificial: IA generativa y tendencias internacionales» celebradas el 5 de junio de 2023 en la Cámara de Diputados de la Nación Argentina por parte del equipo UBA IALAB⁷⁰. Allí, se consideró conveniente el tratamiento de un marco regulatorio general y amplio como una ley de pisos mínimos para evitar anacronismos, que incluya principios éticos y prevea una medición del posible impacto de la IA y, por su particularidad, el abordaje de los riesgos que trae aparejados.

La clave se encontró en diseñar una ley (o cuerpo de leyes) que permita a todos los actores involucrados en el ciclo de vida de la IA tomar decisiones autónomas, conscientes e informadas. En este sentido se consideró que las regulaciones dispersas e implícitas perjudican la visibilidad, por lo cual es recomendable sancionar normas sobre una base de pisos mínimos que de forma explícita sostengan principios, derechos y obligaciones⁷¹.

V. CONCLUSIÓN

Por un lado, en cuanto a la situación regional de los países latinoamericanos, se observa un avance progresivo en la creación de marcos regulatorios. Los planes estratégicos para la IA coexisten con guías éticas que buscan orientar su desarrollo responsable. Para estos documentos, sirven como puntos de referencia las recomendaciones elaboradas por los organismos internacionales como la OCDE, de manera consistente con la tendencia mundial.

A nivel global, a pesar de encontrarse en diferentes puntos de su desarrollo y con diferentes circunstancias económicas y sociales, los países siguen tendencias muy similares en materia de regulación de la IA. La mayor variabilidad se observa en la obligatoriedad o voluntariedad de los distintos marcos y la medida en que el fomento de la innovación se vuelve una prioridad a tener en cuenta para tomar esta decisión.

Esta intuición es coincidente con un estudio llevado a cabo por la consultora Deloitte, en que se analizó una base de datos de más de 1.600 políticas de IA que van desde regulaciones hasta subvenciones de investigación y estrategias nacionales⁷². Allí, en lugar de encontrar conjuntos claros de políticas relacionadas, se descubrió que la mayoría de las políticas se agruparon. Asimismo, se reveló que no solo existe

70. Sobre los consensos y las recomendaciones producto de las Jornadas ver: «Puntos de partida para la regulación de la inteligencia artificial en Argentina» en Corvalán Juan G. (director), «Tratado de Inteligencia Artificial y Derecho» *Thompson Reuters La Ley*, (2023), 2da edición.

71. Sobre los consensos y las recomendaciones producto de las Jornadas ver: «Puntos de partida para la regulación de la inteligencia artificial en Argentina» en Corvalán Juan G. (director), «Tratado de Inteligencia Artificial y Derecho» *Thompson Reuters La Ley*, (2023), 2da edición.

72. El estudio se encuentra disponible en: Mariani Joe, Eggers William D., Kamleshkumar Kishnani Pankaj, «The AI regulations that aren't being talked about» *Deloitte*, disponible en: <https://www2.deloitte.com/xe/en/insights/industry/public-sector/ai-regulations-around-the-world.html> (consultado el 4 de marzo de 2024).

coincidencia en las políticas básicas, sino también en el camino que siguen los países hacia la regulación. Casi todos los países siguen el camino de comprensión de la tecnología, crecimiento y estimulación de la industria y luego dar forma a través de instrumentos regulatorios y estándares.

Este enfoque común sugiere una convergencia en la comprensión global de los desafíos y oportunidades que presenta la IA, así como en la necesidad de un marco regulatorio que fomente la innovación y al mismo tiempo garantice la seguridad y la ética en su aplicación.

**«INTELIGENCIA ARTIFICIAL», ÁMBITO
TERRITORIAL Y ALCANCE DEL REGLAMENTO Y
SU RELACIÓN CON LA PROTECCIÓN DE DATOS**

¿Qué es «inteligencia artificial» para el Reglamento? Análisis, delimitación y aplicaciones prácticas

LORENZO COTINO HUESO

Catedrático de Derecho Constitucional de la Universitat de València. Valgrai

I. LA IMPORTANCIA DEL CONCEPTO DE INTELIGENCIA ARTIFICIAL EN EL REGLAMENTO

El concepto de IA del RIA es clave,¹ entre otros motivos, porque determina esencialmente la aplicación del RIA, que gira en torno a la IA, los sistemas de IA de alto riesgo, determinados sistemas de IA, modelos de IA de uso general o sistemas IA ya introducidos en el mercado. De ahí que delimitar jurídicamente la noción de IA sea la premisa esencial.

Así, cabe recordar que el artículo 1 regula el «objeto» de la norma («promover la adopción de una inteligencia artificial centrada en el ser humano y fiable» y gira en torno a los «sistemas de IA»). De este modo, el RIA establece normas armonizadas, prohibiciones, requisitos específicos para los sistemas de IA de alto riesgo, mecanismos de seguimiento del mercado y vigilancia, así como instrumentos de innovación (art. 1. 2º). El artículo 2 sobre el «ámbito de aplicación» se enfoca en los diversos sujetos de la cadena de valor de los «sistemas de IA» y los de alto riesgo. Además, se hace referencia a los «sistemas o modelos de IA» en investigación (art. 2. 6º y 8º). Así pues, los conceptos de sistema IA y de «alto riesgo» determinan la aplicación de la norma. Esto, obviamente, sin perjuicio de la importancia de los «modelos de IA de uso general» (art. 1. 2º e) y 2. 1º a), Capítulo V) o las obligaciones de transparencia de «determinados sistemas de IA» (art. 50, Capítulo IV).

1. cotino@uv.es. OdiseIA. El presente estudio es resultado de investigación de los siguientes proyectos: MICINN Proyecto «Derechos y garantías públicas frente a las decisiones automatizadas y el sesgo y discriminación algorítmicas» 2023-2025 (PID2022-136439OB-I00) financiado por MCIN/AEI/10.13039/501100011033/; «La regulación de la transformación digital ...» Generalitat Valenciana «Algorithmic law» (Prometeo/2021/009, 2021-24); «Algorithmic Decisions and the Law: Opening the Black Box» (TED2021-131472A-I00) y «Transición digital de las Administraciones públicas e inteligencia artificial» (TED2021-132191B-I00) del Plan de Recuperación, Transformación y Resiliencia. Estancia Generalitat Valenciana CIAEST/2022/1., Grupo de Investigación en Derecho Público y TIC Universidad Católica de Colombia; MICINN; Estancia Generalitat Valenciana CIAEST/2022/1, Convenio de Derechos Digitales-SEDIA Ámbito 5 (2023/C046/00228673) y Ámbito 6. (2023/C046/00229475).

En su caso, cabe hacer referencia a las Disposiciones Finales (Cap. XIII) con relación a «Sistemas de IA ya introducidos en el mercado o puestos en servicio» (art. 111). Ahí se hace referencia a «sistemas de IA que sean componentes de los sistemas informáticos de gran magnitud» en cuanto a la aplicación del RIA. Estos sistemas están determinados en el Anexo X y son relativos al espacio de libertad, seguridad y justicia (Schengen, Visados, Eurodac, Antecedentes Penales, etc.).

II. LAS DIVERSAS DEFINICIONES DE INTELIGENCIA ARTIFICIAL SE HAN BARAJADO

No es una tarea sencilla definir lo que es inteligencia artificial. Desde el siglo XX, se han llegado a señalar más de 55 definiciones,² ello puede tener perspectivas muy diversas como a los efectos de investigación, política e institucional, económica y de mercado. Debe tenerse en cuenta, además, que la «inteligencia artificial» atrae mucha inversión, por lo que muchos llamados sistemas de IA no lo son más que en el nombre y poco o nada tienen que ver con el concepto de IA del RIA, que es el que hay que seguir.

Las dificultades para una definición de IA son mayores cuando se trata de proyectar un régimen jurídico a un sistema de IA. Una definición con fines jurídicos tiene objetivos políticos e institucionales claros y, al mismo tiempo, se requiere la mayor seguridad jurídica posible. En todos los casos, buscar una definición tiene la dificultad de que debe ser adaptativa para los necesarios cambios que la tecnología depara en el futuro.

Así las cosas, la Unión Europea ha seguido una evolución de conceptos por parte de la Comisión Europea en 2018³, por el Alto Grupo de Expertos de la Comisión en 2019⁴, por

2. Así, Samoili, S., y otros, *AI WATCH. Defining Artificial Intelligence*, Publications Office of the European Union, Luxembourg, 2020, doi:10.2760/382730, JRC118163. <https://publications.jrc.ec.europa.eu/repository/handle/JRC118163>
3. Así, en la Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones *sobre Inteligencia Artificial para Europa*, Bruselas, 25.4.2018 COM(2018) 237 final, p. 1, se afirmaba que «La inteligencia artificial (IA) se refiere a los sistemas que muestran un comportamiento inteligente al analizar su entorno y tomar acciones, con cierto grado de autonomía, para lograr objetivos específicos. Los sistemas basados en IA pueden basarse puramente en software, actuando en el mundo virtual (por ejemplo, asistentes de voz, software de análisis de imágenes, motores de búsqueda, sistemas de reconocimiento de voz y rostro) o la IA puede integrarse en dispositivos de hardware (por ejemplo, robots avanzados, automóviles autónomos, drones o aplicaciones de Internet de las Cosas)».
4. Grupo Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial de la Comisión Europea (HLEG), *A definition of AI: main capabilities and disciplines*, Definition developed for the purpose of the AI HLEG's deliverables, Comisión Europea. 8 abril 2019, p. 6. Proponemos utilizar la siguiente definición actualizada de IA: «Los sistemas de inteligencia artificial (IA) son sistemas de software (y posiblemente también de hardware) diseñados por humanos (*) que, dado un objetivo complejo, actúan en la dimensión física o digital al percibir su entorno a través de la adquisición de datos, interpretando los datos recopilados estructurados o no estructurados, razonar sobre el conocimiento, o procesar la información, derivada de estos datos y decidir la(s) mejor(es) acción(es) a tomar para lograr el objetivo dado. Los sistemas de IA pueden usar reglas simbólicas o aprender un modelo numérico, y también pueden adaptar su comportamiento analizando cómo el entorno se ve afectado por sus acciones anteriores. Como disciplina científica, la IA incluye varios enfoques y técnicas, como el

el Parlamento Europeo en 2020⁵. Finalmente, buscando un mayor consenso internacional la UE para el RIA se decantó por el concepto de la OCDE en su Recomendación de los «Principios de IA» en 2019⁶. La definición de la OCDE en 2019 se basó en el concepto de Russell y Norvig de 2009⁷. En este punto hay que destacar que el 8 de noviembre de 2023 el Consejo de la OCDE ha modificado su concepto de IA para alinearse también con las versiones finales del RIA, así como los conceptos de Japón y otros países. El concepto es el siguiente: «Un sistema de IA es un sistema basado en máquinas que, por objetivos explícitos o implícitos, infiere, a partir de la entrada que recibe, cómo generar salidas tales como predicciones, contenidos, recomendaciones o decisiones que pueden influir en entornos físicos o virtuales. Los distintos sistemas de IA varían en sus niveles

aprendizaje automático (de los cuales el aprendizaje profundo y el aprendizaje por refuerzo son ejemplos específicos), el razonamiento automático (que incluye planificación, programación, representación y razonamiento del conocimiento, búsqueda y optimización), y robótica (que incluye control, percepción, sensores y actuadores, así como la integración de todas las demás técnicas en sistemas ciberfísicos)».

5. Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas (2020/2012(INL)). ANEXO B. que establece la *Propuesta de Reglamento del Parlamento Europeo y del Consejo sobre los principios éticos para el desarrollo, el despliegue y el uso de la inteligencia artificial, la robótica y las tecnologías conexas*. En concreto, en su artículo 4. Definiciones: a) «inteligencia artificial», un sistema basado en programas informáticos o incorporado en dispositivos físicos que manifiesta un comportamiento inteligente al ser capaz, entre otras cosas, de recopilar y tratar datos, analizar e interpretar su entorno y pasar a la acción, con cierto grado de autonomía, con el fin de alcanzar objetivos específicos.
6. OECD, *Recommendation of the Council on Artificial Intelligence*, de 22 de mayo de 2019, OECD/LEGAL/0449, adopted by the OECD Council at Ministerial level on 22 May 2019, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449> En dicho documento se «ACUERDA que, a los efectos de la presente Recomendación, los siguientes términos deben entenderse como sigue:
 - Sistema de IA: Un sistema de IA es un sistema basado en una máquina que puede, para un conjunto determinado de objetivos definidos por humanos, hacer predicciones, recomendaciones o decisiones que influyan en entornos reales o virtuales. Los sistemas de IA están diseñados para operar con diferentes niveles de autonomía.
 - Ciclo de vida del sistema de IA: Las fases del ciclo de vida del sistema de IA incluyen: i) “diseño, datos y modelos”; que es una secuencia dependiente del contexto que abarca la planificación y el diseño, la recopilación y el procesamiento de datos, así como la construcción de modelos; ii) “verificación y validación”; iii) “despliegue”; y iv) “operación y seguimiento”. Estas fases a menudo tienen lugar de manera iterativa y no son necesariamente secuenciales. La decisión de retirar un sistema de IA de la operación puede ocurrir en cualquier momento durante la fase de operación y monitoreo.»
7. Ahí se afirmaba que «Un sistema de IA es un sistema basado en una máquina que puede, para un conjunto determinado de objetivos definidos por humanos, hacer predicciones, recomendaciones o decisiones que influyen en entornos reales o virtuales. Los sistemas de IA están diseñados para funcionar con distintos niveles de autonomía». OECD, *Explanatory memorandum on the updated OECD definition of an AI system*, OECD Artificial Intelligence Papers, March 2024 No. 8, Paris, <https://doi.org/10.1787/623da898-en>, p. 4. La referencia es Russell, S. y Norvig P., *Artificial Intelligence: A Modern Approach*, 3rd edition, Pearson, London, 2009 <http://aima.cs.berkeley.edu/>

de autonomía y capacidad de adaptación tras su despliegue.»⁸ Cabe también destacar que la OCDE ha publicado una interesante Exposición de motivos de dicho concepto, aunque no forma parte del texto aprobado.

En Estados Unidos normativamente se maneja un concepto en el Código, 15 U.S.C. 9401(3), reiterado en el apartado Segundo. 3. b, Orden ejecutiva sobre el desarrollo y uso seguro y confiable de la inteligencia artificial, de 30 de octubre de 2023: «un sistema basado en máquinas que puede, para un conjunto dado de objetivos definidos por el ser humano, hacer predicciones, recomendaciones o tomar decisiones que influyan en entornos reales o virtuales. Los sistemas de inteligencia artificial utilizan entradas basadas en máquinas y seres humanos para percibir entornos reales y virtuales; abstraen dichas percepciones en modelos mediante análisis de forma automatizada; y utilizan la inferencia de modelos para formular opciones de información o acción.»⁹ Cabe señalar que hay esfuerzos entre EEUU y la UE para elaborar un mapa de conceptos y taxonomía común, si bien sigue pendiente el concepto de IA.¹⁰

El RIA pretende dotar de una definición única de la IA con suficiente claridad y precisión, que dé seguridad jurídica, que sea funcional y lo más tecnológicamente neutra posible, es decir, que no condicione o favorezca unas tecnologías frente a otras. De igual modo, se pretende una definición que resista al paso del tiempo lo mejor posible dada la dinamicidad tecnológica y del mercado.

8. Original en inglés «An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.» OECD, *Explanatory memorandum... cit.* p. 4.

9. (b) El término «inteligencia artificial» o «IA» tiene el significado establecido en 15 U.S.C. 9401(3): «a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. Artificial intelligence systems use machine — and human-based inputs to perceive real and virtual environments; abstract such perceptions into models through analysis in an automated manner; and use model inference to formulate options for information or action.»

<https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>

[https://uscode.house.gov/view.xhtml?req=\(title:15%20section:9401%20edition:prelim\)#:~:text=\(b\)%20The%20term%20%22artificial,influencing%20real%20or%20virtual%20environments](https://uscode.house.gov/view.xhtml?req=(title:15%20section:9401%20edition:prelim)#:~:text=(b)%20The%20term%20%22artificial,influencing%20real%20or%20virtual%20environments)

10. Así, siguiendo la *Hoja de Ruta de la IA*, Tercera declaración ministerial UE-EE.UU., primera hoja de ruta conjunta sobre herramientas de evaluación y medición para una IA y una gestión de riesgos dignas de confianza (hoja de ruta de la IA). (AI Roadmap) <https://digital-strategy.ec.europa.eu/en/library/ttc-joint-roadmap-trustworthy-ai-and-risk-management>

Ahí, un grupo de expertos se comprometió a preparar un borrador inicial de terminologías y taxonomías de la IA. Se han identificado 65 términos con referencia a documentos clave de la UE y los EE. UU. No obstante, el concepto de inteligencia artificial está en «pendiente». Ver <https://digital-strategy.ec.europa.eu/en/library/eu-us-terminology-and-taxonomy-artificial-intelligence> El documento con los conceptos, EU-U.S. *Terminology and Taxonomy for Artificial Intelligence. First Edition*, mayo 2023, p. 6 <https://ec.europa.eu/newsroom/dae/redirection/document/96104>

III. LA DEFINICIÓN DEL ARTÍCULO 3 REGLAMENTO Y SUS COMPONENTES: TÉCNICAS, AUTONOMÍA, ADAPTACIÓN, ENTRADAS Y SALIDAS Y CONTEXTO

En el listado de 68 definiciones del artículo 3. 1 RIA se inicia con la de «sistema de IA»: «un sistema basado en una máquina que está diseñado para funcionar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación tras el despliegue, y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales». La técnica legislativa empleada para la definición de un sistema IA ha variado a lo largo del proceso legislativo. La propuesta de la Comisión remitía a un Anexo I. No obstante, desde la versión del Consejo UE en diciembre 2022 la definición se contiene íntegra en el listado de definiciones, sin remisión a un anexo. En la evolución cabe destacar que sólo en la última versión se omite la referencia a un «programa informático».

Un elemento central del concepto de IA son las técnicas utilizadas, que como recuerda la OCDE son variadas y en rápido desarrollo¹¹. Bajo «IA» se incluyen categorías de técnicas «como el aprendizaje automático y los enfoques basados en el conocimiento, y áreas de aplicación como la visión por computadora, el procesamiento del lenguaje natural, el reconocimiento de voz y los sistemas inteligentes de apoyo a la toma de decisiones, sistemas robóticos inteligentes, así como la novedosa aplicación de estas herramientas a diversos dominios. Las tecnologías de IA se están desarrollando a un ritmo rápido y es probable que surjan técnicas y aplicaciones adicionales en el futuro».¹²

El Considerando 12 RIA excluye del concepto IA a «los sistemas basados en las normas definidas únicamente por personas físicas para ejecutar automáticamente operaciones». Precisamente para distinguirlo de «los sistemas de software o los planteamientos de programación tradicionales y más sencillos los sistemas de software o los planteamientos de programación tradicionales y más sencillos» se delimitan las técnicas.

Entre las técnicas que definen lo que es un sistema IA destaca el «aprendizaje automático» «que infieren a partir de conocimientos codificados o de una representación simbólica de la tarea que debe resolverse.» (Cons. 12). Como se recuerda desde OCDE, «el aprendizaje automático es un conjunto de técnicas que permite a las máquinas mejorar su rendimiento y generalmente generar modelos de manera automatizada». El proceso de mejorar el rendimiento de un sistema mediante técnicas de aprendizaje automático se conoce como «entrenamiento».

El aprendizaje puede ser:

- supervisado: a partir de anotaciones/etiquetado humano de datos.
- no supervisado: a partir de instancias/puntos de datos que no han sido etiquetados por un humano.
- por refuerzo: se basa en «recompensar» al sistema (mediante ensayo y error), no en conjuntos de datos etiquetados o no etiquetados. De igual modo,

11. OECD, *Explanatory memorandum cit.*, p.1.

12. *Ibidem*, p. 6.

cabe distinguir la variedad de métodos para el aprendizaje automático y el aprendizaje profundo, esto es, basado en redes neuronales (a menudo muy complejas y opacas).

Se recuerda asimismo por la OCDE que «el aprendizaje automático puede continuar adaptándose después de la fase de construcción inicial, mejorando su rendimiento al interactuar directamente con nuevas entradas y datos. Además, los sistemas de IA pueden actualizarse/reentrenarse, volverse a probar y volverse a implementar periódicamente como nuevas versiones».¹³

Así, por ejemplo, se puede considerar que es aprendizaje profundo aquel que utiliza redes neuronales artificiales con múltiples capas, como los sistemas reconocimiento de voz y de imágenes, como el uso de redes neuronales convolucionales para identificar objetos en fotografías. En el caso del aprendizaje por refuerzo, en el que el sistema aprende a tomar decisiones mediante ensayo y error, pueden situarse los sistemas de IA en juegos de estrategia como el ajedrez o *Go*, donde el sistema mejora su juego a través de partidas repetidas y ajustando su estrategia en función de las recompensas obtenidas.

Además del aprendizaje, no deben excluirse las técnicas de razonamiento o modelado, a las que se hacía expresa referencia en las versiones iniciales del RIA.¹⁴ La modelización predictiva está estrechamente relacionada con el aprendizaje automático. Se incluyen técnicas estadísticas de aprendizaje e inferencia (incluida la estimación bayesiana) y métodos de búsqueda y optimización.¹⁵ Como ejemplo, cabe mencionar los motores de IA de ajedrez (búsqueda y optimización), que generan un árbol de búsqueda mostrando algunas de las posibles jugadas de las blancas.¹⁶

También se señala como ejemplo de sistemas de IA simbólicos o basados en conocimientos un sistema que razona sobre procesos de fabricación, que podría tener variables que representen fábricas, bienes, trabajadores, vehículos, máquinas, etc.¹⁷

Igualmente, son sistemas de IA los basados en el razonamiento que pueden razonar (usando operaciones como ordenar, buscar, emparejar y encadenar) en la base del conocimiento codificado. Estos métodos son más interpretables que los sistemas de aprendizaje, pero también pueden presentar sesgo, complejidad, imprevisibilidad y autonomía.¹⁸

Entre las distintas técnicas podemos poner ejemplos de sistemas de planificación y programación que crean planes de acción para alcanzar objetivos específicos y la

13. *Ibidem*, p. 8.

14. Siguiendo a la OCDE un modelo se define como una «representación física, matemática o lógica de otro modo de un sistema, entidad, fenómeno, proceso o datos» en la norma ISO/IEC 22989. Y se indica que se incluirían entre otros, modelos estadísticos y diversos tipos de funciones de entrada y salida (como árboles de decisión y redes neuronales). Los modelos de IA pueden ser contruidos manualmente por programadores humanos o automáticamente mediante, por ejemplo, métodos no supervisados, supervisados o Técnicas de aprendizaje automático de refuerzo.

15. En versiones anteriores eran técnicas de modelado, letra c de Anexo 1, ahora hay expresa referencia en Considerandos.

16. Ejemplo de modelado del JRC en Samoili, S., y otros., *AI WATCH. Defining Artificial Intelligence...* cit.

17. *Ibidem*, p. 8.

18. Se sigue de *Ibidem*.

programación de tareas para ser ejecutadas por una máquina, como la planificación de rutas para vehículos autónomos. Respecto de las técnicas de representación y razonamiento, se pueden citar los sistemas expertos que diagnostican enfermedades basándose en los síntomas y el historial médico del paciente.

Otro elemento sustancial del concepto de IA es que el sistema cuente con un mínimo grado de «autonomía», en concreto «distintos niveles de autonomía». Esto implica «cierto grado de independencia» [...] ciertas capacidades para funcionar sin intervención humana (Cons. 12). La OCDE define la autonomía como el «grado en que un sistema puede aprender o actuar sin la participación humana luego de la delegación de autonomía y automatización de procesos por parte de los humanos».¹⁹ Existen distintos niveles de autonomía, como los seis niveles estándar generados en 2016 para los vehículos autónomos:²⁰ Nivel 0 (sin automatización de conducción), Nivel 1 (asistencia al conductor), Nivel 2 (automatización de conducción parcial), Nivel 3 (automatización de conducción condicional), Nivel 4 (alta automatización de conducción) y Nivel 5 (automatización de conducción completa).

De modo paralelo, se subraya la «capacidad de adaptación», esto es, «capacidades de autoaprendizaje que permiten al sistema cambiar mientras está en uso» (Cons. 12). Desde la OCDE se apunta que los sistemas pueden seguir evolucionando y modificando su comportamiento.²¹ Así, un sistema se puede entrenar y desarrollar nuevas formas de inferencia no imaginadas por los desarrolladores. Un sistema con alta capacidad de adaptación puede cambiar su funcionamiento en respuesta a cambios en su entorno, lo que le permite mantenerse eficaz y relevante en situaciones dinámicas. Por ejemplo, los asistentes como *Siri* o *Alexa*, en sus versiones iniciales, respondían a comandos predefinidos, pero no aprendían de las interacciones, mientras que los asistentes virtuales avanzados utilizan aprendizaje profundo para personalizar respuestas y recomendaciones basadas en el historial de interacción del usuario. En el caso de los sistemas de diagnóstico médico basados en IA, hay herramientas de apoyo a la decisión clínica basadas en reglas predefinidas que no serían consideradas IA, mientras que los sistemas de diagnóstico basados en IA que utilizan grandes bases de datos y algoritmos de aprendizaje automático pueden diagnosticar enfermedades de forma autónoma y mejorar sus recomendaciones con el tiempo.

También es un elemento esencial definitorio la capacidad de inferencia de un sistema de IA. Por un lado, un sistema genera una salida a partir de sus entradas, generalmente después de la implementación. Se trata de la «obtención de información de salida, como predicciones, contenidos, recomendaciones o decisiones, que puede influir en entornos físicos y virtuales, y a la capacidad de los sistemas de IA para deducir modelos o algoritmos a partir de información de entrada o datos» (Cons. 12). En cuanto a los tipos de salida, se señala que las categorías corresponden a diferentes niveles de participación humana, siendo las «decisiones» el tipo de resultado más autónomo (el sistema de IA afecta su entorno directamente o dirige a otra entidad

19. OECD, *Explanatory memorandum cit.*, p. 6.

20. <https://www.sae.org/news/2019/01/sae-updates-j3016-automated-driving-graphic> y https://www.sae.org/standards/content/j3016_202104/

21. OECD, *Explanatory memorandum cit.*, p. 6.

para que lo haga) y las «predicciones» menos autónomo.²² Por ejemplo, un sistema de asistencia al conductor puede «predecir» que una región de píxeles en la entrada de su cámara es un peatón; podría «recomendar» frenar o podría «decidir» aplicar el freno. Por su parte, los sistemas de IA generativa que producen «contenidos» (incluidos texto, imágenes, audio y vídeo).

Por otro lado, especialmente durante la fase de construcción, la capacidad de inferencia «permite el aprendizaje, el razonamiento o la modelización» (Cons. 12). Así, las salidas del sistema IA se utilizan para evaluar una versión de un modelo y derivar un modelo a partir de entradas/datos.²³

Por cuanto a la entrada que se facilita al sistema, puede incluir datos relevantes para la tarea a realizar o tomar la forma de un mensaje de usuario o una consulta de búsqueda. Los sistemas de IA pueden tener uno o más tipos de objetivos y «pueden funcionar con arreglo a objetivos definidos explícitos o a objetivos implícitos» (Cons. 12). Los objetivos explícitos están definidos por humanos. Los objetivos implícitos pueden estar en reglas (normalmente especificadas por humanos) o implícitos en los datos de entrenamiento. En estos casos, los objetivos no se conocen completamente de antemano. Asimismo, las indicaciones del usuario pueden complementar los objetivos.²⁴

Finalmente, cabe señalar que los sistemas IA funcionan en máquinas («basado en máquinas») y en ambientes o contextos físicos o virtuales, e incluyen entornos que describen aspectos de la actividad humana.²⁵ Asimismo, «pueden utilizarse de manera independiente o como componentes de un producto, con independencia de si el sistema forma parte físicamente del producto (integrado) o contribuye a la funcionalidad del producto sin formar parte de él (no integrado)» (Cons. 12).

IV. EJEMPLOS DE SISTEMAS QUE SÍ QUE SON O NO INTELIGENCIA ARTIFICIAL

Hay que partir de que no todo sistema informático que permite procesos o decisiones automatizados es *automáticamente* IA. No se considerarán «inteligencia artificial» aunque puedan razonar o modelar matemáticamente. Ello es importante; si el sistema de decisiones automatizadas no es un Sistema de IA, aunque se utilice para un caso de uso y finalidades de alto riesgo (anexo III), quedará también fuera de la aplicación del RIA.

En esta dirección el JRC sitúa algunos ejemplos.²⁶ Un sistema de puntuación de créditos que tiene por objetivo estimar el riesgo asociado a la concesión de un préstamo. Este sistema utiliza datos sobre características del prestatario, situación económica (ingresos, gastos mensuales), importe del préstamo, finalidad, datos demográficos. El resultado de este sistema es una categoría de riesgo, por ejemplo, clientes fiables, clientes que pueden tener problemas para pagar. En este caso,

22. *Ibidem*, p. 8.

23. *Ídem*.

24. *Ibidem*, p. 7.

25. *Ídem*.

26. Se sigue del JRC en Samoili, S., y otros., *AI WATCH. Defining Artificial Intelligence...* cit. aunque tales afirmaciones pueden ser cuestionables.

pueden darse los requisitos de arquitectura de aprendizaje profundo entrenada en información histórica, con técnicas de aprendizaje automático. Puede darse un razonamiento basado en un histórico de decisiones humanas con técnicas de estrategias basadas en la lógica y el conocimiento. Puede haber una regresión logística sobre datos históricos con estrategias estadísticas. Sin embargo, no habría que considerarlo como IA si el sistema se basa en un conjunto fijo de reglas, definidas manualmente por un humano.

Otro ejemplo es un algoritmo para calificar a los estudiantes de A Level y GCSE en Inglaterra, Gales e Irlanda del Norte basándose en información histórica (calificaciones anteriores).²⁷ Así, sí que se considerará que es un sistema de IA si los criterios relevantes para determinar el resultado (las calificaciones de los alumnos) son elegidos por los humanos: las estimaciones de los profesores sobre las calificaciones, el rendimiento de la escuela en años anteriores, el rendimiento de la cohorte en años anteriores. En este caso, en cuanto al enfoque de los resultados, el modelo estadístico combina los datos históricos con las estimaciones de los profesores. Las estimaciones de los profesores se ajustan de acuerdo con el modelo estadístico para adaptarse a una distribución de calificaciones anteriores.

Por el contrario, y en este mismo contexto, no habría que considerar como un sistema de IA, por no contar con los requisitos del artículo 3, un algoritmo para decidir la inscripción en la escuela basado en variables del estudiante como el nivel de educación de la madre, la situación económica, la distancia del hogar a la escuela o la preferencia del estudiante. No sería un sistema de IA si son los humanos quienes seleccionan los criterios que son relevantes para la decisión y también deciden a qué criterios dar más importancia y qué pesos asignar a los criterios de categorización.

Para concluir, este análisis del concepto de IA en el RIA subraya la importancia que tiene una definición clara y precisa de la IA para garantizar la seguridad jurídica y funcionalidad tecnológica. Se ha destacado la adopción de la definición de la OCDE como base para el RIA, en la búsqueda de un enfoque estandarizado y armonizado en el ámbito internacional. La definición del RIA incluye diversas técnicas que constituyen la IA, como el aprendizaje automático, el procesamiento del lenguaje natural y los sistemas de apoyo a la toma de decisiones. Se ha intentado dar una visión práctica de las mismas. En cualquier caso, se ha subrayado la autonomía y capacidad de adaptación de un sistema IA. En cualquier caso y como se explicará en el apartado relativo a los sistemas de alto riesgo, el RIA prevé la evolución y su adaptación a los rápidos avances tecnológicos.

27. <https://bera-journals.onlinelibrary.wiley.com/doi/full/10.1002/berj.3705>

El ámbito de aplicación territorial del Reglamento de inteligencia artificial

ALFONSO ORTEGA GIMÉNEZ

Profesor Titular de Derecho internacional privado de la Universidad Miguel Hernández de Elche (Alicante)¹

I. PLANTEAMIENTO

El RIA supone la primera regulación jurídica de la Inteligencia Artificial de carácter global, directamente aplicable en todos los Estados miembros de la Unión Europea (en lo sucesivo, UE). Una norma de prevención dirigida a los fabricantes/ desarrolladores de sistemas de IA para que los mismos no impacten en los derechos fundamentales de las personas. Al mismo tiempo, aspira a tener una eficacia universal, como ya ha sucedido con el Reglamento General de Protección de Datos², es decir, con repercusión más allá de las fronteras de la UE. Se aplicará a sistemas de IA que funcionen como componentes de productos o que sean productos en sí mismos, que se pretendan comercializar o poner en servicio en el mercado de la UE (artículo 2.1 del RIA).

Esta nueva norma persigue desarrollar un ecosistema de confianza mediante el establecimiento de un marco jurídico destinado a lograr que la IA sea fiable y respete el Derecho. Se basa en los valores y derechos fundamentales de la UE que tienen por objeto esencial inspirar confianza a los ciudadanos y otros usuarios para que adopten soluciones basadas en la IA; al tiempo que se trata de animar a las empresas a que desarrollen e inviertan en este tipo de soluciones. La IA debe ser un instrumento para las personas y una fuerza positiva en la sociedad, y su fin último debe ser incrementar el bienestar humano, respetando los derechos de las personas.

La técnica utilizada para la regulación de esta materia, que está inspirada por el RGPD, se caracteriza por cuatro elementos³: a) La utilización de un Reglamento

1. 0000-0002-8313-2070.
2. Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos). DOUE n.º 119, de 4 de mayo de 2016.
3. Ver Gascón Macén, A., «El Reglamento General de Protección de Datos como modelo de las recientes propuestas de legislación digital europea», *CDT*, Vol. 13(2) (2021), pp. 209-232, disponible en: <https://doi.org/10.20318/cdt.2021.6256>; Papakonstantinou,

en lugar de una Directiva como técnica jurídica⁴; b) El establecimiento de rígidos requisitos y obligaciones para distintas categorías de posiciones para el acceso a la actividad y la prestación de cualquier servicio digital; c) El nombramiento por parte de los Estados Miembros de autoridades nacionales competentes para que las empresas encuentren una vía más directa cuando deseen reclamar por el incumplimiento del Reglamento; y d) El establecimiento de órganos colegiados a nivel europeo, aunque con diferentes papeles en función de cada Reglamento⁵.

Se configura el RIA como un instrumento jurídico que busca armonizar las normas en este campo y establecer un marco regulatorio confiable, no limitado a sectores concretos; y con la finalidad de ofrecer respuestas sujetas, entre otros, al principio de proporcionalidad en función de los riesgos que ocasione la IA. La IA está diseñada para ser utilizada en cualquier sector de la actividad, dando lugar a que las normas reguladoras de los distintos sectores se apliquen en relación con el diseño y desarrollo de IA; por ejemplo, siendo de aplicación la legislación en materia de protección de datos de carácter personal, la normativa sobre secretos empresariales, o la legislación sobre protección de los consumidores y prácticas comerciales desleales, entre otros⁶.

El RIA no sólo está diseñado para fomentar la adopción de sistemas de IA en el mercado interior, sino que también tiene la ambición de posicionar a la UE como un líder mundial en el desarrollo de una IA confiable y ética. Este marco legislativo responde a la necesidad de ofrecer, a nivel global, un alto nivel de protección de los intereses públicos, como la salud y la seguridad, mientras se asegura el respeto de los derechos fundamentales.

El artículo 2.1 del RIA⁷ se convierte en uno de los artículos fundamentales, ya que delinea el ámbito de aplicación de la ley, especificando quiénes estarán sujetos a las

V. y DE HERT, P., «Post GDPR EU laws and their GDPR mimesis. DGA, DSA, DMA and the EU regulation of AI», *European Law Blog*, (2021), disponible en: <https://europeanlawblog.eu/2021/04/01/post-gdpr-eu-laws-and-their-gdpr-mimesis-dga-dsa-dma-and-the-eu-regulation-of-ai/>

4. A pesar de que en las propuestas legislativas se les denomine «leyes». *Vid.*, sobre el particular, Papakonstantinou, V. y De Hert, P., «EU lawmaking in the Artificial Intelligent Age: Actification, GDPR mimesis, and regulatory brutality», *European Law Blog*, (2021), disponible en: <https://europeanlawblog.eu/2021/07/08/eu-lawmaking-in-the-artificial-intelligent-age-act-ification-gdpr-mimesis-and-regulatory-brutality/>
5. Comité Europeo de Inteligencia Artificial (artículo 56 RIA), si bien el Parlamento propone cambiar su nombre al de European Artificial Intelligence Office (AI Office) e incrementar considerablemente sus funciones. Otros organismos colegiados previstos en las leyes digitales son: Comité Europeo de Innovación en materia de Datos (artículo 29 RGPD), Junta Europea de Servicios Digitales (artículo 61 RSD), Grupo de Alto Nivel (artículo 40 RMD), que se unen al ya existente Comité Europeo de protección de datos (artículo 68 RGPD).
6. Ver Miguel Asensio, P., *Manual de Derecho de las Nuevas Tecnologías. Derecho Digital*, Aranzadi, Cizur Menor, Navarra 2023, pp. 121.
7. El artículo 2.1 del RIA señala: «1. El presente Reglamento se aplicará a: a) los proveedores que introduzcan en el mercado o pongan en servicio sistemas de IA o que introduzcan en el mercado modelos de IA de uso general en la Unión, con independencia de si dichos proveedores están establecidos o ubicados en la Unión o en un tercer país; b) los responsables del despliegue de sistemas de IA que estén establecidos o ubicados en la Unión; c) los proveedores y responsables del despliegue de sistemas de IA que estén

nuevas regulaciones; y, por tanto, quiénes deben acatar las obligaciones contenidas en el Reglamento. Los proveedores y usuarios de sistemas de IA, ya sea dentro de la UE o en terceros países, se verán afectados por este Reglamento cuando la información de salida del sistema de IA se utilice en la UE. Esta disposición garantiza que la regulación tenga un alcance transfronterizo, abarcando no solo a los actores dentro de la UE sino también aquellos cuyos sistemas de IA puedan afectar a los ciudadanos de la UE. La extraterritorialidad de la norma debe ser regulada y analizada con detenimiento debido a las múltiples implicaciones que trae consigo, siendo una de las mayores novedades de esta propuesta.

Uno de los aspectos más destacados del RIA es su enfoque tecnológicamente neutral y su intento de ser resistente al tiempo, teniendo en cuenta la rápida evolución de la tecnología y el mercado de la IA. Esto es fundamental para una regulación duradera y adaptable que pueda mantenerse al día con los avances tecnológicos sin necesidad de cambios frecuentes.

El RIA también proporciona una definición clara de los principales actores en la cadena de valor de la IA, tales como proveedores⁸, implementadores⁹, representantes autorizados¹⁰, importadores¹¹ y distribuidores de sistemas de IA¹², así como los fabricantes de productos que comercialicen o pongan en servicio un sistema de IA junto con su producto y bajo su propio nombre o marca comercial. Este enfoque detallado es esencial para clarificar las responsabilidades y garantizar una igualdad de condiciones en toda la industria. Por otro lado, los sistemas de IA se encuentran clasificados en función de su capacidad para dañar y poner en peligro la seguridad y los derechos fundamentales de las personas.

establecidos o ubicados en un tercer país, cuando la información de salida generada por el sistema de IA se utilice en la Unión. d) los importadores y distribuidores de sistemas de IA; e) los fabricantes de productos que introduzcan en el mercado o pongan en servicio un sistema de IA junto con su producto y con su propio nombre o marca comercial; f) los representantes autorizados de los proveedores que no estén establecidos en la Unión; g) las personas afectadas que estén ubicadas en la Unión».

8. «Proveedor»: una persona física o jurídica o autoridad, órgano u organismo de otra índole públicos que desarrolle un sistema de IA o un modelo de IA de uso general o para el que se desarrolle un sistema de IA o un modelo de IA de uso general y lo introduzca en el mercado o ponga en servicio el sistema de IA con su propio nombre o marca comercial, previo pago o gratuitamente.
9. «Implementador» o «responsable del despliegue»: una persona física o jurídica o autoridad, órgano u organismo de otra índole públicos que utilice un sistema de IA bajo su propia autoridad, salvo cuando su uso se enmarque en una actividad personal de carácter no profesional.
10. «Representante autorizado»: una persona física o jurídica ubicada o establecida en la Unión que haya recibido y aceptado el mandato por escrito de un proveedor de un sistema de IA o de un modelo de IA de uso general para cumplir las obligaciones y llevar a cabo los procedimientos establecidos en el presente Reglamento en representación de dicho proveedor.
11. «Importador»: una persona física o jurídica ubicada o establecida en la Unión que introduzca en el mercado un sistema de IA que lleve el nombre o la marca comercial de una persona física o jurídica establecida en un tercer país.
12. «Distribuidor»: una persona física o jurídica que forme parte de la cadena de suministro, distinta del proveedor o el importador, que comercialice un sistema de IA en el mercado de la Unión.

Sin duda alguna, e RIA representa un esfuerzo ambicioso para establecer un equilibrio entre la promoción de la innovación tecnológica y la protección de los ciudadanos y sus derechos. Lo que permanece claro es que la UE busca una posición de liderazgo en el establecimiento de estándares globales para la gobernanza de la IA, destacando su compromiso con una IA que sea segura, ética y bajo el control humano.

En el Título I del RIA se define el ámbito de aplicación de las nuevas normas que abarcan la introducción en el mercado mediante la comercialización, la puesta en servicio y la utilización de sistemas de IA.

El artículo 2.1 del RIA puede ser considerada desde el punto de vista del Derecho internacional privado como una norma de derecho aplicable, y concretamente, una norma de conflicto unilateral cuyo objetivo es determinar a qué situaciones de la UE es aplicable el RIA. Estas situaciones serán diferentes en función de si se analizan desde la posición de los operadores económicos, de las autoridades nacionales competentes o de los tribunales de justicia¹³.

II. APLICACIÓN DEL ARTÍCULO 2.1 POR LOS OPERADORES ECONÓMICOS

La primera perspectiva de análisis es la de los operadores económicos, es decir, la de los «proveedores» que comercialicen o pongan en servicio sistemas de IA o comercialicen modelos de IA de propósito general en la UE, «implantadores o responsables del despliegue de sistemas de IA», «representantes autorizados», «importadores y distribuidores de sistemas de IA» y «personas afectadas que estén establecidas en la UE». Para estos operadores resulta fundamental, antes de llevar a cabo su actividad económica, saber si el Reglamento les resultará aplicable. En principio, la respuesta podría parecer sencilla: la futura normativa será de aplicación a sistemas IA que funcionan como componentes de productos o que son productos en sí mismos que se pretenden comercializar o poner en servicio en el mercado UE. Por consiguiente, cualquier sistema IA desarrollado por proveedores o utilizados por usuarios establecidos en terceros Estados resultan accesibles por potenciales clientes ubicados en Europa. ¿Resulta esto suficiente para que el RIA resulte aplicable? A tenor del artículo 2.1 del RIA, la respuesta debe ser negativa. Esta disposición establece unos criterios de conexión que, en principio, implican que únicamente se va a aplicar a proveedores y usuarios que llevan a cabo actividades que presentan una vinculación estrecha con la UE¹⁴.

El artículo 2.1 del RIA desempeña un papel crucial para determinar los operadores económicos que estarán sujetos a las obligaciones del Reglamento y cómo deben interpretar su aplicación prospectiva. La discusión sobre la aplicación prospectiva es esencial para los operadores económicos, ya que les proporciona la claridad necesaria para planificar y adaptar sus estrategias de negocio en concordancia con los nuevos requisitos regulatorios.

13. Ver López-Tarruella Martínez, A., «El reglamento de Inteligencia Artificial y las relaciones con terceros estados», *Revista Electrónica de Estudios Internacionales (REEI)*, n.º 45 (2023), pp. 5-11.

14. *Vid.* en sentido amplio, *ibidem*.

En el contexto del RIA, los operadores económicos incluyen a los proveedores, a los implementadores de sistemas de IA, a los usuarios de sistemas de IA, a los representantes autorizados, a los importadores y distribuidores de sistemas de IA que operan en el mercado único y a las personas afectadas que estén establecidas en la UE. El artículo 2.1 especifica que el Reglamento se aplicará a los operadores económicos que pongan en el mercado de la UE o pongan en servicio sistemas de IA. Este enfoque prospectivo significa que los operadores económicos deben, en primer lugar, conocer si el Reglamento les será aplicable y, en caso afirmativo, anticipar cómo las disposiciones reglamentarias impactarán en sus productos de IA aún no comercializados, así como en los servicios que planean ofrecer en el futuro.

Para conocer si incide en su ámbito de aplicación, la disposición establece unos criterios de conexión que van a implicar que el Reglamento se aplique a sistemas de IA desarrollados por proveedores o utilizados por usuarios establecidos en terceros estados. El criterio de conexión será el de la vinculación estrecha con la UE.

Estos criterios de conexión plantean consecuencias: la primera de ellas, inseguridad jurídica para los operadores que encontrarán dificultades para determinar si están sujetos a ciertas obligaciones; y, la segunda, que el RIA se le aplique injustificadamente por no presentar suficientes vínculos con la UE¹⁵ (por ejemplo, la aplicación de la legislación europea a empresas establecidas en terceros estados en aquellos supuestos que presentan una mínima vinculación con la UE).

Para lograr una comprensión integral de la aplicación prospectiva del artículo 2.1 del RIA, es esencial analizarlo desde varias perspectivas: la del cumplimiento proactivo; la de la evaluación de impacto de la IA y el diseño por privacidad; la del enfoque basado en riesgos; y la de los principios de ética en IA.

Veamos cada una de esas perspectivas:

A) Cumplimiento proactivo.

El cumplimiento proactivo en el contexto del artículo 2.1 del RIA refleja la necesidad de que los operadores económicos aborden las cuestiones reglamentarias desde las fases tempranas del diseño y desarrollo de los sistemas de IA. Esta aproximación no sólo garantiza la conformidad con las regulaciones emergentes, sino que también se alinea con los principios éticos y las expectativas de la sociedad respecto a la tecnología responsable.

Un componente esencial de un enfoque de cumplimiento proactivo es la capacitación interna. Los operadores económicos deben invertir en programas de formación para asegurar que su personal esté al tanto de los requisitos del Reglamento y entienda cómo aplicar las prácticas de desarrollo de sistemas de IA conforme al RIA. Esto es especialmente crítico para aquellos involucrados en el diseño y la implementación de los sistemas de IA, así como para el personal encargado del cumplimiento normativo y la gestión de riesgos.

La fase de prueba de los sistemas de IA es un momento crucial para el cumplimiento proactivo. Los operadores económicos deberán implementar procedimientos de prueba rigurosos que no sólo evalúen la funcionalidad técnica sino también la conformidad con los estándares éticos y legales. Esto puede requerir

15. Ver *ibídem*, p. 6.

la colaboración con partes interesadas externas, como organismos de certificación o grupos de consumidores, para validar la eficacia y seguridad de los sistemas de IA.

Los operadores económicos deben evaluar cómo sus sistemas podrían ser mal utilizados o cómo podrían fallar y las consecuencias que estos tendría para los usuarios y la sociedad en general.

B) Evaluaciones de impacto de la Inteligencia Artificial y diseño por privacidad.

Una estrategia de cumplimiento proactivo implica que las evaluaciones de impacto de la IA deben convertirse en una parte integral del ciclo de vida del desarrollo de los productos de IA. Estas evaluaciones deben ir más allá de las consideraciones técnicas y abordar también las implicaciones sociales, éticas y legales de los sistemas de IA. El enfoque de la «privacidad por diseño» exige que los operadores económicos integren medidas de protección de datos personales desde la etapa de conceptualización del producto, asegurando que la privacidad del usuario no sea una consideración secundaria sino un pilar central de los sistemas de IA.

C) Enfoque basado en riesgos.

El RIA categoriza los sistemas de IA según el nivel de riesgo que presentan. Un enfoque proactivo requiere que los operadores económicos identifiquen el nivel de riesgo de sus sistemas de IA desde el inicio; es decir, con anterioridad a su introducción en el mercado, puesta en servicio y/o uso, adaptando sus procesos de desarrollo para cumplir con los estándares necesarios. Por ejemplo, un sistema de IA de alto riesgo requerirá del cumplimiento de los siguientes requisitos: 1) establecimiento de un sistema de gestión de riesgos; 2) calidad de los conjuntos de datos empleados; 3) documentación y registro; 4) transparencia y divulgación de información a los usuarios; 5) supervisión humana; y 6) la solidez, precisión y ciberseguridad.

Una herramienta clave en este proceso es la evaluación de impacto en la IA, que examina las consecuencias potenciales antes de que los sistemas sean desplegados. Estas evaluaciones deben considerar no sólo los casos de uso previstos, sino también escenarios hipotéticos en los que el sistema podría ser empleado de manera no intencionada. Al anticipar estos escenarios, los operadores pueden diseñar sistemas más resilientes.

Una vez identificados los riesgos, es esencial desarrollar estrategias de mitigación. Esto puede incluir la implementación de salvaguardias técnicas (como sistemas de supervisión y alertas tempranas), así como la creación de políticas y procedimientos que limiten el uso de la IA a aplicaciones seguras y éticas. La formación continua y la actualización de habilidades del personal que maneja sistemas de IA también son cruciales para una efectiva mitigación de riesgos.

D) Principios de ética en Inteligencia Artificial.

Los operadores económicos también deben asegurarse de que los sistemas de IA estén en consonancia con los principios éticos reconocidos, como la transparencia, la equidad y la no discriminación. Esto implica un compromiso con la creación de sistemas de IA que sean comprensibles para los usuarios y cuyas decisiones puedan ser justificadas. Además, debe haber salvaguardias para prevenir sesgos y discriminación, lo cual requiere una constante revisión y actualización de los modelos de IA para reflejar los valores sociales.

El diseño ético y responsable de la IA debe considerar el impacto potencial en individuos y en la sociedad. Los sistemas de IA deben ser diseñados de manera que respeten los derechos humanos, la dignidad y los valores democráticos. Esto se refiere a evitar sesgos y discriminación y garantizar que las decisiones automatizadas sean justas y no discriminatorias.

III. APLICACIÓN DEL ARTÍCULO 2.1 POR LAS AUTORIDADES NACIONALES COMPETENTES

La aplicación del artículo 2.1 del RIA por las autoridades nacionales competentes es cuanto menos «curiosa»; pues contrariamente a lo establecido en el RGDP, el RIA no establece un derecho para las personas físicas o jurídicas de presentar una reclamación ante las autoridades nacionales de supervisión por el incumplimiento de las disposiciones del Reglamento por parte de proveedores, usuarios o cualquier otro operador de sistemas IA. Además, su aplicación lleva consigo la necesidad de establecer un marco de competencia internacional entre dichas autoridades. Este artículo 2.1. del RIA establece el ámbito de aplicación material de la normativa, definiendo lo que se entiende por sistemas de IA y estableciendo las bases para su regulación, supervisión y control. Hay una serie de factores imprescindibles para que la aplicación del artículo 2.1 resulte efectiva: la necesidad de cooperación internacional; la armonización de estándares regulatorios; y la interacción con el Derecho internacional.

Analicemos cada uno de esos factores:

1. NECESIDAD DE COOPERACIÓN INTERNACIONAL

La cooperación internacional en la supervisión de la IA es un componente crucial en la aplicación del artículo 2.1 del RIA por parte de las autoridades nacionales competentes. Dado que la IA no conoce fronteras y puede tener impactos significativos en múltiples jurisdicciones, la necesidad de un enfoque coordinado es imperativa. Las autoridades nacionales competentes deben, por tanto, construir puentes de colaboración y compartir información y recursos para garantizar una supervisión efectiva.

Uno de los mayores desafíos para la cooperación internacional es la diversidad de marcos regulatorios. Cada país puede tener su propio enfoque para la regulación de la IA, basado en sus valores culturales, normas sociales y prioridades políticas. Esto puede llevar a discrepancias en la interpretación y aplicación de las regulaciones de la IA. Por tanto, las autoridades deben trabajar hacia la armonización de estos enfoques para permitir un ecosistema regulador más homogéneo¹⁶.

La cooperación internacional no se detiene en la regulación formal; también implica la capacitación y el conocimiento compartido. Las autoridades nacionales pueden beneficiarse enormemente de programas de capacitación conjuntos, intercambios de personal que promuevan un entendimiento común de los desafíos y las mejores prácticas en la supervisión de la IA.

16. Ver Corcoy, M., «La inteligencia artificial en el derecho español», *Revista de Derecho y Genoma Humano*, n.º 54, (2021).

Mirando hacia el futuro, se espera que la cooperación internacional en la regulación de la IA se fortalezca aún más. Con la rápida evolución de la tecnología, las autoridades nacionales deberán permanecer proactivas en su colaboración.

La aplicación del artículo 2.1 del RIA por las autoridades nacionales competentes dentro del contexto internacional requiere un esfuerzo continuo y concertado. Al trabajar en conjunto, las autoridades pueden garantizar que la regulación de la IA sea eficaz, justa y no discriminatoria, protegiendo a los ciudadanos y fomentando la innovación responsable a nivel global.

2. ARMONIZACIÓN DE ESTÁNDARES REGULATORIOS

La armonización de estándares regulatorios para la IA a nivel internacional es un proceso complejo, pero fundamental para asegurar que la tecnología se desarrolle y se aplique de manera segura y ética en diversos contextos socioeconómicos. La aplicación del artículo 2.1 del RIA por las autoridades nacionales competentes se ve directamente influenciada por el grado de coherencia que puedan alcanzar los estándares regulatorios a nivel global.

Los estándares regulatorios de la IA abarcan una amplia gama de consideraciones, desde la seguridad y la privacidad hasta la equidad y la transparencia. La existencia de numerosos enfoques regulatorios refleja la diversidad de valores y objetivos de las sociedades a lo largo del mundo. Sin embargo, esta diversidad también puede resultar en un paisaje fragmentado que dificulta la cooperación internacional y el comercio¹⁷.

Un componente clave de la armonización es el desarrollo de esquemas de certificación y pruebas estandarizados. Estos esquemas permitirán a las autoridades evaluar y certificar sistemas de IA de acuerdo con criterios acordados internacionalmente. Facilitan así la confianza mutua y el reconocimiento de la conformidad de productos y servicios de IA.

Para lograr una armonización efectiva es crucial que los diferentes marcos normativos sean interoperables. Esto significa que las regulaciones de una jurisdicción no deben entrar en conflicto con las de otra y que los operadores económicos deben poder navegar fácilmente entre diferentes sistemas regulatorios sin tener que cumplir con requisitos contradictorios. Una de las vías para lograrlo, como ya se comentará con posterioridad es mediante el uso de acuerdos multilaterales y bilaterales entre terceros estados y la UE.

La mayor preocupación en la armonización de estándares regulatorios es la protección de los derechos fundamentales. La legislación europea pone un énfasis particular en la protección de datos y la privacidad, y cualquier esfuerzo de armonización debe garantizar que estos derechos no se vean comprometidos.

En la práctica, la armonización de estándares regulatorios puede implicar la creación de grupos de trabajo internacionales, la redacción de documentos de consenso y la realización de estudios comparativos.

17. Ver International Standards Organization (ISO), *ISO Standards for Artificial Intelligence* (2022).

La armonización de los estándares regulatorios es esencial para crear un entorno global seguro y equitativo para el desarrollo y uso de la IA. A través del esfuerzo colectivo de las autoridades nacionales competentes, la cooperación internacional, y la participación activa de entidades de todos los sectores, se puede lograr un marco regulatorio que no sólo proteja los derechos individuales, sino que también promueva la innovación y el crecimiento económico.

3. INTERACCIÓN CON EL DERECHO INTERNACIONAL

La interacción del marco regulatorio de la IA con el Derecho internacional es un campo que está evolucionando rápidamente, con múltiples implicaciones para el comercio, la diplomacia, y la gobernanza global. La implementación del artículo 2.1 del RIA por las autoridades nacionales debe considerar cómo las normativas locales se alinean, complementan o, en algunos casos, pueden entrar en conflicto con las obligaciones internacionales existentes.

Una de las primeras consideraciones es cómo las regulaciones de IA se insertan en el tejido de tratados internacionales previamente establecidos. Por ejemplo, los tratados de la Organización Mundial del Comercio incluyen disposiciones que podrían interpretarse para abordar aspectos de la comercialización y los estándares de IA. Las regulaciones de la UE deberán respetar estos acuerdos existentes o buscar su modificación cuando se refieran a la IA.

Además, el Derecho internacional humanitario y los derechos humanos establecen límites en el desarrollo y uso de tecnologías que pueden emplearse en contextos militares o que puedan afectar los derechos individuales. Las regulaciones de la IA deben garantizar que los sistemas de IA sean consistentes con estos principios, prohibiendo usos que violen el Derecho internacional¹⁸.

La UE, con su enfoque progresista en la regulación de la IA, tiene la oportunidad de liderar en la formulación de una legislación internacional en esta materia. Las políticas y normativas que desarrolle pueden servir como un modelo para tratados internacionales futuros y para la legislación de otros países, promoviendo estándares altos en la protección de datos, la privacidad y la seguridad.

IV. APLICACIÓN DEL ARTÍCULO 2.1. POR LOS TRIBUNALES DE JUSTICIA

El RIA también plantea cuestiones de jurisdicción y aplicación extraterritorial. La UE debe trabajar para garantizar que sus regulaciones sean respetadas más allá de sus fronteras, lo cual es un reto significativo en el espacio digital globalizado. Esto puede requerir acuerdos bilaterales o multilaterales, así como un diálogo constante con otras jurisdicciones para asegurar la cooperación en la supervisión y cumplimiento de estas regulaciones.

El Consejo de Europa y otras organizaciones internacionales de derechos humanos son foros críticos para el diálogo sobre cómo las aplicaciones de IA pueden afectar los derechos humanos. Las regulaciones de la UE pueden influir en la creación de

18. Ministerio de Asuntos Exteriores, Unión Europea y Cooperación de España, Diplomacia regulatoria y la IA, (2023).

directrices globales para garantizar que la IA se desarrolle de manera que respete la dignidad humana y los derechos fundamentales.

Es fundamental que las regulaciones de la UE consideren el impacto en los países en desarrollo, que pueden carecer de la infraestructura para cumplir con estándares estrictos. La cooperación internacional para el desarrollo y la asistencia técnica serán cruciales para asegurar que la IA sea una herramienta de avance y no una fuente de división.

El RIA tiene el potencial de configurar no solo el panorama regulatorio europeo sino también la gobernanza global de la IA. Sin embargo, para que sea efectivo y justo debe articularse dentro del marco del Derecho internacional, respetando los tratados existentes y contribuyendo al desarrollo de nuevos estándares y principios. Esto requiere un esfuerzo concertado para la cooperación internacional, la diplomacia tecnológica y la promoción de un enfoque inclusivo y holístico que abarque todas las regiones y sectores de la sociedad.

El RIA por parte de la UE introduce desafíos significativos en términos de jurisdicción y alcance territorial. La naturaleza inherentemente global de la IA y su industria asociada exige un escrutinio detallado de cómo la normativa de un territorio puede influir en (o ser implementada por) entidades que operan internacionalmente. La aplicación del artículo 2.1 del RIA hace evidente la necesidad de un enfoque holístico y globalizado para la regulación, con un énfasis en la cooperación internacional y la armonización regulatoria.

La preocupación primordial en cuanto a la jurisdicción es el alcance extraterritorial del RIA. Es decir, la UE debe definir cómo sus normas afectarán a las empresas y entidades fuera de su territorio que producen o proveen sistemas de IA utilizados dentro de la UE.

La aplicación del artículo 2.1 del RIA por las autoridades nacionales competentes enfatiza la necesidad de un diálogo global y la colaboración para desarrollar un enfoque armonizado y equitativo hacia la regulación de la IA. En última instancia, el éxito de la UE en la regulación de la IA no se medirá sólo por la eficacia de su legislación interna, sino también por su capacidad para influir y formar parte de un marco normativo global cohesivo.

Las autoridades nacionales competentes no son meros ejecutores de la legislación de la UE en materia de IA; son participantes activos en el escenario regulatorio global. Al aplicar el artículo 2.1 del RIA, estas entidades contribuyen a la formación de un paisaje internacional que es más cohesivo, justo y equilibrado. Su papel va más allá de la implementación de políticas, extendiéndose a la influencia de la gobernanza de la IA a nivel mundial, lo cual es crucial para abordar los retos que la tecnología presenta en una sociedad interconectada.

El propio artículo 2.1 del RIA señala además que, en ocasiones, puede ocurrir que la aplicación del reglamento se plantee en el marco de una acción civil presentada ante un órgano judicial relativa, por ejemplo, a una responsabilidad civil extracontractual derivada del funcionamiento defectuoso de un sistema IA, o al incumplimiento de un contrato celebrado entre un proveedor de sistemas IA y un usuario, o a una disputa entre cualquiera de estos y un particular que es parte de un contrato de prestación de servicios en los que se utilizan estos sistemas. En dichos litigios, el incumplimiento

de los requisitos u obligaciones establecidos en el Reglamento para las distintas categorías de sistemas IA puede ser invocado como fundamento de la demanda.

Tratándose de litigios en materia civil o mercantil, la competencia judicial internacional del tribunal del Estado miembro ante el que se presente la demanda vendrá determinada por el Reglamento «Bruselas I bis» —siendo competentes los Tribunales del estado donde el presunto perjudicado tenga su residencia habitual, los del lugar de trabajo o donde se produjo la infracción del RIA—; y la ley aplicable por el Reglamento «Roma II» —si se trata de un litigio por responsabilidad civil extracontractual— o el Reglamento «Roma I» —si el litigio es sobre el incumplimiento de un contrato internacional—. Eso sí, el derecho sustantivo aplicable (la *lex causae*) será el propio RIA, y no podrán aplicar derecho extranjero de un tercer Estado.

V. APLICACIÓN EXTRATERRITORIAL DEL REGLAMENTO EUROPEO DE INTELIGENCIA ARTIFICIAL

El objetivo de este amplio ámbito territorial es que la protección que ofrece el RGPD «viaje» con los datos personales allá donde vayan en una sociedad globalizada donde los datos cruzan fronteras con un simple clic. La UE se guía por el razonamiento de que ofrecer protección solo para el procesamiento de datos que tiene lugar dentro de las fronteras europeas no sería suficiente. Esta medida también busca ofrecer igualdad de condiciones para las empresas europeas sin crear una regulación más estricta que supusiera cargas solo para ellas. La aplicación extraterritorial del RGPD significa que cualquier empresa que desee acceder al mercado europeo para ofrecer sus servicios y bienes y tratar datos personales «europeos» debe cumplir con estas reglas aunque tenga su sede en un tercer país¹⁹.

La aplicación extraterritorial de la legislación no es algo nuevo²⁰, pero sí que se puede ver que está cobrando mucha fuerza en los aspectos relativos a la regulación de Internet²¹.

El RGPD ha sido duramente criticado, ya que con la cantidad de empresas que se encuadran en estos criterios en todo el mundo es más fácil para las multinacionales

-
19. Ver Gascón Marcén, A., «The extraterritorial application of European Union Data Protection Law», *Spanish Yearbook of International Law*, n.º 23 (2019), pp. 413-425, p. 415.
 20. Ver Dover, R. y Frosini, J., *The Extraterritorial Effects of Legislation and Policies in the EU and US*, European Union, Brussels 2012. Según Gallego, a pesar de que la UE nunca ha sido una completa defensora de la extraterritorialidad, comienza a redoblar su ejercicio a través de la extensión territorial, la cual permite controlar aquellas conductas que, aunque se lleven a cabo en el extranjero, repercutan en los intereses generales de la UE. Véase Gallego Hernández, A. C., «La aplicación de la extensión territorial del Derecho de la Unión Europea», *Cuadernos Europeos de Deusto*, n.º 63 (septiembre) (2020), pp. 297-313, disponible en: <https://doi.org/10.18543/ced-63-2020pp297-313>
 21. Ver Internet Society, *The Internet and extra-territorial effects of laws*, Internet Society, 2018, p. 1. Esta organización advierte que muchos Estados están imponiendo reglas que se extienden a Internet en otros lugares, obstaculizan la innovación, disuaden la inversión en sus propios países y corren el riesgo de crear nuevas brechas digitales que perjudiquen a sus propios ciudadanos.

adaptarse a él mientras que es muy costoso para las *pymes*²². Además, las autoridades de protección de datos en los Estados miembros tienen recursos limitados, por lo que, como habrá más empresas extranjeras que no cumplan con el RGPD que recursos para investigarlas, la aplicación real del mismo necesariamente será arbitraria, lo que socavaría la legitimidad de cualquier acción de ejecución que se adopte²³. Sin embargo, se podría considerar legítima esta aplicación extraterritorial y argumenta que la UE está equipada con las herramientas relevantes para hacer cumplir el RGPD en el exterior, aunque haya que desarrollarlas más²⁴. Este enfoque, aunque no sin inconvenientes y desafíos para los intereses estatales y los derechos individuales, resuelve uno de los mayores problemas a los que se enfrentaba hasta entonces la normativa europea de protección de datos, que era la falta de jurisdicción sobre los responsables del procesamiento de datos en terceros países que afectaban a un número considerable de datos de europeos²⁵.

Los legisladores europeos eran bastante conscientes de que la aplicación extraterritorial de las leyes podía tener impactos indeseables. El propio RGPD en su considerando 115 establece que la aplicación extraterritorial de algunas leyes, reglamentaciones y otros actos jurídicos puede ser contraria al Derecho internacional e impedir la protección de las personas físicas garantizada en la UE en virtud del RGPD; y, por tanto, las transferencias de datos sólo deben hacerse respetando las condiciones del mismo. Así, vemos que el RGPD establece su propia aplicación extraterritorial, pero excluye la de las leyes extranjeras en muchos casos. Un conflicto de esta naturaleza puede darse, por ejemplo, cuando las autoridades estadounidenses requieran datos en el marco de una investigación penal a una compañía situada en su territorio, pero que sean referentes a un residente de la UE en contra de lo establecido en el RGPD, por lo que la empresa puede encontrarse con obligaciones legales contradictorias.

Los problemas son múltiples y los críticos tienen buenas razones para estar preocupados, pero la dificultad para garantizar la aplicación del RGPD o la falta de recursos para ello no pueden hacer que apuntemos a estándares más bajos de protección de los derechos fundamentales; sobre todo teniendo en cuenta cómo el RGPD ha servido para elevar este nivel de protección no sólo en Europa.

Para comprender la naturaleza de la extraterritorialidad en el RIA, es esencial analizar los dos elementos clave que subyacen a su aplicación.

El primer elemento es el criterio de «oferta» y «uso». Según el Reglamento, las regulaciones se aplicarán no sólo a las entidades que ofrecen servicios de IA en la UE,

22. Ver Scott, M; Cerulus, L; y Kaya LI, L., «Six months in, Europe’s privacy revolution favors Google, Facebook», *Político.eu*, 23 de noviembre de 2018.

23. Ver Svantesson, D. J. B., «European Union Claims of Jurisdiction over the Internet — an Analysis of Three Recent Key Developments», *Journal of Intellectual Property, Information Technology and E-Commerce Law*, vol. 9, n.º 2 (2018), pp. 113-125, p. 118.

24. Ver Azzi, A., «The Challenges Faced by the Extraterritorial Scope of the General Data Protection Regulation», *Journal of Intellectual Property, Information Technology and E-Commerce Law*, vol. 9, n.º 2 (2018), pp. 126-137, p. 137.

25. Ver De Hert, P. y Czerniawski, M., «Expanding the European data protection scope beyond territory: Article 3 of the General Data. Protection Regulation in its wider context», *International Data Privacy Law*, vol. 6, n.º 3 (2016), pp. 230-243, p. 230.

sino también a aquellas cuyos sistemas de IA se utilizan en la UE, independientemente de si esa entidad está establecida o no en la UE.

El segundo elemento clave es el principio de «efecto». El principio de efecto implica que, si un sistema de IA tiene un impacto significativo en los individuos o entidades en la UE, entonces la ley se aplicará. Esto se extiende incluso a sistemas desarrollados y operados completamente fuera de la UE, lo que destaca la intención del Reglamento de proteger a sus ciudadanos de riesgos potenciales independientemente de la ubicación de la empresa de IA.

La extraterritorialidad en el RIA también se refleja en las obligaciones de las entidades no europeas. Estas empresas deberán designar a un representante legal en la UE para asegurarse de que cumplen con la ley y actuar como un punto de contacto con las autoridades regulatorias. Esto es similar a los requisitos establecidos por el RGPD y es fundamental para asegurar que las entidades no europeas puedan ser sujetas a supervisión y sanciones si no cumplen con los estándares establecidos.

Este enfoque tiene implicaciones significativas para la gobernanza global de la IA: a) por un lado, establece un alto estándar que podría inspirar a otras jurisdicciones a seguir su ejemplo, promoviendo una forma de «diplomacia regulatoria»; y, b) por otro lado, también plantea preguntas sobre la soberanía y el equilibrio de poder en la regulación de las tecnologías emergentes.

La extraterritorialidad, sin embargo, no está exenta de críticas y preocupaciones. Algunos argumentan que esta podría conducir a conflictos de leyes, donde las empresas se encuentran atrapadas entre regulaciones incompatibles de diferentes jurisdicciones. Sin duda alguna, la carga administrativa y financiera de cumplir con múltiples sistemas regulatorios puede ser onerosa, especialmente para las *startups* y las *pymes*. Para abordar estas preocupaciones, la UE puede necesitar colaborar con socios internacionales para desarrollar estándares comunes o mecanismos de reconocimiento mutuo que faciliten el cumplimiento transfronterizo. Además, la UE debe considerar los impactos económicos de sus regulaciones extraterritoriales y equilibrar la protección de los consumidores con un entorno propicio para la innovación y el comercio.

El artículo 2.1 a) del RIA es un criterio de conexión informado por la doctrina jurisprudencial de las «actividades dirigidas», utilizado por ejemplo en materia de contratos celebrados por los consumidores en Internet, o de infracción en línea de títulos unitarios de propiedad industrial. Este criterio garantiza que el propio Reglamento resulta aplicable en situaciones que presentan una estrecha vinculación con la UE.

El artículo 2.1 b) del RIA es un criterio criticable por dos razones: a) para empezar, la utilización de los términos «estén situados en la UE» otorga al propio Reglamento un ámbito de aplicación extremadamente amplio. La aplicación resulta injustificada pues la situación presenta una vinculación muy escasa con la UE. Este problema se solucionaría con una modificación de la disposición que limite su aplicación a usuarios establecidos o con residencia habitual en la UE; y,

Efectivamente, el RIA no resulta aplicable a proveedores establecidos en la UE que comercializan sus sistemas IA exclusivamente en terceros Estados; en cambio, si resulta aplicable a usuarios establecidos en la UE que prestan sus servicios en terceros Estados. La diferencia de trato resulta injustificada. En ambos casos la vinculación

con la UE es la misma. Si la intención es que los usuarios europeos de sistemas IA respeten los estándares previstos en el Reglamento con independencia del país en el que ofrezcan sus servicios, los proveedores establecidos en la Unión Europea que comercialicen sistemas IA en terceros Estados también deberían cumplir con esos estándares.

Alternativamente, se podría defender una modificación del artículo 2.1 b) para que el Reglamento únicamente fuera aplicable a implementadores de sistemas IA cuando la información de salida generada por el sistema se utilice en la UE, independientemente de si tienen su residencia habitual o establecimiento en territorio europeo o no.

El criterio del artículo 2.1 c) del RIA es un criterio que puede conllevar una aplicación extraterritorial injustificada del propio Reglamento; y puede resultar aplicable en situaciones difícilmente previsibles para proveedores de sistemas IA establecidos en terceros Estados²⁶.

VI. REFLEXIÓN FINAL

El RIA se configura como un nuevo estándar global regulatorio en materia de IA, gracias al ámbito de aplicación territorial extremadamente amplio que le otorga su artículo 2.1. Esa aplicación territorial extremadamente amplia del RIA no siempre está justificada, pues los criterios de conexión con la UE, como ya hemos señalado, son «escasos», en ocasiones.

Quizás el recurso a la vía convencional, bilateral o multilateral, para extender los estándares regulatorios europeos más allá de nuestras fronteras sería lo más óptimo para ayudar a garantizar el cumplimiento del RIA por parte de proveedores y usuarios establecidos en terceros Estados; sobre todo, en consonancia con la tradición europea en política exterior de búsqueda de consensos a través de negociaciones bilaterales o multilaterales. La cooperación internacional en el contexto de la RIA es un paso crucial hacia la creación de un entorno global seguro y ético para el desarrollo y la aplicación de la IA. A medida que la tecnología avanza y su impacto se globaliza, trabajar juntos se vuelve indispensable para manejar sus desafíos y maximizar sus beneficios.

En definitiva, la aplicación extraterritorial del RIA nos lleva a reflexionar, a mayores, si debemos abandonar «la idea de Europa» de STEINER: Europa siempre ha creído (y creará) que perecerá; que se puede consolidar y progresar; y, en definitiva, servir de espejo y de competencia para otros países.

Ojalá la extraterritorialidad del RIA (y su carácter de «norma de contención y de reorientación») no convierta a la UE en una isla rezagada en el mundo que no nos deje avanzar en innovación. No obstante, tendremos que esperar unos cuantos años más para analizar el verdadero impacto extraterritorial del RIA.

26. Ver López-Tarruella Martínez, A., «El reglamento...» cit. pp. 14-17.

La exclusión de los sistemas inteligencia artificial de seguridad nacional, defensa y militares del Reglamento y el Derecho aplicable

ÁNGEL GÓMEZ DE ÁGREDA

Doctor Universidad Politécnica de Madrid. Ministerio de Defensa de España.
Odiseia¹

I. INTRODUCCIÓN

La inteligencia artificial (IA), como concepto, resulta un campo muy amplio para considerar su regulación en conjunto. La aproximación que están siguiendo todos los países es la de estudiar los distintos empleos que tiene. Aún esta es una labor difícil por la amplia casuística de cada uno de ellos, por la evolución constante a que están sometidas estas tecnologías y, no en menor medida, por la ventaja sustancial que ofrece su empleo a aquellos que se encuentran en vanguardia de su estudio.

Esto último es particularmente cierto cuando se trata de usos militares —en sentido amplio—. Igual que ha ocurrido históricamente con todos los avances tecnológicos, aquellos que hacen un uso más intensivo de ellos suelen resistirse a que se coarte este, mientras que los menos adelantados aducen todo tipo de riesgos y amenazas. El caso de la prohibición de ballestas y arcos largos en el Segundo Concilio de Letrán aduciendo la indignidad de una muerte (de los nobles y caballeros) a distancia tiene notables paralelismos con la actualidad.

La capacidad de muchas tecnologías —y, muy especialmente, de la IA— para ser empleadas en aplicaciones civiles de carácter beneficioso y en otras de índole bélica y naturaleza destructiva se conoce como uso dual de la tecnología. Es doble naturaleza obliga a una visión mucho más amplia de las posibles aplicaciones y, por consiguiente, de los aspectos a regular.

Al mismo tiempo, el mero hecho de incluir los usos bélicos de estos medios es susceptible de generar un cierto rechazo social a su desarrollo o, al menos, condicionar

1. ORCID: 0000-0003-1036-6324. El presente trabajo se realiza en el marco del Proyecto “Derechos y garantías públicas frente a las decisiones automatizadas y el sesgo y discriminación algorítmicas” 2023-2025 (PID2022-136439OB-I00) financiado por MCIN/AEI/10.13039/501100011033/ FEDER, UE.

el mismo a un escrutinio mayor. En un contexto de avances muy rápidos, el gran poder acumulado por las corporaciones tecnológicas y el de sus estados de referencia se combina, en este caso, para ralentizar cualquier proceso regulador que pueda ofrecer una ventaja competitiva a un rival.

Por todas estas causas, y por la complejidad y especificidad del tema, la mayor parte de la legislación y de los códigos éticos relativos a la IA deja de lado el tratamiento de su uso bélico.

En muchos casos, directamente ignorando en su desarrollo estos posibles usos. En otros, negando la legitimidad de este empleo y abogando por su prohibición con poca o ninguna argumentación o visos de realismo. Finalmente, otros muchos códigos reconocen la posibilidad de este uso dual, pero declinan incluir los aspectos militares dentro de su tratamiento. De esta manera, distorsionan —y blanquean— la imagen de la IA de forma consciente.

Si esto ocurre en los códigos de carácter «civil» (no militar), los códigos que se han desarrollado recientemente para abordar de forma específica este empleo también tienen tendencia a adolecer de dos problemas comunes. Por un lado, aunque muchos recogen en su articulado la necesidad de no hacerlo, condicionan el desarrollo beneficioso de estas tecnologías. Por otro, tienden a asociar los usos militares de la IA con los sistemas de armas autónomos letales (SALA), esto es, con los «robots asesinos».

En realidad, la mayor parte de las aplicaciones de IA al campo militar no están asociadas a la conducción de plataformas o vectores autónomos, sino al análisis de datos, a la toma de decisión o a labores logísticas. A pesar de su carácter muchas veces no letal, la especificidad del marco en que se desarrollan (y de la legislación aplicable al mismo) y los efectos que pueden provocar en acciones bélicas requieren también de un tratamiento específico.

De hecho, se ha argumentado que los principios éticos y las normas jurídicas que se desarrollen para el ámbito militar podrían ser susceptibles de encerrar importantes enseñanzas y conclusiones para la regulación de los sistemas civiles. La letalidad asociada a muchas de estas aplicaciones y la visibilidad de los efectos que provocan ilustran mejor que otros usos factores comunes a cualquier sistema dotado de IA.

En este sentido, es importante destacar que es preciso diferenciar entre el diseño, desarrollo y puesta en servicio de los sistemas dotados de IA, y el empleo que se hace de estos sistemas. A modo de ejemplo, la regulación específica de la tecnología debería centrarse en que los sistemas de tratamiento de datos biométricos estuvieran libres de sesgos raciales, de género, de clase o de cualquier otro tipo, y que tuvieran presente la necesidad de la protección de la privacidad de los individuos. Mientras tanto, otros códigos deberían contemplar si esta aplicación se utiliza para la selección de personal de una empresa, para el apoyo a los tribunales en relación con una lista de sospechosos, o para la selección de blancos para un sistema de defensa perimetral autónomo en una instalación militar.

Pretender ignorar las posibles aplicaciones bélicas de los desarrollos tecnológicos es tan peligroso como forzar la renuncia a estos mismos desarrollos en función de su potencial empleo dañino.

Otro tanto puede decirse de la diferenciación necesaria entre la regulación de la tecnología y la de las técnicas de empleo de esta. Uno de los principales avances que introduce el estudio de los sistemas dotados de IA es el mayor conocimiento de los mecanismos cognitivos humanos y una mayor capacidad para utilizar los algoritmos para afectarles. A eso hay que añadir que la menor adaptabilidad de los sistemas mecánicos hace que se esté rediseñando el entorno para favorecer la comprensión de este por parte de las máquinas, otorgando ventaja competitiva a las inteligencias artificiales sobre las naturales.

Muchos de los códigos éticos —y la práctica totalidad de las empresas y gobiernos— terminan por enfatizar la necesidad de evitar el coste de oportunidad de demorar la investigación y el desarrollo de la IA con carácter general y en beneficio de la humanidad en función de las consecuencias negativas que un uso maligno de estas pueda derivar. Es urgente, por consiguiente, establecer los controles necesarios para minimizar los efectos perversos del uso dual de estas tecnologías desde el momento mismo en que se están diseñando.

El Convenio sobre inteligencia artificial, derechos humanos, democracia y Estado de Derecho del Consejo de Europa se mantiene en la misma línea de excluir la defensa nacional del ámbito de sus competencias. En este sentido, apenas si matiza el RIA en un recordatorio de que el Derecho, o los derechos, en su caso, sigue siendo aplicable a estos sistemas mientras no se excluya explícitamente. Se trata de una vaga referencia que ya se ha utilizado previamente en otros casos como la ciberseguridad (Manual de Tallin) o la privacidad.

En el presente documento se tratará sobre el tratamiento que se otorga a los usos militares de la IA en textos relativos al primer aspecto, el tecnológico. Adicionalmente, se efectuará una revisión de la legislación internacional aplicable a los sistemas dotados de IA y su empleo *ad bellum* y *in bello*. En este último caso, las referencias son generalmente indirectas y generales, no aplicables exclusivamente al caso de la IA.

II. CÓDIGOS ÉTICOS PARA INTELIGENCIA ARTIFICIAL DE USO GENERAL

En los primeros códigos éticos de la IA [2016 con Satya Nadella (Nadella, 2016) a 2019] la tendencia es a ignorar los posibles usos bélicos o duales de la IA. En algunos casos, se excluye explícitamente, en otros, simplemente no se menciona o se aboga por su prohibición. La excepción son el Ethically Aligned Design V.2 del IEEE (IEEE, 2016) y el COMEST (la Comisión Mundial para la ética del conocimiento científico de la UNESCO) (COMEST, 2017). Ambos abogan por reforzar el control y la responsabilidad humana sobre las máquinas, y por el desarrollo de códigos éticos y jurídicos específicos para este tipo de sistemas autónomos.

En casi todos los casos se asocia erróneamente el uso bélico de la IA con los sistemas autónomos letales, cuando la mayor parte —y la más peligrosa— tiene que ver con el proceso de toma de decisión, y con el apoyo logístico y de inteligencia. Su empleo en ciberataques, no siendo normalmente letal, también reviste gran peligrosidad.

En (Gómez-de-Ágreda, 2020) se resumen estos primeros códigos éticos relativos a la IA y se detalla su relación con las aplicaciones militares de esta. Con posterioridad se han elaborado numerosos códigos de carácter específico para el uso bélico. A

diferencia de los primeros —lógicamente— estos últimos están redactados por los ministerios de Defensa de los distintos gobiernos y, por lo tanto, adolecen de una falta de visión plural.

En el estudio de los distintos códigos es preciso, por consiguiente, tener muy presente quién es el autor y cuáles son sus intereses. Mientras que las empresas tienden a justificar la necesidad de mantener el desarrollo y la innovación, la sociedad civil suele mostrarse muy cauta ante los usos nocivos de la IA. Sin embargo, la cautela está más relacionada con sus posibles usos «civiles» perversos que con la dualidad en la posibilidad de empleo de los algoritmos.

III. USO DUAL DE LAS TECNOLOGÍAS

La ética de la IA se ha relacionado en numerosas ocasiones con la de otras tecnologías, como la nuclear. En ambos casos existen numerosas y muy importantes aplicaciones civiles que justifican sobradamente la investigación en dicho campo. También comparten ambas la criticidad de los posibles usos nocivos de sus versiones militares: un ataque nuclear o las posibles consecuencias de una IA de carácter general.

No obstante, mientras que las dos están presentes en el imaginario popular, la IA sigue vinculada al ámbito de la ficción, de la hipótesis y de la distopía. No existe, en realidad, una percepción de peligrosidad real del uso de los algoritmos. El relato dominante se centra en los riesgos derivados de desarrollos futuros a medio o largo plazo, mientras que deja de lado —de forma interesada en muchos casos— las amenazas reales de los usos presentes de esta tecnología. De este modo, más que preocupación, se crea un cierto morbo desligado de la realidad que aleja, para el gran público, la urgencia de regular su uso.

Otras tecnologías asociadas a lo militar —como las empleadas en la guerra electrónica— están muy alejadas del día a día de la población como para generar alarma social. Al mismo tiempo, la falta de un mercado eficiente en el sector civil limita mucho la dualidad de su uso y la necesidad de control desde ese ámbito.

La intangibilidad de la IA, distribuida a menudo como código abierto con escasa o nula supervisión sobre en qué aplicaciones se integra, contribuye a su invisibilidad.

La ética y la legislación relativa al uso de la IA debe, por el contrario, asociarse a otras tecnologías más vinculadas a la libertad y la voluntad humana. Entre ellas estaría la neurociencia, o la biotecnología. En ambos casos se trata de disciplinas que afectan a la naturaleza misma del ser humano, algo que la IA puede llevar a cabo de manera indirecta. Como queda dicho, la programación de los algoritmos se ha beneficiado mucho de los avances en la comprensión de la faceta cognitiva humana, y viceversa.

Dentro de este paralelismo, cabría aplicar conceptos como el «Dual use research of concern», utilizado en The Human Brain Project (Aicardi, 2018), que se aplica a las fases iniciales del desarrollo de estas tecnologías. La tendencia actual es, además, a hacer un uso conjunto de todas estas disciplinas para conseguir modelar al individuo en todas sus facetas, desde la genética a la cognitiva.

En cualquier caso, ninguna de estas disciplinas tiene tampoco el carácter «democrático» de la IA que hace que su evolución esté ampliamente distribuida

entre numerosos actores, muchos de ellos ni siquiera profesionales del sector. La capacidad de la IA para crecer de forma incontrolada no tiene paralelismos en otras ciencias modernas.

Ni siquiera los sistemas autónomos empleados en combate reciben un rechazo frontal por parte de la opinión pública, que se va adaptando según las necesidades operativas y el relato oficial explican las ventajas —indudables, por otra parte— de su empleo.

Esta falta de percepción de los riesgos vinculados al uso dual de la IA está relacionada con el hecho de que los sistemas se interpretan en su conjunto, sin desglosar los distintos componentes dotados de IA entre sí ni de estos con las plataformas que los albergan. Un dron de combate autónomo comparte muchas características comunes con un automóvil no tripulado. Pero, además, un despiece mayor permite identificar subsistemas que, por sí mismos, no parecen resultar una amenaza para los seres humanos: sistemas de identificación de imágenes, por ejemplo.

La dualidad del uso de la IA no se circunscribe, por consiguiente, solo a sistemas completos, sino a los distintos algoritmos que les permiten funcionar y que, aplicados sobre plataformas diferentes o en conjunción con otros algoritmos, pueden generar amenazas distintas (en el caso de la identificación de imágenes, por seguir con el ejemplo, en relación con la privacidad).

IV. EL DECÁLOGO DE PRINCIPIOS ÉTICOS DEL CCW PARA SISTEMAS DE ARMAS AUTÓNOMOS LETALES

La Convención de Naciones Unidas para ciertas armas (que pueden ser especialmente dañinas) lleva manteniendo reuniones semestrales desde 2013 para intentar llegar a acuerdos internacionales que regulen este tipo de armas. Este tipo de foros aporta una gran legitimidad al estar representados los Estados y la sociedad civil (ICRC, por ejemplo), pero carece de capacidad coercitiva y de un liderazgo claro. Esta circunstancia queda patente en el carácter desiderativo más que impositivo de sus enunciados.

De hecho, las principales naciones e industrias implicadas son las primeras en dilatar artificialmente cualquier proceso de adopción de legislación que pueda constreñir su uso mientras ellas tengan la ventaja.

El CCW solo ha conseguido elaborar un decálogo (en 2018) en el que viene a afirmar la aplicabilidad de Derecho Internacional Humanitario a cualquier tipo de acción bélica con independencia del armamento que se utilice en ella.

En su reunión de 2018, Austria, Brasil y Chile elevaron una propuesta que culminó en la puesta en marcha de un grupo de trabajo abierto para negociar un acuerdo vinculante (Austria et al., 2018). De forma simultánea, el CCW llegó a un acuerdo —en este caso, no legalmente vinculante— sobre un decálogo de principios (CCW, 2018b) que, si bien no tienen exclusivamente un carácter ético, representan el mayor avance del grupo de trabajo hasta el momento y siguen suponiendo un punto de partida a tener en cuenta para su posterior expansión:

1. El Derecho Internacional Humanitario sigue siendo de completa aplicación a todos los sistemas de armas, incluyendo el potencial desarrollo y utilización de sistemas de armas autónomos letales.

2. La responsabilidad humana sobre las decisiones relacionadas con el uso de sistemas de armas tiene que mantenerse ya que la responsabilidad jurídica no puede transferirse a las máquinas. Este principio debería ser de aplicación a lo largo del ciclo de vida completo del sistema de armas.

3. Se debe asegurar la responsabilidad jurídica por el desarrollo, despliegue y uso de cualquier sistema de armas emergente en el marco de la CCW de conformidad con el Derecho Internacional aplicable, incluyendo por la operación de dicho sistema de armas dentro de una cadena humana de mando y control responsable.

4. De conformidad con las obligaciones estatales bajo el Derecho Internacional, durante el estudio del desarrollo, adquisición o adopción de una nueva arma, medio o método de guerra se debe determinar si su empleo podría, en algunas o todas las circunstancias, estar prohibido por el Derecho Internacional (aplicabilidad de la cláusula Martens) (Hague Convention (II) on the Laws and Customs of War on Land, 1899; Ticehurst, 1997).

La cláusula Martens, mencionada de forma reiterada a lo largo de las discusiones del CCW, establece la necesidad de aplicar a tipos novedosos de tácticas o armamento los mismos criterios que, por sentido común, se puedan extrapolar de la norma aplicable de forma general. Esta lógica difusa del sentido común está todavía más lejos del alcance —al menos en el estado del arte actual— de los sistemas autónomos y está detrás de buena parte de la opinión pública hostil a estos sistemas (M. C. Horowitz, 2016). De forma similar, no cabe esperar que se respete el principio de humanidad —y las muestras de compasión— por parte de un sistema diseñado para optimizar la ventaja en el combate.

5. Durante el desarrollo o adquisición de nuevos sistemas de armas basados en tecnologías emergentes en el área de los SALAS se deberían considerar las salvaguardas físicas y no físicas (incluyendo la ciberseguridad contra hackeo o suplantación de datos), el riesgo de adquisición por grupos terroristas y el riesgo de proliferación.

6. La evaluación de riesgos y de medidas de mitigación deberían ser parte del ciclo de diseño, desarrollo, pruebas y despliegue de tecnologías emergentes en cualquier sistema de armas.

7. Se debería prestar consideración al uso de tecnologías emergentes en el área de los SALAS en el aseguramiento del cumplimiento de las normas de Derecho Internacional Humanitario y otras obligaciones legales internacionales.

8. En el desarrollo de potenciales medidas normativas, las tecnologías emergentes en el área de los SALAS no deberían ser antropomorfizadas.

9. Las deliberaciones y cualquier medida normativa potencial que se tome en el contexto de los SALAS no deberían evitar el progreso de o el acceso a usos pacíficos de tecnologías autónomas inteligentes.

10. El CCW (...) busca conseguir un equilibrio entre la necesidad militar y las consideraciones humanitarias.

Es interesante observar cómo el decálogo hace acertadamente referencia a aspectos vinculados al empleo de las tecnologías, y no a las tecnologías mismas. Desde su redacción, no obstante, la polarización política internacional ha impedido ulteriores avances o, incluso, la implementación práctica de los preceptos que contiene.

La mayor parte de los códigos éticos generales incluyen la beneficencia como su primer principio. En el caso de los sistemas de armas se suele hablar de beneficencia relativa, siendo este también un concepto muy discutible. Para apoyar esta «relatividad» —respecto de una acción equivalente llevada a cabo por una inteligencia humana— se apela en la mayor capacidad de discriminación de los sensores y algoritmos artificiales que los sentidos y razonamiento humanos.

De este modo, la aplicación de, por ejemplo, el principio de distinción (entre combatientes y no combatientes) se beneficiaría de la mayor agudeza de las cámaras y micrófonos. Muchas organizaciones no gubernamentales y académicos cuestionan este argumento y la capacidad real de discernir entre actitudes agresivas o pacíficas. Esta falta de habilidad para ir más allá de lo directamente mensurable es un argumento más para mantener viva la revisión periódica de los criterios éticos aplicables al uso de sistemas dotados de IA.

V. DIFERENCIAS SIGNIFICATIVAS ENTRE LOS USOS CIVILES Y MILITARES DE LA INTELIGENCIA ARTIFICIAL

El grado de complejidad del entorno militar —y, especialmente, del bélico— es muy superior al civil. En primer lugar, porque el escenario en el que tiene que desarrollar su acción es mucho más amplio en el caso del militar (como, en el caso de los vehículos no tripulados, la necesidad de navegar por terrenos no preparados y no solamente por carreteras convencionales). Además, los sistemas militares tienen que hacer frente a una potencial acción adversaria, en lugar de una colaborativa con el resto de los sistemas como ocurre en el mundo civil.

Evidentemente, en muchas ocasiones, los sistemas militares estarán vinculados a decisiones y acciones críticas; circunstancia que solo se da en contados casos en el entorno social no bélico. Esto no concierne únicamente a los sistemas de armas autónomos con capacidad letal, sino también a las tomas de decisión o a la coordinación de las operaciones militares. En muchos de estos casos, la conclusión afecta a vidas humanas y, por lo tanto, implica criterios éticos que pueden no ser tan relevantes en otros empleos.

Esta criticidad fuerza una interpretación peculiar del uso militar de los sistemas dotados de IA. Un caso concreto puede encontrarse en la búsqueda de predictibilidad de los resultados de los algoritmos que, en muchos códigos éticos, aparece como un requisito a cumplir. Sin embargo, un sistema predecible resulta enormemente vulnerable en un entorno adversarial. Si el enemigo puede prever la reacción del sistema, también puede contrarrestarla.

Igual que con la «beneficencia» —otro de los principios éticos más generalizados— que se transforma en «beneficencia relativa» en el ámbito militar, la predictibilidad se transforma en mera fiabilidad («reliability») (ICRC, 2018) deployment and use of emerging technologies in the area of lethal autonomous weapons systems (Additional remarks. Es decir, se exige una consistencia en el cumplimiento de los objetivos, pero se rechaza la repetición automática de los medios para alcanzarlos. Los sistemas tienen que ser previsibles solo para el usuario, pero opacos para el adversario.

Por descontado, en el ámbito del Derecho, la actividad militar se ve afectada por una legislación diferenciada específica y con un tratamiento muy distinto del civil o el penal. La aplicación de normas de Derecho Internacional Humanitario,

las convenciones específicas (Haya y Ginebra) y el Derecho de la Guerra (Ley de los conflictos armados) son específicas del entorno militar y bélico. La letalidad, proscrita en el ámbito civil, se convierte en un dato de partida en el militar. No es su exclusión lo que se persigue, sino su restricción a unas circunstancias concretas y a unos fines tasados.

Un riesgo adicional viene dado por el uso de sistemas dotados de IA todavía no maduros en su desarrollo. Este es más probable en el ámbito bélico que en el civil al ser mayor la tolerancia al riesgo y menor la supervisión jurídica previa al uso. Históricamente, esta mayor flexibilidad ha dado lugar a numerosas innovaciones que, posteriormente, se han trasladado al mundo civil. Sin embargo, estas han venido muchas veces acompañadas de un doloroso precio y en graves atentados contra la dignidad humana según se entendía en ese momento.

VI. APLICABILIDAD DEL DERECHO INTERNACIONAL A LOS SISTEMAS DOTADOS DE INTELIGENCIA ARTIFICIAL

Queda, por tanto, sentada la premisa de que el Derecho Internacional, en concreto el Derecho Internacional Humanitario, tiene absoluta e indiscutible aplicación en el uso de los sistemas dotados de IA. El consenso alcanzado en el CCW no deja margen a duda alguna al respecto.

No cabe discrepancia con el criterio del CCW toda vez que el preámbulo del Convenio de La Haya (II) de 1899 relativo a las leyes y costumbres de la guerra terrestre ya incorpora la llamada «Cláusula Martens» que indica: «Hasta que se publique un código más completo de las leyes de la guerra las Altas Partes Contratantes estiman oportuno declarar que en los casos no incluidos en los Reglamentos adoptados por ellas, las poblaciones y los beligerantes permanecen bajo la protección y el imperio de los principios del derecho internacional tal como resultan de los usos establecidos entre naciones civilizadas, de las leyes de la humanidad y de las exigencias de la conciencia pública».

Para aquellos tentados de argüir que, desde entonces, se han publicado códigos más completos como los que reclama el enunciado de la cláusula hay que recordar que tanto los Convenios de Ginebra de 1949 como los dos Protocolos adicionales de 1977 la reafirman y la resaltan llevándola, incluso, a su Preámbulo.

El mismo Convenio de Ginebra recoge en su artículo 36 un argumento adicional y complementario a la Cláusula Martes. El enunciado de este artículo demanda que, «en el estudio, desarrollo, adquisición o adopción de una nueva arma, medio o método de guerra, una Alta Parte Contratante tiene la obligación de determinar si su empleo estaría, en algunas o en todas las circunstancias, prohibido por este Protocolo o por cualquiera otra norma de derecho internacional aplicable a la Alta Parte Contratante».

La redacción de normas específicas para los sistemas dotados de IA está generando, no obstante, una importante pugna entre las grandes potencias, la industria y aquellos países con menor acceso a estas tecnologías. Siempre desde la premisa de la universalidad del Derecho establecido, las interpretaciones siguen divergiendo considerablemente en función de los intereses de unos y otros.

VII. RESPONSABILIDAD Y CONTROL HUMANO SIGNIFICATIVO

La responsabilidad en los sistemas de IA resulta, en muchas ocasiones, muy difícil de establecer al combinarse la acción de varios de ellos en una tarea compleja. La obtención de información, su procesamiento y la adopción de decisiones (y su ejecución) pueden estar asignados a sistemas diferentes en momentos distintos.

Aparece aquí el concepto de responsabilidad compartida. En él, la responsabilidad se sigue retrotrayendo a los humanos detrás de cada proceso, pero también detrás de cada parte de la cadena de creación de los dispositivos y sistemas. De este modo, el diseñador retiene su parte de responsabilidad por un uso no conforma a derecho de su obra. La omisión de la creación de salvaguardias para evitar este escenario recae sobre él. Igual ocurre con el desarrollador del diseño, con el integrador, el distribuidor y, naturalmente, con el operador final del sistema y con toda la cadena de mando que contribuye a que se utilice.

Es, en todo caso, siempre el humano —como género— el que ostenta la responsabilidad por los actos cometidos por las máquinas. Igual que ocurre en la investigación de accidentes de aviación, incluso los fallos mecánicos pueden atribuirse a acciones u omisiones humanas en las fases de diseño, elaboración, operación, formación y entrenamiento, mantenimiento y demás.

No cabe, por lo tanto, ampararse en la intencionalidad benéfica de un diseño, sino que se debe prever un posible uso que no lo sea. La IA no es una mera herramienta para alterar un entorno, sino que supone potencialmente un entorno en sí mismo y, como consecuencia, supone una mayor responsabilidad en sus creadores.

Contencioso con el concepto de control humano significativo. Definiciones de distintos organismos.

CNAS ²		Artículo 36	ICRAC ³	ICRC ⁴
Participación humana	Decisiones conscientes informadas	Juicio y acción humana oportuna	Participación cognitiva. Percepción y acción	El humano interviene en todas las fases
Información requerida	Información suficiente sobre el armamento, el objetivo y el contexto	Información precisa sobre la tecnología, el objetivo y el contexto	Naturaleza del objetivo y daños colaterales. Conciencia completa de la situación y contexto.	Información sobre el sistema de armas y el contexto

2. Center for New American Security, think tank estadounidense.

3. International Committee for Robot Arms Control.

4. Comité Internacional de la Cruz Roja.

CNAS ²		Artículo 36	ICRAC ³	ICRC ⁴
Diseño del armamento	Armamento probado. Humano entrenado	Tecnología predecible, fiable y transparente	Suspensión o aborto del ataque	Predictibilidad y fiabilidad
Requisitos legales	Información suficiente para garantizar la legalidad	Accountability hasta cierto punto	Necesidad de que el ataque sea apropiado. Cumplimiento del DIH	Accountability y cumplimiento del DIH

Tabla. Diferencias en los conceptos básicos por parte de distintos organismos. El autor en (Gómez-de-Ágreda, 2020)

La responsabilidad no recae tanto en el ejecutor final de la acción como en el que adopta la decisión de llevarla a cabo. La autonomía humana, el concepto anglosajón de agencia, tiene que ver con esa capacidad de decidir y debe desvincularse del acto físico de «apretar el gatillo». De hecho, la atribución de responsabilidad al ejecutor puede derivar en un descargo injusto de esta. El operador se convierte en el «chivo expiatorio» de una decisión que no ha tomado, o cuya adopción resulta viciada por los sesgos introducidos por los algoritmos que han obtenido, seleccionado e interpretado una información que a él le llega ya digerida.

Los grados de autonomía no pueden, por consiguiente, definirse en función de la proximidad del punto de intervención humana a la decisión final. En muchos modelos, por ejemplo, la ejecución corre siempre de cuenta de la máquina, pero la decisión ha sido tomada por un humano con distintos grados de libertad.

Tampoco la letalidad es un factor a tener en cuenta en la elaboración de códigos éticos y jurídicos para el uso de sistemas dotados de IA en el ámbito militar. La letalidad es un factor inherente al hecho bélico y es el punto de partida de la legislación que lo regula.

En el caso que nos ocupa, la mayor parte de los sistemas dotados de IA de uso en las Fuerzas Armadas no están directamente relacionados con el uso de la fuerza y, mucho menos, son letales o forman parte de un arma. Sin embargo, siguiendo el razonamiento que acabamos de hacer, no procede diferenciar entre los principios aplicables a unos y otros.

De un modo muy especial, hay que tener en cuenta que buena parte de los conflictos —y la vida en general— modernos se libra en el ámbito virtual y en el cognitivo. Con independencia de su efecto directo sobre el mundo material, los actos del ámbito cibernético y de desinformación son ejemplos claros de acciones agresivas que están siendo empleadas como parte de las operaciones militares. Las herramientas que se emplean en ellas deben, por consiguiente, tener una consideración análoga a la del armamento que produce muerte o destrucción directamente.

VIII. USO DE LA INTELIGENCIA ARTIFICIAL EN SEGURIDAD Y DEFENSA SEGÚN EL CONVENIO MARCO SOBRE INTELIGENCIA ARTIFICIAL, DERECHOS HUMANOS, DEMOCRACIA Y ESTADO DE DERECHO

El artículo 3 del primer capítulo del Convenio establece el alcance este. El cuarto punto del artículo, el más conciso de todos, simplemente excluye los asuntos relacionados con la defensa nacional del ámbito de competencias del documento.

Previamente, el segundo punto del artículo también exime a las partes de aplicar el contenido del convenio a las actividades relacionadas con su seguridad nacional. La única matización que contiene es que se entiende que estas actividades tienen que llevarse a cabo respetando el derecho internacional y los derechos humanos, así como las instituciones y procesos democráticos; una suerte de Clausula Marteens aplicada a la inteligencia artificial.

Si bien esto último resulta algo más restrictivo que la exclusión total de la aplicación del convenio que se contempla para los casos propiamente vinculados a la defensa, la propia indefinición ética y normativa, y las diferentes interpretaciones que se han visto en este capítulo proporcionan un escaso anclaje a dichas restricciones.

Resulta significativo que no se incluya en la redacción del apartado relativo a la defensa una referencia similar a la aplicabilidad del Derecho a cualquier sistema bélico, con independencia de su carácter digital o basado en inteligencia artificial.

El RIA ya la exclusión de seguridad y defensa de su ámbito de aplicación. No obstante, cabe apreciar en su redacción algunos matices que el legislador ha introducido.

La no afectación del reglamento a las competencias estatales en materia de seguridad nacional se hace extensiva a cualquier entidad a la que se encomienden estas tareas. Se deja así abierta la puerta a la posibilidad de externalización de las tareas relacionadas con la seguridad nacional en empresas o entidades ajenas a la administración en una suerte de contrapunto a la extensión de la responsabilidad estatal a otras entidades cuando estas actúen en su nombre.

Este aspecto está lejos de resultar menor o intrascendente ya que refleja la participación cada vez mayor de contratistas y corporaciones en tareas relacionadas con la seguridad nacional, sobre todo en tareas tecnológicas o apoyadas en la tecnología. La extensión del paraguas a estos casos podría llegar a dar lugar a abusos o a interpretaciones sesgadas de su espíritu.

En segundo lugar, cabe destacar el hecho de que las exenciones se apliquen únicamente a sistemas que se empleen de forma exclusiva con fines militares, de defensa o de seguridad nacional. La naturaleza abiertamente dual de los desarrollos tecnológicos digitales y, en concreto, los vinculados con la inteligencia artificial, permite cuestionar la posibilidad de que haya diseños cuyo empleo esté cerrado a actividades distintas de las originalmente previstas. Ninguna provisión al respecto se recoge en la legislación, aunque sí se contempla en distintos códigos de conducta o códigos éticos de los mencionados en este capítulo.

IX. CONCLUSIONES

A pesar de que la comunidad internacional no cuestiona la aplicabilidad del Derecho Internacional Humanitario en el uso de los sistemas de armas dotados de

inteligencia artificial, no es previsible que se pueda alcanzar un consenso ejecutivo sobre su empleo en los próximos años.

Por el momento, el CCW de Naciones Unidas se ha concedido hasta finales de 2025 para seguir estudiando el asunto de cara a elaborar criterios éticos y jurídicos sobre este (Lipton, 2023). Son, precisamente, los países más avanzados los que se muestran más reticentes a regular la actividad de unos medios de los que ellos mismos obtienen el mayor partido.

Estados Unidos ha dejado clara su postura de no aceptar imposiciones al tiempo que enfatiza la importancia de que, por el momento, sea preciso ejercer un control humano significativo sobre las acciones de estos sistemas. Rusia discrepa de que esto sea una prioridad. La mayor parte de los países consideran que una demora en la legislación —a la que aboca la falta de consenso y de voluntad— resultará en una presencia masiva de estos sistemas (Hicks, 2023) y una falta de control sobre el grado de autonomía de que pueden estar dotados.

Washington parece haber optado por la vía de los hechos publicando políticas internas, tanto del Departamento de Defensa (*Autonomy in Weapon Systems*, 2023) como del Departamento de Estado (US Department of State, 2023), e invitando al resto de los países a adoptarlas tal y como están redactadas.

Con los precedentes de armas equiparables cabe esperar que no se produzca ningún avance significativo hasta que los sistemas hayan sido empleados en un conflicto entre grandes potencias o hasta que la tecnología esté lo suficientemente madura como para considerarla no disruptiva. Las responsabilidades exigibles lo serán en base a la legislación existente, incluyendo el artículo 36 y la cláusula Marteens como apoyos más significativos.

Entre tanto, el uso dual que potencialmente tienen estas tecnologías —no solo los sistemas completos, sino también componentes particulares— queda igualmente falto de una regulación clara en el ámbito civil.

La urgencia en la regulación de los sistemas de armas autónomos letales no deriva solo de su uso propio, sino de la capacidad que tienen los principios éticos y jurídicos que se adopten de suponer una referencia para otros empleos de la IA.

Tanto el RIA como el Convenio del Consejo de Europa excluyen, no obstante, el ámbito militar —y, en muy buena medida, el de la seguridad nacional— de sus competencias. De un lado, la coyuntura geopolítica actual y, de otro, los intereses de la industria hacen inviable una aproximación más precisa en estos momentos. En el primer caso, por las dificultades para limitar los desarrollos armamentísticos en un ambiente de narrativa prebélica como el actual y por las escasas posibilidades de acuerdo entre las partes rivales implicadas. En cuanto a la industria, lo incipiente de los desarrollos y el vertiginoso ritmo de los mismos favorece, en todo caso, una potenciación de la investigación y la innovación a toda costa, más que la adopción de medidas restrictivas.

No parece probable que la legislación vaya a imponer cortapisas a los desarrollos de la inteligencia artificial en general por su valor económico y estratégico. Mucho menos que, en estos momentos, se pueda plantear cualquier medida que restrinja la posibilidad de obtener ventajas militares de la aplicación de estas tecnologías.

El Reglamento de inteligencia artificial y el Reglamento general de protección de datos

JESÚS JIMÉNEZ LÓPEZ

Director del Consejo de Transparencia y Protección de datos de Andalucía

I. INTRODUCCIÓN

Siendo cierto que el derecho fundamental a la protección de los datos personales, y su estructura normativa e institucional de garantía, ha de aplicarse cualquiera que sea la herramienta tecnológica que se utilice en su tratamiento, también lo es que la Inteligencia Artificial supone desafíos específicos en este ámbito. Es objeto de estas líneas comprobar en qué medida el RIA responde al mismo, en el sentido de posibilitar, coadyuvar o dificultar la aplicación del RGPD a estos sistemas. Por tanto, y desde esta perspectiva, es objeto del presente documento un breve análisis el RIA en su relación con el Reglamento General de Protección de Datos.

Una somera comparación de ambos textos normativos, en concreto los preceptos relativos a su objeto, nos permite comprobar que el RGPD se refiere en todo caso a los tratamientos de datos personales, sea cual sea la forma o tecnología mediante el cual esto tenga lugar, bien por ser necesarios para el desarrollo y uso de sistemas de inteligencia artificial (SIAs), bien por ser tales sistemas el medio para llevar a cabo el tratamiento, como cometido específico.

Por tanto, la primera aclaración que ha de realizarse es que no se pretende hacer un estudio sobre la aplicación del RGPD a los sistemas de inteligencia artificial, en atención a sus peculiaridades, durante su *ciclo de vida* o en el contexto de su *cadena de valor*, trabajo avanzado por autorizada doctrina y por autoridades independientes de control¹.

Queremos en cierta forma contrastar dos textos normativos, RGPD y RIA. En qué medida se relacionan, superponen, complementan o modifican, de forma manifiesta o no, siempre desde el punto de vista de la seguridad jurídica necesaria para la preservación del derecho a la protección de datos personales, en el entorno del imprescindible desarrollo tecnológico. Se trata, en definitiva, de esbozar un marco

1. Palma Ortigosa, Adrián. (2022); AEPD. (2020). Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial; ICO UK. (2023). GUIA INTELIGENCIA ARTIFICIAL Y PROTECCIÓN DE DATOS.

<https://ico.org.uk/media/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection-2-0.pdf>

de certeza en la aplicación del RGPD², no sólo en un nuevo entorno tecnológico, propiciado por los sistemas de inteligencia artificial, sino también en el nuevo entorno normativo que se ha construido en torno a estos.

Como se observará, no siempre resulta fácil discernir qué singularidad en la aplicación del RGPD resulta del texto normativo aprobado (RIA) y cuál se vincula realmente a las características de la tecnología empleada. Por ello, en la descripción de cualquier supuesto de hecho relativo a los SIAs —como tecnología—, se asumirán los presupuestos que hayan sido aceptados como supuesto de hecho a los efectos del RIA, lo que nos permitirá centrar el objeto de estudio.

Señalado lo anterior, nuestro punto de partida es que una regulación ética de la IA pasa por garantizar, con un adecuado marco de seguridad jurídica, el cumplimiento de las normas de protección de datos personales, más aún si consideramos que para el derecho fundamental, para los principios en torno a los cuales se ha estructurado su protección, los SIAs plantean retos específicos, y generan espacios de riesgo, ciertos y actuales³. Nos fue recordado en el proceso de elaboración y aprobación de RIA que *«los datos (personales y no personales) en IA son en muchos casos la premisa clave para las decisiones autónomas, que inevitablemente afectarán la vida de las personas en varios niveles»*.⁴

II. SISTEMAS DE INTELIGENCIA ARTIFICIAL Y TRATAMIENTO DE DATOS PERSONALES

1. SISTEMAS DE INTELIGENCIA ARTIFICIAL

La definición y caracterización de los SIAs, como objeto de regulación por el RIA se había considerado ya indispensable al objeto de garantizar la seguridad jurídica, teniendo en cuenta no obstante la necesaria flexibilidad en el continuo avance tecnológico (Considerando 12 RIA)⁵. En nuestro caso, nos aportará el contexto de aplicación del RGPD, si bien, al ser objeto de otros trabajos en la presente obra, solo identificaremos brevemente aquellos elementos de la definición de SIA que estimamos relevantes.

Dispone el art. 3.1 RIA que el SIA *«es un sistema basado en una máquina diseñado para funcionar con diversos niveles de autonomía y que puede mostrar capacidad de adaptación tras su despliegue y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar información*

2. El cualquier caso, debemos aclarar que el marco jurídico de protección de los datos personales no sólo se encuentra establecido en el RGPD. Consideraremos a estos efectos el artículo 16 TFUE, el artículo 8 CDFUE; el RGPD, el Reglamento (UE) 2018/1725; y la Directiva (UE) 2016/680 (Directiva sobre la aplicación de la ley). Con carácter general, las referencias a la materia se realizarán al RGPD, evitando dar complejidad a las reflexiones generales.
3. De interés, sobre las limitaciones de la protección de datos para dar respuesta a las nuevas necesidades de la IA, COTINO HUESO 2022 pp. 85 y ss.
4. EDPS y EDPB, 2021, p. 8.
5. Puede ponerse en relación con la previsiones relativas a los Actos Delegados de la Comisión Europea (Considerando (52), artículos 6.6, 7, 43.5 y 6, 47.5, 51, 53.6 en relación con el artículo 97 (“Ejercicio de la delegación”) RIA.

de salida, como predicciones, contenidos, recomendaciones o decisiones, que puede influir en entornos físicos o virtuales». Esta definición debe completarse, de modo introductorio de nuevo, con los denominados modelos de IA de uso general, referido a «*un modelo de IA, también uno entrenado con un gran volumen de datos utilizando la autosupervisión a gran escala, que presenta un grado considerable de generalidad y es capaz de realizar de manera competente una gran variedad de tareas distintas, independientemente de la manera en que el modelo se introduzca en el mercado, y que puede integrarse en diversos sistemas o aplicaciones posteriores.*»⁶ (Art. 3.63). Por último, se refiere el RIA al SIA de uso general, como «*sistema de IA que se basa en un modelo de IA de uso general, que tiene capacidad para servir a diversos fines, tanto para su uso directo como para su integración en otros sistemas de IA.*» (Art. 3.66).

En esta definición destacamos ahora⁷:

- Su capacidad para inferir entendida como:
- El proceso de obtención de los resultados, como predicciones, contenido, recomendaciones o decisiones, que pueden influir en entornos físicos y virtuales.
- La capacidad de los sistemas de IA para derivar modelos o algoritmos a partir de entradas o datos.
- Su capacidad de actuar con diferentes niveles de autonomía, por referencia a la intervención o participación humana.
- Su capacidad de adaptación tras su despliegue o implementación, y, mediante autoaprendizaje, de cambio mientras está en uso.
- Persigue objetivos definidos explícitamente o de forma implícita en un entorno específico, operando en un contexto determinado.

Los SIAs se pueden utilizar «*de manera independiente o como componentes de un producto, con independencia de si el sistema forma parte físicamente del producto (integrado) o contribuye a la funcionalidad del producto sin formar parte de él (no integrado)*».

2. Ciclo de vida y cadena de valor de los SIA

Para confirmar y actualizar la aplicación del RGPD a los SIAs en el contexto del RIA, es necesario atender a estos no solo desde un punto de vista estático, sino también considerando todas las fases en su ciclo de vida, y todos sus intervinientes y agentes, así como sus interacciones, su *cadena de valor*, tal y como ha sido expuesto por quienes han abordado esta cuestión.⁸ Ambos conceptos, *ciclo de vida* y *cadena de valor* son mencionados en la RIA⁹ sin ser objeto de una definición precisa. Si resulta del

6. “*excepto los modelos de IA que se utilizan para actividades de investigación, desarrollo o creación de prototipos antes de su comercialización*” (art. 3.63 *in fine* RIA).

7. Considerando (12) RIA.

8. (“AI Watch. Inteligencia artificial para el sector público. Informe del «3er Taller de aprendizaje entre pares sobre el uso y el impacto de la IA en los servicios públicos», 24 de junio de 2021”).

9. Considerando (65); Considerando (69), Considerando (73); Considerando (74); Considerando (110); Considerandos (114 y 115); Artículo 9.2, relativo al sistema de gestión de riesgos; artículo 12.1, relativo a la conservación de registros; artículo 15.1 relativo a la precisión robustez y ciberseguridad; Artículo 40.2 relativo a las normas armoni-

RIA, y de la doctrina, que son exponentes de la realidad técnica, compleja, subyacente en los Sistemas de Inteligencia Artificial.

Sobre el *ciclo de vida* de los SIA, el Consejo aportó una propuesta de definición —no aprobada—, referida a su duración desde diseño hasta su retirada o hasta su modificación sustancial (*propuesta de art. 3.1a. del Consejo*)¹⁰. Destaca en este punto el Considerando (69), que impone la obligación de garantizar el derecho a la privacidad y la protección de los datos personales en todo el ciclo de vida completo del sistema de IA, proponiendo incluso medidas técnicas y organizativas con este cometido específico¹¹.

En segundo lugar, hablar de la *cadena de valor* en los SIA nos obliga a identificar sus diferentes agentes, operadores, que intervienen en su ciclo de vida y las interacciones que se producen entre ellos. Ello ha de servir de base para garantizar el derecho a la protección de los datos personales, el cumplimiento de obligaciones y la exigencia de responsabilidades conforme al RGPD¹².

-
- zadas y documentos de normalización; y el ANEXO IV. Documentación técnica a que se refiere el artículo 11, apartado 1.
10. Otras definiciones en: AI HLEG (2020), p. 34: “Ciclo de vida: El ciclo de vida de un sistema de IA incluye varias fases interdependientes que van desde su diseño y desarrollo (incluyendo subfases como análisis de requisitos, recopilación de datos, capacitación, pruebas, integración), instalación, implementación, operación, mantenimiento y eliminación. Dada la complejidad de los sistemas de IA (y en general de la información), se han definido varios modelos y metodologías para gestionar esta complejidad, especialmente durante las fases de diseño y desarrollo, como cascada, espiral, desarrollo de software ágil, prototipado rápido e incremental.”—; AI HLEG (2019). Ap. 147: “El ciclo de vida de un sistema de IA abarca las fases de desarrollo (incluidas las tareas de investigación, diseño, provisión de datos y realización de ensayos limitados), despliegue (incluida la aplicación) y utilización de dicho sistema.”; Lazcoz y Hert, (2023), p. 8: “Al observar las definiciones del artículo 3, aprendemos que la fase de desarrollo y la fase de uso son las dos fases o etapas principales del ciclo de vida de la IA, cuyos participantes clave son los proveedores y los responsables del despliegue, respectivamente.”; y por último, CONSEJO DE EUROPA. (2023), en su propuesta de artículo 10 incluye en el ciclo de vida del SIA su desmantelamiento (en el mismo sentido AEPD.2020).
 11. “... A este respecto, los principios de minimización de datos y de protección de datos desde el diseño y por defecto, establecidos en el Derecho de la Unión en materia de protección de datos, son aplicables cuando se tratan datos personales. Las medidas adoptadas por los proveedores para garantizar el cumplimiento de estos principios podrán incluir no solo la anonimización y el cifrado, sino también el uso de una tecnología que permita llevar los algoritmos a los datos y el entrenamiento de los sistemas de IA sin que sea necesaria la transmisión entre las partes ni la copia de los datos brutos o estructurados, sin perjuicio de los requisitos en materia de gobernanza de datos establecidos en el presente Reglamento...”. Ver también AI HLEG. “Directrices éticas para una Inteligencia artificial fiable.”, abril de 2019.
 12. Conviene recordar en este punto que el Considerando 79 RGPD se puede relacionar con el Considerando (83) RIA: “Teniendo en cuenta la naturaleza y la complejidad de la cadena de valor de los sistemas de IA y en consonancia con los principios del nuevo marco legislativo, es esencial garantizar la seguridad jurídica y facilitar el cumplimiento del presente Reglamento. Por lo tanto, es necesario aclarar el papel y las obligaciones específicas de los operadores pertinentes a lo largo de la cadena de valor, como los importadores y distribuidores que pueden contribuir al desarrollo de sistemas de IA. En determinadas situaciones, estos operadores podrían desempeñar más de una función al mismo tiempo y, por lo tanto, deben cumplir acumulativamente todas las obligaciones pertinentes asociadas a dichas funciones. Por ejemplo, un operador podría actuar como distribuidor y importador al mismo tiempo.

Desde este punto de vista, el RIA identifica como agentes o intervinientes a lo largo del ciclo de vida y en la cadena de valor, entre otros, al *proveedor* (art. 3.3 RIA), al *responsable del despliegue* (art. 3.4 RIA), al *representante autorizado* (art. 3.5 RIA), al *importador* (art. 3.6 RIA), al *distribuidor* (art. 3.7 RIA), *proveedores de modelos y SIAs de uso general* (Considerandos (97) y (101), entre otros muchos, y arts. 53 y siguientes RIA), *suministradores de sistemas, herramientas, servicios de IA, componentes o procesos incorporados por el proveedor al SIA*, para entre, otros objetivos, el entrenamiento, el reentrenamiento, la prueba y evaluación de modelos, la integración en el software u otros aspectos del desarrollo de aquellos (Considerandos (88) y (90) RIA y art. 25 RIA). Además, el artículo 25 RIA, relativo a las «*responsabilidades a lo largo de la cadena de valor de la IA*» se refiere también a aquellos que pongan en el SIA su nombre o una marca comercial, quienes los modifiquen sustancialmente, o quienes modifiquen la finalidad prevista, entre otros.

En relación con estas interacciones y agentes se han descrito diferentes modelos de cadena de valor, a modo de ejemplo: desarrollo o implementación de un SIA interno, coincidiendo proveedor y usuario; desarrollo a medida de un SIA para otra entidad; una entidad escribe el código y entrena el sistema, y lo comercializa a través de un acceso restringido al SIA, de modo que el usuario no puede hacer cambios, solo enviar datos de entrada y recibir resultados; una entidad vende modelos pre-entrenados y la entidad que adquiere el modelo incorpora datos de entrenamiento; proveedor vende un SIA actualizable cuando los responsables del despliegue introducen nuevos datos; un desarrollador de un SIA lo vende a otro desarrollador SIA, para continuar entrenando, para mejorarlo, o para adaptarlo tareas más específicas —trabajan por tanto diferentes conjuntos de datos—; una entidad integra diferentes SIAs (p.e. el SIA decide a que SIA se derivan los datos de entradas)¹³.

Este análisis del SIA, en un contexto de cadena de valor, ya había sido considerado por la AEPD, en la identificación del tratamiento de datos personales, por referencia a componentes IA que son incluidos o utilizados, e integrados a su vez por otros componentes referidos a la recogida de datos, sistemas de archivos, módulos de seguridad, interfaces de usuario, entre otros¹⁴.

3. TRATAMIENTO DE DATOS PERSONALES Y SIAS

Como premisa debemos considerar, que «*el desarrollo y uso de sistemas de IA implicará en muchos casos el tratamiento de datos personales*»¹⁵, siendo precisamente este el planteamiento del RIA.

El tratamiento de datos personales en el ciclo de vida de un SIA puede producirse en distintos momentos y funcionalidades. Efectivamente, como supuesto de hecho que debemos considerar, se prevén en la RIA diferentes operaciones con finalidades inmediatas definidas, sin perjuicio de poder integrarse, junto con otras operaciones, en un tratamiento de datos más complejo. Estas operaciones pueden tener lugar durante el diseño, el desarrollo o el uso de

13. Engler, A. C., & Renda, A. (2022).

14. AEPD, 2020, p. 12.

15. EDPS y EDPB, 2021, p. 14 —pár. 15—.

sistemas de IA¹⁶. Así, podemos hacer referencia a las operaciones de tratamiento realizadas sobre los denominados *datos de entrenamiento* (Art. 3.29 RIA)¹⁷, los *datos de validación* (art. 3.30 RIA)¹⁸, los *datos de prueba* (art. 3.32 RIA)¹⁹, y los *datos de entrada* (art. 3.33 RIA)²⁰ proporcionados al SIA o adquiridos por este, a partir de los cuales el SIA produce la información de salida. Incluso se ha llegado a describir la existencia de datos personales en el propio algoritmo, como consecuencia de la posibilidad técnica de recuperar aquellos datos utilizados en el entrenamiento, y avanzar en la identificación de los interesados²¹.

También podemos hacer referencia a la operación de tratamiento a la que el SIA sirve de instrumento, entendiendo que pueden coincidir la finalidad del SIA, referida a la información de salida (art. 3.1 RIA) como contenidos, predicciones, recomendaciones, e incluso decisiones, con la finalidad del tratamiento, ya sea como una operación o un conjunto de operaciones de tratamiento de datos personales.

Son abundantes las referencias a los datos personales en el RIA e incluso a tratamientos, definidos por referencia al objetivo perseguido —explícito o implícito—, que no necesariamente ha de coincidir con la finalidad a los efectos de su licitud conforme al RGPD. Se habla de *datos biométricos* (Art. 3.34) RIA, *identificación y verificación biométrica* (artículo 3, apartado 35 y 36 RIA); *datos basados en biometría; categorías especiales de datos personales; ultrafalsificación; datos resultado de perfilado; datos de puntuación social; datos de comportamiento social; emociones...*²². Del mismo modo, podemos considerar a su vez el SIA constitutivo de una operación de tratamiento que se integra en un conjunto de operaciones de tratamiento, con una finalidad superior, y ello aun cuando se integre en un tratamiento con otras fases u operaciones que no utilicen herramientas tecnológicas como SIA.

En todos estos supuestos, el punto de conexión con el marco jurídico de protección es la existencia de datos personales y su tratamiento, incluso integrados en un conjunto de datos no personales²³. Por otro lado, si bien es cierto que los datos anónimos, o anonimizados, no deben considerarse datos personales a los efectos del RGPD²⁴, debemos recordar, por un lado, que la anonimización de datos personales

16. Considerando (10) RIA, párrafo segundo.

17. «los datos utilizados para entrenar un sistema de IA mediante el ajuste de sus parámetros entrenables».

18. «los datos usados para proporcionar una evaluación del sistema de IA entrenado y adaptar sus parámetros no entrenables y su proceso de aprendizaje para, entre otras cosas, evitar el ajuste insuficiente o el sobreajuste».

19. «los datos usados para proporcionar una evaluación independiente del sistema de IA, con el fin de confirmar el funcionamiento previsto de dicho sistema antes de su introducción en el mercado o su puesta en servicio».

20. «datos proporcionados a un sistema de IA o adquiridos directamente por éste, a partir de los cuales el sistema produce la información de salida».

21. Sobre este riesgo se recomienda lectura de Veale, M., Binns, R., & Edwards, L. (2018). También Considerando (76).

22. Considerando (18) RIA.

23. Considerando 10 RIA.

24. Apartados 55 y siguientes de la Sentencia TJUE (Gran Sala) de 5 de diciembre de 2023, Cuestión prejudicial. Deutsche Wohnen SE y Staatsanwaltschaft Berlin Asunto C-683/21 (JUR 2023\432912).

es una operación de tratamiento, con sus propios riesgos, sujeta al mismo, y, por otro, que los avances tecnológicos también tienen impacto en las posibilidades de reidentificación, lo cual ha de ser tenido en cuenta²⁵.

No tenemos espacio para analizar los datos personales citados en el RIA, en cualquiera de sus funcionalidades en el ciclo de vida o cadena de valor del SIA, o incluso los tratamientos que en fase de desarrollo o en explotación —en su caso como finalidad misma del SIA— se producen²⁶. Si debemos recordar que una de las características de los SIAs, reconocida así por la RIA, es su capacidad de *inferir*²⁷, de modo que considerando datos personales²⁸, o incluso anónimos —para el entrenamiento del algoritmo en este momento— y a la vista de concreta de unos datos de entrada —aportado u obtenidos por el sistema— se infieren como resultado nuevos datos personales.

Como se observa, hablamos de un tratamiento de datos personales respecto del cual el RGPD tiene especial cautela cual es la *elaboración de perfiles*²⁹. Y ello es así en la medida en que el resultado de un perfilado —en el sentido descrito—,

-
25. CONSEJO DE EUROPA (2010). «99. El texto desea responder a la objeción planteada de que la recomendación va más allá del ámbito de aplicación del Convenio N.º 108 en la medida en que abarca o podría abarcar, al menos en las etapas 1 y 2, el tratamiento de datos no personales, a saber, datos anónimos. Como se explicó en la introducción, en relación con esta objeción, se pretendía que esta recomendación abarcara, aunque solo fuera incidentalmente, la recopilación y el tratamiento de datos anónimos en la medida en que el tratamiento de estos datos anónimos en la primera y segunda fase puede ser crucial para determinar la legitimidad y la seguridad del tratamiento en la tercera etapa, y que las tres etapas en realidad constituyen un proceso continuo. Así, por ejemplo, parece innecesario exigir a los responsables del tratamiento que utilicen datos anónimos exactos, auténticos y actualizados durante la primera fase de almacenamiento de datos, especialmente porque, a primera vista y en principio, el Convenio N.º 108 no abarca los datos anónimos. De hecho, la sustancia real de estos datos anónimos puede, en cierta medida, como resultado de la elaboración de perfiles, encontrarse, posteriormente e inesperadamente, en el perfil de una persona identificada o identificable».
26. Es conocido el debate en torno a la aplicación de la prohibición de tratamientos que se contienen en el artículo 9.1 RGPD, relativo a los datos biométricos, cuando «esté dirigido a identificar de manera unívoca a una persona física». Efectivamente, se discutía el alcance la prohibición del artículo 9.1 RGPD, según se tratase de la verificación —autenticación de la identidad de una persona («uno – uno») —Art. 3.36 RIA—, y la identificación («uno – varios») —art. 3.35 RIA. Con carácter general las autoridades de control consideran que en ambos casos nos encontramos con tratamientos sujetos a la prohibición, y sus excepciones, del artículo 9 RGPD —Directrices 5/2022 EDPS, en su apartado 12 y AEPD (Nov 2023) apartado IV.A—. Señalado lo anterior, el Considerando (17) RIA con relación a la verificación— autenticación les atribuye una probable repercusión menor en los derechos fundamentales de las personas físicas que los sistemas de identificación biométrica remota que puedan utilizarse para el tratamiento de los datos biométricos de un gran número de personas sin su participación activa, a los efectos de su exclusión de los SIA prohibidos y los SIA de AR, Artículo 5.1.h) y 2 RIA. A esta afirmación y régimen jurídico no debe anudarse una revisión de los criterios indicados.
27. A modo de ejemplo, Considerandos (12), (30), (31), entre otros muchos.
28. Tiene importancia su alta disponibilidad en términos de volumen, variedad y velocidad ICO UK (2019) p. 6.
29. Un análisis más completo sobre sus riesgos específicos (Recomendación 2010) Párrafos 49.2 y siguientes.

cualquiera que sea la tecnología empleada³⁰, incluida la IA, no deja de ser un dato personal en sí mismo, respecto del cual ha de garantizarse el pleno respeto del RGPD, en todos sus principios y normas³¹. En cualquier caso, el RIA considera que la elaboración de perfiles en el contexto de un SIA, más allá de los riesgos que le son inherentes³², puede determinar un riesgo significativo y para la salud, la seguridad o los derechos fundamentales de las personas físicas, siendo elemento determinante de su prohibición (Art. 5.1.c) y d) RIA), de su consideración de SIA de AR (Anexo III RIA³³) e incluso de la cautela a que se refiere el artículo 6.3, último párrafo, esto es, que «(n)o obstante lo dispuesto en el párrafo primero, los sistemas de IA a que se refiere el anexo III siempre se considerarán de alto riesgo cuando el sistema de IA lleve a cabo la elaboración de perfiles de personas físicas...». Y ello aun cuando no sean soporte o condición de la decisión basada únicamente en el tratamiento automatizado en el sentido del Art. 22 RGPD, en cualquiera de las interpretaciones que hubiera de darse a la garantía de la intervención humana, pues el impacto sobre la persona es indudable³⁴. Sobre las decisiones automatizadas en el marco del artículo 22 RGPD en relación con el RIA se realizarán observaciones en el Apartado III.6 del presente documento.

4. IDENTIFICACIÓN DE RESPONSABLES DE TRATAMIENTO DE DATOS PERSONALES EN LOS SIAS

Uno de los elementos estructurales en torno a la protección de los datos personales, como derecho fundamental, es la adecuada identificación de todos aquellos que participan en el tratamiento de datos personales, a los efectos del cumplimiento de las prescripciones legales contenidas en el RGPD, con el alcance y extensión que, de modo indicativo, se expondrá en el apartado III.2.

En este punto tiene una posición nuclear la identificación del responsable de tratamiento («solo o junto con otros»)³⁵. Para ello, podemos inicialmente identificar dos

30. Puede verse las normas contenidas al efecto en la DSA, particularmente el artículo 28.2 relativo a la protección de menores en línea, impidiendo la presentación a menores de anuncios de interfaz basados en elaboración de perfiles. De interés también, su Considerando (94).

31. Considerando (10). *También conviene aclarar que los interesados siguen disfrutando de todos los derechos y garantías que les confiere dicho Derecho de la Unión, incluidos los derechos relacionados con la toma de decisiones individuales exclusivamente automatizada, incluida la elaboración de perfiles.*

32. «Considerando que la falta de transparencia, o incluso la “invisibilidad”, de la elaboración de perfiles y la falta de precisión que puede derivarse de la aplicación automática de normas de inferencia preestablecidas pueden suponer riesgos significativos para los derechos y libertades del individual.» [CONSEJO DE EUROPA (2010) p. 5].

33. Considerando (53). «En cualquier caso, debe considerarse que los sistemas de IA a que se refiere el anexo III plantean riesgos significativos de perjuicio para la salud, la seguridad o los derechos fundamentales de las personas físicas si el sistema de IA implica la elaboración de perfiles en el sentido del artículo 4, apartado 4, del Reglamento (UE) 2016/679 y del artículo 3, apartado 4, de la Directiva (UE) 2016/680 y del artículo 3, apartado 5, del Reglamento 2018/1725.»

34. Aun cuando se cumplan las condiciones señaladas en el Considerando (53), para que se entienda que el SIA no influye sustancialmente en el resultado de la toma de decisiones.

35. «El principio consagrado en el artículo 5, apartado 2, del RGPD es, en nuestra opinión, el principio más pertinente del RGPD, ya que está íntimamente vinculado a los otros seis prin-

cuestiones relevantes, objeto de regulación en el RGPD, y que, para la resolución de esta cuestión, han sido ampliamente analizada por la Jurisprudencia comunitaria.

Así, por un lado, podemos hacer referencia a la indicación contenida en el artículo 4.2 RGPD en la definición de *tratamiento*, que comprende las «operaciones o conjunto de operaciones realizadas sobre datos personales o conjuntos de datos personales». Nos interesa en este punto la actuación realizada sobre datos personales que pudiera considerarse compleja, integradas por diferentes fases o etapas todas ellas constitutivas de un tratamiento de datos personales³⁶. Este contexto debe completarse con la definición de *responsable de tratamiento* que se contiene en el artículo 4.7 RGPD referido a quien determine los fines y medios del tratamiento, así como, con este mismo alcance introductorio, el *encargado de tratamiento*, quien trata *datos personales por cuenta del responsable del tratamiento* (art. 4.8 RGPD).

A partir de aquí, en nuestra labor de analizar los tratamientos de datos personales que se producen en el contexto de los SIAs, la realidad que debemos considerar está determinada por un lado por su *ciclo de vida* y por su *cadena de valor* (Apartado II.2) y por otro por la necesidad, también dificultad, de establecer un marco de seguridad jurídica. Esto ha sido puesto de manifiesto por diversos autores³⁷ y durante el procedimiento normativo, especialmente por el SEPD y CEPD, en su informe conjunto³⁸.

De acuerdo con lo anterior, el RGPD establece normas referidas a las interacciones entre responsables, corresponsables y encargados, las cuales pueden servir al cometido que nos ocupa, pero que pudieran resultar insuficientes:

- El artículo 26 RGPD, relativo al reparto de responsabilidades entre los corresponsables del tratamiento, les impone la obligación de establecer de manera transparente, mediante acuerdo —cuya esencia se pondrá a disposición de los interesados— sus respectivas responsabilidades, funciones

... cipios del RGPD (...) La responsabilidad exige que los controladores tomen la responsabilidad de lo que hacen con los datos personales, para cumplir con todos los demás principios del RGPD y para demostrar este cumplimiento.» (Lazcoz y Hert, 2023, p. 20).

36. En este punto puede analizarse el apartado 72 de la Sentencia TJUE (Sala Segunda) de 29 de julio de 2019, Fashion ID en el asunto C-40/17 (TJCE 2019\148): «De esta definición —con el mismo contenido en la Directiva— resulta que un tratamiento de datos personales puede estar constituido por una o varias operaciones, cada una de ellas referida a una de las distintas fases que puede contener un tratamiento de datos personales».
37. «Para que los mecanismos de gobernanza y rendición de cuentas responsabilicen a los responsables del desarrollo, el despliegue y el uso de tecnologías de IA para rendir cuentas de su funcionamiento y sus efectos, debe abordarse urgentemente la dinámica de las cadenas de suministro.» (Cobbe et al., 2023, p. 1197).
38. «6. El CEPD y el SEPD acogen con satisfacción la participación en la regulación de todas las partes interesadas de la cadena de valor de la IA y la introducción de requisitos específicos para los proveedores de soluciones, ya que desempeñan un papel importante en los productos que utilizan sus sistemas. Sin embargo, las responsabilidades de las distintas partes (usuario, proveedor, importador o distribuidor de un sistema de IA) deben estar claramente circunscritas y asignadas. En particular, al tratar datos personales, debe prestarse especial atención a la coherencia de estas funciones y responsabilidades con las nociones de responsable del tratamiento y encargado del tratamiento que lleva a cabo el marco de protección de datos, ya que ambas normas no son congruentes». (EDPS y EDPB, 2021, p. 10).

- y relaciones, en el cumplimiento del RGPD, en particular, con excepciones, en lo que respecta al ejercicio de los derechos del interesado y sus respectivas obligaciones de facilitar la información, pudiendo designar un punto de contacto para los interesados.
- En cualquier caso, el interesado podrá ejercer sus derechos conforme al RGPD, «*con respecto a y contra cada uno de los responsables del tratamiento*». Parece, en definitiva, que cualquiera que sea la distribución de las funciones que se hayan asignado los corresponsables de tratamiento mediante acuerdo, ello no vincula al interesado que podrá ejercer sus derechos (ex RGPD) ante cualquiera de ellos³⁹
 - El artículo 17, apartado 2, RGPD, en relación con el «*derecho de supresión (“derecho al olvido”)*» se refiere al mismo en un entorno de responsabilidad en el tratamiento en términos de uso secundario de datos personales⁴⁰ exigiendo la adopción de *medidas razonables*, para informar a otros responsables⁴¹.
 - Por tanto, en un entorno *en línea* el RGPD impone al responsable, en caso de ejercicio del derecho, la obligación de informar e indicar a los demás que supriman todo «*enlace a ellos, o las copias o réplicas de tales datos*», debiendo tomar medidas razonables, teniendo en cuenta la tecnología y los medios a su disposición, incluidas las medidas técnicas (Considerando 66 RGPD). En desarrollo y aplicación del artículo 17.2 RGPD el Artículo 70.d) RGPD atribuye al CEPD la función de garantizar la aplicación coherente del precepto.
 - El artículo 82 («*Derecho a indemnización y responsabilidad*»), apartados 4 y 5, RGPD, se refiere al supuesto de corresponsabilidad, previendo, en caso de ser responsables de daños y perjuicios causados por el tratamiento, que cada uno de los corresponsables será considerado responsable de todos ellos para «*garantizar la indemnización efectiva del interesado*», sin perjuicio del derecho de repetición sobre el resto, en su caso conforme al art. 82.2 RGPD⁴². No se contiene norma similar en el artículo 83 RGPD («*Condiciones generales para la imposición de multas administrativas*»).

39. Mahieu et al., 2018, p. 52.

40. Brown, 2023, p. 36.

41. «2. Cuando haya hecho públicos los datos personales y esté obligado, en virtud de lo dispuesto en el apartado 1, a suprimir dichos datos, el responsable del tratamiento, teniendo en cuenta la tecnología disponible y el coste de su aplicación, adoptará medidas razonables, incluidas medidas técnicas, con miras a informar a los responsables que estén tratando los datos personales de la solicitud del interesado de supresión de cualquier enlace a esos datos personales, o cualquier copia o réplica de los mismos.»

42. No se pretende con esta cita del art. 82 RGPD entrar en el debate sobre responsabilidad civil en el contexto de la Inteligencia Artificial, y su relación con el marco de responsabilidad civil de los corresponsables de tratamiento de datos personales. Sobre la responsabilidad civil e IA de interés el documento «INFORME DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO Y AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO, sobre las repercusiones en materia de seguridad y responsabilidad civil de la inteligencia artificial, el internet de las cosas y la robótica», de 19 de febrero de 2020 (COM (2020) 64 final), y la propuesta de DIRECTIVA DEL PARLAMENTO EUROPEO Y DEL CONSEJO relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial (Directiva sobre responsabilidad en materia de IA (COM/2022/496 final).

Frente a ello se han descrito como dificultades, en el contexto de los SIAs:

- *En muchos casos, ningún actor tendrá suficiente conocimiento o control sobre la producción y el despliegue para poder evaluar o mitigar de manera fiable los impactos y riesgos*⁴³; o incluso, si consideramos que el derecho de acceso incluye el derecho a conocer la identidad del destinatario de los datos personales⁴⁴ permitiendo conocer el flujo de datos, con las finalidades previstas en el RGPD, este derecho se limitaría a categorías de destinatarios en función del contexto de la cadena de valor.
- La asunción de responsabilidades en cadenas de valor complejas puede dificultarse como consecuencia de su naturaleza transfronteriza, a pesar de las previsiones del RGPD⁴⁵.
- La falta de estandarización o de especificaciones comunes, la interacción entre los distintos componentes, las interacciones ocultas de datos y el desequilibrio en la relación entre los distintos implicados en la cadena de valor.
- Las incertidumbres en la asignación de responsabilidades en función de las etapas en las operaciones de tratamiento, y diversos grados de responsabilidad, en el caso de corresponsables de tratamiento, en los que no se ha establecido un marco de colaboración y coordinación entre ellos⁴⁶.
- La eventual falta de documentación relativa a los agentes e interacciones en el ciclo de vida y cadena de valor puede dar lugar a dificultades para verificar el cumplimiento del RGPD en las diferentes operaciones de tratamiento de datos personales, aun cuando se encuentren conectadas, o formen parte de un conjunto⁴⁷.
- Se dificulta la obtención del consentimiento de los interesados conforme al RGPD, por falta de interacción o de conocimiento de estos⁴⁸.

Siendo estas las dificultades, entre otras, no puede renunciarse al cumplimiento del RGPD, debiendo atender a «una visión amplia de las cadenas de suministro, buscando identificar todos los operadores y agentes, cuál es su función e intervención, y cómo asignar responsabilidades entre ellas⁴⁹». Así, los responsables del despliegue de SIAs deben desplegar una especial diligencia, de manera que adopten las medidas

43. Cobbe et al., 2023, p. 1195.

44. Sentencia TJUE (Sala Primera) de 12 de enero de 2023. Österreichische Post. En el asunto C-154/21, (TJCE 2023\4) (ECLI:EU:C:2023:3).

45. «*Si bien algunas leyes, como la ley de protección de datos de la UE y la Ley de IA y la Ley de privacidad del consumidor de California han buscado un efecto extraterritorial para abordar el arbitraje regulatorio, la naturaleza transfronteriza de las cadenas de suministro y las dificultades de ejecución siguen siendo un desafío significativo de rendición de cuentas.*» (Cobbe et al., 2023, p. 1196).

46. Mahieu et al., 2018, pp. 50 y 51.

47. Brown, 2023, p. 19.

48. Brown, 2023, p. 19.

49. Cobbe et al., 2023, p. 1197. En el mismo sentido EDPS y EDPB, 2021, p. 10: «*las responsabilidades de las distintas partes (usuario, proveedor, importador o distribuidor de un sistema de IA) deben estar claramente circunscritas y asignadas. En particular, al tratar datos personales, debe prestarse especial atención a la coherencia de estas funciones y responsabilidades con las nociones de responsable del tratamiento y encargado del tratamiento que lleva a cabo el marco de protección de datos, ya que ambas normas no son congruentes.*»

necesarias, en su caso contractuales, para asegurar el cumplimiento del RGPD en el contexto del ciclo de vida y la cadena de valor, enervando por tanto el riesgo que supone la integración en mismo de *servicios* no conformes con la norma, no sujetos a rendición de cuentas, o carentes de documentación necesaria⁵⁰, pudiendo determinar incluso la no integración del servicio.

A esta solución se aproxima la obligación de declarar el cumplimiento del RGPD, como contenido de la evaluación de conformidad (ANEXO V («Declaración UE de Conformidad»)), apartado 5⁵¹, RIA) cuando el SIA a que se refiere suponga el tratamiento de datos personales.

Debemos recordar en este punto que la Jurisprudencia de TJUE y las autoridades de control, han completado el marco de identificación de los responsables de tratamiento en entornos complejos, de acuerdo con las siguientes pautas generales:

- Se ha de partir de una interpretación amplia en la identificación del responsable de tratamiento, en la medida que el Derecho de la Unión exige la protección efectiva y completa de los interesados en los términos del RGPD⁵².
- Se ha de analizar caso a caso, desde un punto de vista fáctico, la determinación de los fines y medios de tratamiento de los datos personales, en su conjunto y cada una de las operaciones individuales de tratamiento. Se asienta en la influencia en el tratamiento de datos personales participando así en la determinación de sus fines y medios pueden ser considerada responsable del tratamiento⁵³.
- Cuando se distinguen varias fases del tratamiento o varios conjuntos de operaciones de tratamiento no se está haciendo referencia a las distintas actividades de ejecución material del tratamiento sino a la existencia de fases del tratamiento con diferente diseño —qué datos, con qué fines y en virtud de qué medios—⁵⁴.
- La corresponsabilidad no supone necesariamente que, con respecto a un mismo tratamiento de datos personales, los diversos agentes tengan una responsabilidad equivalente, sino que pueden estar implicados en distintas etapas del tratamiento y en distintos grados. Los agentes en este caso únicamente pueden ser responsables, conjuntamente con otros, de las

50. En este sentido Mahieu et al., 2018, pp. 50 y 51.

51. «La declaración UE de conformidad a que se hace referencia en el artículo 47 contendrá toda la información siguiente:... 5. Cuando un sistema de IA conlleve el tratamiento de datos personales, la declaración de que el sistema de IA se ajusta al Reglamento (UE) 2016/679, al Reglamento (UE) 2018/1725 y a la Directiva (UE) 2016/680.»

52. Apartados 34 y 66 Sentencia TJUE (Gran Sala) de 13 de mayo de 2014, Google Spain S.L.y AEPD en el asunto C-131/12, (TJCE 2014\85), si bien referido al artículo 2.d) de la Directiva; apartados 26, 27 y 28 Sentencia TJUE (Gran Sala) de 5 de junio de 2018, Wirtschaftsakademie Schleswig-Holstein GmbH, asunto C-210/16, (TJCE 2018\120) (ECLI:EU:C:2018:388); apartados 65 y 66 Sentencia Fashion ID (TJCE 2019\148).

53. Apartado 68 Sentencia TJUE Google Spain S.L.y AEPD (TJCE 2014\85) citada; y apartado 68 de la Sentencia TJUE (Gran Sala) de 10 de julio de 2018, Jehovan todistajat, en asunto C-25/17. (TJCE 2021\63) (ECLI:EU:C:2018:551).

54. Sentencia TJUE, Jehovan todistajat (TJCE 2021\63) citada, apartado 71. También de interés Sentencia del Tribunal Constitucional español 42/2022, de 21 de marzo de 2022. Recurso de amparo 4011-2020 (RTC 2022\42).

operaciones de tratamiento cuyos fines y medios determine conjuntamente. No podrán ser considerados responsable de las operaciones anteriores o posteriores de la cadena de tratamiento respecto de las que no determine los fines ni los medios⁵⁵. Así, el control ejercido por una entidad determinada puede extenderse a la totalidad del tratamiento en cuestión, o a una etapa particular del tratamiento.

- La corresponsabilidad no supone que cada uno de los corresponsables tenga acceso a los datos personales en cuestión, si externaliza la actividad de tratamiento teniendo una influencia determinante en el propósito y los medios esenciales⁵⁶.

III. LA RELACIÓN DE REGLAMENTO DE INTELIGENCIA ARTIFICIAL Y EL REGLAMENTO GENERAL DE PROTECCIÓN DE DATOS

1. NECESIDAD DE UN MARCO DE RELACIÓN ENTRE AMBOS CUERPOS NORMATIVOS

Si nos centramos, como hemos anticipado, en la relación entra ambos marcos normativos, debemos destacar el objeto y el objetivo diverso de una y otra regulación.

Efectivamente la RIA se refiere a la introducción en el mercado, puesta en servicio y utilización de SIA, en la UE, estableciendo al efecto, prohibiciones de determinadas prácticas de IA, estableciendo requisitos específicos y obligaciones en el caso de SIA de AR, normas de transparencia para determinados SIA, normas armonizadas para introducción en el mercado de determinados modelos de IA de uso general y normas sobre la vigilancia del mercado, gobernanza y la ejecución (Artículo 1 RIA). Todo ello en su ámbito de aplicación (Art. 2 RIA).

Por su parte, el RGPD se refiere a la protección de las personas físicas, sus derechos y libertades, en lo que respecta al tratamiento de los datos personales, y su libre circulación (Art. 1 RGPD, sin perjuicio de su concreto ámbito material (art. 2 RGPD) en todo su ámbito de aplicación territorial (Art. 3 RGPD).

Debemos recordar en cualquier caso que el RGPD, si bien referido al tratamiento de datos personales, a su protección como derecho fundamental, realmente se aplica al servicio de la persona, su libertad, su dignidad, y sus derechos fundamentales, particularmente en aquellos casos en los que, como consecuencias de la elaboración de perfiles y decisiones automatizadas, su desarrollo, su posibilidad de

55. Apartados 70 y 74 Sentencia TJUE Fashion ID (TJCE 2019\148) citada, con aceptación de la conclusión 101 del Abogado General; apartado 66 Sentencia TJUE, Jehovan todistajat (TJCE 2021\63) citada. En el mismo sentido la reciente Sentencia TJUE (Sala Cuarta), de 7 de marzo de 2024, IAB Europe y Gegevensbeschermingsautoriteit en el asunto C-604/22. (JUR 2024\73167), apartado 78.2 (parte dispositiva).

56. Apartado 38 de la Sentencia TJUE Wirtschaftsakademie Schleswig-Holstein GmbH (TJCE 2018\120) citada; apartado 69 de la Sentencia TJUE, Jehovan todistajat (TJCE 2021\63) citada; apartado 69 de la Sentencia TJUE Fashion ID (TJCE 2019\148) citada; y apartado 78 (parte dispositiva) de la Sentencia TJUE IAB Europe y Gegevensbeschermingsautoriteit (JUR 2024\73167) citada.

autodeterminación o de elección se encuentra condicionado, en muchos casos, sobre la base de una inferencia, de una probabilidad⁵⁷.

Por ello, y en definitiva, desde este punto de vista, el RGPD es de plena aplicación a todo el ciclo de vida del SIA y a toda su cadena de valor, si en su contexto y con cualquier alcance se produce un tratamiento de datos personales o, a partir de ellos, el interesado resulta afectado por decisiones automatizadas⁵⁸, siendo el cumplimiento normativo, también en lo referido al RGPD, uno de los cimientos de una IA ética. Por ello, la relación entre los SIAs, sus elementos constitutivos, y el tratamiento de datos personales, ha sido considerada por la doctrina⁵⁹ y las autoridades de control en los informes emitidos con ocasión del proceso de elaboración de la RIA.

En cualquier caso, han sido autorizadas las voces que, considerando la importancia de los datos personales en los SIAs y para evitar la puesta en peligro, directa o indirectamente, el derecho fundamental a su protección, pidieron en la tramitación del RIA:

— Una relación claramente definida entre la RIA propuesta y la legislación vigente en materia de protección de datos, evitando cualquier incoherencia y posible conflicto, evitando cualquier afección o interferencia en el mismo, incluyendo sobre la competencia de las autoridades de supervisión y la gobernanza.

— Respetar el acervo generado en la interpretación y aplicación de la Directiva 95/46/CE y el RGPD, por el Grupo de Trabajo del Artículo 29, el Comité Europeo de Protección Datos, el Supervisor Europeo de Protección de Datos, las diferentes autoridades de control, y los órganos jurisdiccionales⁶⁰.

El marco jurídico de protección de los datos personales se inserta en el SIA en la medida en que este se sirve o soporta el tratamiento de datos personales. El artículo

57. GPA. 2020. p.3.

58. EDPS 2022, p.8. «*son en muchos casos la premisa clave para las decisiones autónomas que inevitablemente afectarán a la vida de las personas a diversos niveles*». En el mismo sentido EDPS y EDPB, 2021, p. 8.

59. Martínez Martínez, Ricard. «INTELIGENCIA ARTIFICIAL DESDE EL DISEÑO. RETOS Y ESTRATEGIAS PARA EL CUMPLIMIENTO NORMATIVO». *Revista Catalana de Dret Públic*, n.º 58 (2019) pp. 73 y ss.

60. EDPS. 2022. «Dictamen 20/2022 sobre la Recomendación de Decisión del Consejo por la que se autoriza la apertura de negociaciones en nombre de la Unión Europea para un Convenio del Consejo de Europa sobre inteligencia artificial, derechos humanos, democracia y Estado de Derecho»; EDPS, y EDPB. 2021. p. 14 y p. 24 —Apartado 3, párrafo 56 y 57—: «15. ...*Es de suma importancia garantizar la claridad de la relación de la presente propuesta con la legislación vigente de la UE en materia de protección de datos. La propuesta no prejuzga y complementa el RGPD, el RPDUE y el LED. Si bien los considerandos de la propuesta aclaran que el uso de sistemas de IA debe seguir cumpliendo la legislación en materia de protección de datos, el CEPD y el SEPD recomiendan encarecidamente aclarar en el artículo 1 de la propuesta que la legislación de la Unión para la protección de los datos personales, en particular el RGPD, el RPDUE, la Directiva 10 sobre la privacidad y las comunicaciones electrónicas y el LED, se aplicarán a todo tratamiento de datos personales que entre en el ámbito de aplicación de la propuesta. Un considerando correspondiente debe aclarar igualmente que la propuesta no pretende afectar a la aplicación de la legislación vigente de la UE que regula el tratamiento de datos personales, incluidas las funciones y competencias de las autoridades de control independientes competentes para supervisar el cumplimiento de dichos instrumentos.*»

8 CDFUE precisa que los datos personales se tratarán de modo leal, para fines concretos y sobre la base del consentimiento de la persona afectada o en virtud de otro fundamento legítimo previsto por la ley, que toda persona tiene derecho a acceder a los datos recogidos que le conciernan y a obtener su rectificación y que el respeto de estas normas estará sujeto al control de una autoridad independiente. Aplican estos requisitos, en particular, diferentes preceptos del RGPD.

Se impone al tratamiento de datos personales el marco jurídico de su protección, como derecho fundamental, y que tiene como ejes principales:

— El establecimiento de requisitos para el tratamiento de datos personales, que se concreta en los denominados principios de tratamiento (Artículo 5 y siguientes RGPD), tanto para el entrenamiento de modelos IA, como para los datos de entrada a los efectos generar una predicción o inferencia, como ejemplos. Tienen en este punto especial importancia los principios de transparencia, limitación de finalidad, minimización y exactitud de datos, limitación del plazo de conservación o integridad, entre otros.

— El reconocimiento de derechos a los interesados (Artículos 12 y siguientes RGPD), particularmente en relación con las decisiones automatizadas conforme al artículo 22 RGPD, sin restar importancia a los derechos de información y acceso (arts. 12 a 15 RGPD), a los derechos de rectificación y supresión (arts. 16 y 17 RGPD) derechos a la limitación de tratamiento (art. 18 RGPD) y el derecho a la oposición al tratamiento (art. 21 RGPD) todo ellos referido a los datos personales tratados en cualquier momento del ciclo de vida de los SIAs.

— El establecimiento de obligaciones para aquellos que intervienen de una y otra manera en el tratamiento de los datos personales. Entre estas obligaciones destacan la protección de datos desde el diseño y por defecto (art. 25 RGPD), y la Evaluación de Impacto en la Protección de los Datos personales (Art. 35 RGPD).

De especial importancia, se impone también a todos los responsables, no solo respecto de tratamientos en SIAs de AR, la obligación de aplicar, revisar y actualizar medidas técnicas y organizativas apropiadas para garantizar y poder demostrar que el tratamiento es conforme con el RGPD, teniendo en cuenta la naturaleza, el ámbito, el contexto y los fines del tratamiento así como los riesgos de diversa probabilidad y gravedad para los derechos y libertades de las personas físicas (Art. 24.1 RGPD).

Del mismo modo, también como obligación de todos los responsables, la notificación y gestión de las brechas de seguridad de los datos personales a que se refieren los artículos 33 y 34 RGPD:

— Como resulta del artículo 5.2 (RGPD) el responsable deberá ser capaz de demostrar el cumplimiento de las obligaciones asumidas en el marco del RGPD («responsabilidad proactiva»).

— El establecimiento de un marco institucional de garantía, destacando en este sentido la intervención de las *autoridades independientes de control en materia de protección de datos*, en un entorno de certeza y seguridad jurídica.

En este marco de garantía del derecho fundamental es una pieza clave la identificación del responsable de tratamiento de datos personales, incluso en los casos en los que determine los fines y medios del tratamiento de forma conjunta con otros responsables, o en los que el tratamiento se lleve a cabo por cuenta de un responsable, esto es,

interviniendo un encargado, en su nombre. Se ha considerado así indispensable para la protección de los derechos y libertades de los interesados, para la atribución clara de responsabilidades, también para la supervisión por parte de las autoridades de control⁶¹.

2. APLICACIÓN DEL REGLAMENTO SIN PERJUICIO DE LA APLICACIÓN DEL RGPD

Un presupuesto básico considerado por el RIA es la plena aplicación del RGPD al SIA, exclusivamente y en la medida que suponga o determine un tratamiento de datos personales, o una decisión automatizada, incluido en su ámbito de aplicación. Desde este punto de vista, es coherente y puede analizarse el Considerando 15 RGPD según el cual *«A fin de evitar que haya un grave riesgo de elusión, la protección de las personas físicas debe ser tecnológicamente neutra y no debe depender de las técnicas utilizadas»*.

Este enunciado general también ha sido previsto en el RIA. De este modo el Considerando (10), nos dice: *«...El presente Reglamento no pretende afectar a la aplicación del Derecho de la Unión vigente que regula el tratamiento de datos personales, incluidas las funciones y competencias de las autoridades de supervisión independientes competentes para vigilar el cumplimiento de dichos instrumentos...»*. El mismo Considerando nos recuerda que el RIA *«no afecta a las obligaciones de los proveedores y los responsables del despliegue de sistemas de IA en su calidad de responsables o encargados del tratamiento de datos derivadas del derecho de la Unión en materia de protección de datos personales en la medida en que el diseño, el desarrollo o el uso de los sistemas de IA impliquen el tratamiento de datos personales»*, y continúa aclarando que *«los interesados siguen disfrutando de todos los derechos y garantías que les confiere dicho Derecho de la Unión, incluidos los derechos relacionados con las decisiones individuales totalmente automatizadas, como la elaboración de perfiles»*, y finaliza, *«(u)nas normas armonizadas para la introducción en el mercado, la puesta en servicio y la utilización de sistemas de IA establecidas en virtud del presente Reglamento deben facilitar la aplicación efectiva y permitir el ejercicio de los derechos y otras vías de recurso de los interesados garantizados por el Derecho de la Unión en materia de protección de datos personales, así como de otros derechos fundamentales.»*

A partir de estas declaraciones, sin perjuicio de diversos Considerandos con el mismo contenido⁶², el Artículo 2, apartado 7, RIA declara: *«El Derecho de la Unión en materia de protección de datos personales, privacidad y confidencialidad de las comunicaciones se aplica a los datos personales tratados en relación con los derechos y obligaciones establecidos en el presente Reglamento. El presente Reglamento no afectará a los Reglamentos (UE) 2016/679 y (UE) 2018/1725 y Directivas 2002/58/CE y (UE) 2016/680, sin perjuicio de los mecanismos previstos en el artículo 10, apartado 5, y en el artículo 59 del presente Reglamento»*. Sobre el sentido y alcance del último inciso, hablaremos en siguientes apartados.

A modo de ejemplo, de manera más precisa, con relación a la *identificación biométrica a distancia para la búsqueda selectiva de una persona condenada o sospechosa de haber cometido un delito* el marco de obligaciones contenidas en el art. 29, apartado

61. Pueden verse en este sentido los Considerandos 13 y 79 RGPD.

62. Considerandos (95), (157) —sobre las autoridades independientes de control en materia de protección de datos personales—.

10, para los responsables del despliegue de SIAs de AR, se entiende «sin perjuicio del artículo 9 del Reglamento (UE) 2016/679 y del artículo 10 de la Directiva (UE) 2016/680 para el tratamiento de los datos biométricos.».

Por último, debemos considerar las llamadas que el RIA realiza a los preceptos del RGPD (de la DEP y del Reglamento (UE) 2016/679) para definir o completar la definición a sus propios efectos, pudiendo destacarse la definición de datos personales y no personales, categorías especiales de datos personales o elaboración de perfiles⁶³.

3. EL REGLAMENTO COMO «LEX SPECIALIS» CON FINES DE POLICÍA

El RIA contiene entre sus disposiciones normas que completan el marco normativo del derecho a la protección de los datos personales como derecho fundamental, y ello se produce precisamente como consecuencia de los retos y desafíos específicos que supone para aquellos la IA. De acuerdo con lo anterior, el RIA ya desde la propuesta de la Comisión, se fundamenta en algunas de sus previsiones en el artículo 16 TFUE y en artículo 8.1. CDFUE, proponiendo desde este punto de vista normas de carácter complementario al RGPD.

Puede leerse, en este sentido, la propuesta inicial de la Comisión Europea que debe entenderse sin perjuicio del RGPD y la DEP, «a los que complementa con un conjunto de normas armonizadas aplicables al diseño, el desarrollo y la utilización de determinados sistemas de IA de alto riesgo y con restricciones de determinados usos de los sistemas de identificación biométrica remota»⁶⁴.

Se refiere en concreto a la prohibición, salvo excepciones, del uso de SIAs para la identificación biométrica remota en tiempo real en espacios de acceso público con fines policiales, para el uso de sistemas de IA para la evaluación del riesgo de las personas físicas a efectos de la aplicación de la ley y para el uso de sistemas de IA de categorización biométrica a efectos de la aplicación de la ley. Se regula así de forma exhaustiva dicho uso y el tratamiento de datos implicado, soportado en el art. 16 TFUE y debe considerarse *lex specialis* respeto de las normas sobre el tratamiento de datos contenidas en el artículo 10 DEP.

El ámbito de esta *lex specialis* es el concreto definido, no extendiéndose a otros tratamientos similares para fines distintos de la aplicación de la Ley, aún por autoridades competentes⁶⁵. En definitiva, el establecimiento de prohibiciones, o

63. Art. 3, apartados 37 y 50 a 53 RIA.

64. COMISIÓN EUROPEA 2021. p.4. Así se traslada además al Considerando (3) RIA: «...En la medida en que el presente Reglamento contiene normas específicas sobre la protección de las personas físicas en lo que respecta al tratamiento de datos personales relativos a las restricciones del uso de sistemas de IA para la identificación biométrica remota a efectos de la aplicación de la ley, para el uso de sistemas de IA para la evaluación del riesgo de las personas físicas a efectos de la aplicación de la ley y para el uso de sistemas de IA de categorización biométrica a efectos de la aplicación de la ley, procede basar el presente Reglamento, en la medida en que se refiera a dichas normas específicas, en el artículo 16 del TFUE. A la luz de esas normas específicas y del recurso al artículo 16 del TFUE, procede consultar al Consejo Europeo de Protección de Datos».

65. Considerando (39): «...En la aplicación del artículo 9, apartado 1, del Reglamento (UE) 2016/679, el uso de la identificación biométrica remota para fines distintos de la aplicación de

incluso requisitos adicionales resultado de la clasificación del SIA, ante el riesgo que supone para la protección de los datos personales, deben considerarse complementarios del marco jurídico de la protección de datos vigente, adicionales a los requisitos y obligaciones establecidos en el mismo (DEP).

Sobre estos extremos se ha de tener en cuenta:

— En ningún caso puede considerarse que proporciona base legítima para el tratamiento de datos personales conforme al artículo 8 DEP⁶⁶. En general, no pueden confundirse las reglas que delimitan la existencia de un SIA prohibido o SIA de Alto riesgo, con la base legítima de tratamiento⁶⁷.

— Sobre esta base legítima de tratamiento, parece referirse a ella el art. 5, apartado 5, RIA, con remisión a las normas estatales, que habrán de entenderse vinculadas, no solo por el art. 5 RIA, sino también por las condiciones impuestas por el marco jurídico de protección de datos personales.

— Debe garantizarse la aplicación del RGPD, en la medida del riesgo que determinan para la protección de los datos personales, y atenderse a los principios de necesidad, proporcionalidad, limitación de finalidad, entre otros.

— Se actualiza la obligación de garantizar la existencia de una autoridad independiente a la que se someta el control del cumplimiento de estas normas⁶⁸.

4. SUPUESTOS ESPECÍFICOS DE BASE JURÍDICA DE TRATAMIENTO DE DATOS PERSONALES EN EL REGLAMENTO

En tercer lugar, debemos considerar la existencia de normas en el RIA que expresamente introducen una base legítima de tratamiento conforme a las previsiones del RGPD, como anticipa el artículo 2, apartado 7, RIA:

Artículo 10.5 Reglamento, «para para garantizar la detección y la corrección de los sesgos asociados a los sistemas de inteligencia artificial de alto riesgo».

la ley ya ha sido objeto de decisiones de prohibición por parte de las autoridades nacionales de protección de datos».

66. Considerando (38) RIA.

67. Considerando (63) RIA, y el precitado Considerando (38). También CEPS, 2022, p. 14: «El considerando 41 de la propuesta establece que los operadores de sistemas de IA deben atenerse al régimen de protección de datos de la UE, afirmando, en particular, que las categorías basadas en el riesgo de la Ley de IA no deben interpretarse en el sentido de que proporcionan fundamentos jurídicos para el tratamiento de datos personales. Esta disposición sienta las bases para la compatibilidad entre la Ley de IA y el RGPD, pero su generalidad exige una mayor especificación de las normas en ambos actos, como lo demuestran las siguientes secciones.»

68. EDPS y EDPB. 2021 p. 22 —apartado 49—. En este punto es relevante también el informe en el apartado 50: «Sin embargo, no hay ninguna disposición explícita en la propuesta que asigne competencias para garantizar el cumplimiento de estas normas al control de las autoridades independientes. La única referencia a las autoridades de control de protección de datos competentes en virtud del RGPD, o LED, es el artículo 63, apartado 5, de la propuesta, pero solo como organismos de “vigilancia del mercado” y, alternativamente, con otras autoridades. El CEPD y el SEPD consideran que esta creación no garantiza el cumplimiento del requisito de control independiente establecido en el artículo 16, apartado 2, del TFUE y en el artículo 8 de la Carta».

El artículo 10 RIA, relativo a datos y gobernanza de datos, integrado en la Sección 2, del Capítulo III («Requisitos de los Sistemas de IA de Alto Riesgo») prevé en su apartado 5 la base legítima necesaria para el tratamiento de datos personales conforme al art. 6 RGPD, y levanta la prohibición a la que se refiere el art. 9.1 RGPD, Art. 10 DEP y Art. 10.1 Reglamento (UE) 2018/1275. Para ello el tratamiento, que se habilita excepcionalmente, debe ser estrictamente necesario para garantizar la detección y la corrección de los sesgos negativos asociados a los SIA de AR, considerándose a estos efectos como responsables del tratamiento a los proveedores de dichos sistemas.

Siendo esto así y conforme al artículo 9 RGPD, se impone la obligación de adoptar las medidas de garantía que se consideren necesarias, citándose expresamente el establecimiento de limitaciones técnicas a la reutilización y la utilización de las medidas de seguridad y protección de la privacidad más recientes, tales como la seudonimización o el cifrado, si la anonimización pudiera afectar significativamente al objetivo perseguido. En este sentido, la redacción inicial fue completada (Enmienda 290 Propuesta de Reglamento Artículo 10 - apartado 5), resaltando su carácter excepcional, referido a sesgos negativos e incorporando garantías adicionales⁶⁹. Se requiere, de forma cumulativa:

— La detección y corrección de sesgos no pueden cumplirse eficazmente mediante el tratamiento de otros datos, incluidos los datos sintéticos o anónimos.

— El tratamiento de categorías especiales de datos está sujetas a limitaciones técnicas para la reutilización y a las medidas de seguridad y protección de privacidad más avanzadas, incluida la seudonimización.

— Los datos personales tratados estarán protegidos, sujetos a garantías adecuadas, incluidos controles estrictos y documentación del acceso, para evitar el uso indebido y garantizar que solo las personas autorizadas tengan acceso a ellos con las obligaciones de confidencialidad adecuadas, y no deben ser transmitidos, transferidos o ser accesibles por terceros de otro modo.

— Los datos personales tratados se suprimen una vez corregido el sesgo o al llegar al final de su período de conservación, cualquiera que sea el primero.

Por último, el tratamiento de datos personales en este caso requiere una previsión específica en los *registros de las actividades de tratamiento*⁷⁰ que han de incluir una justificación de por qué el tratamiento de categorías especiales de datos personales era estrictamente necesario para detectar y corregir sesgos y este objetivo no pudo lograrse mediante el tratamiento de otros datos.

69. «73. ... Al mismo tiempo, el artículo 10, apartado 5, de la propuesta dice “los proveedores de dichos sistemas podrán tratar categorías especiales de datos personales”. Además, la misma disposición requiere salvaguardias adicionales, también dando ejemplos. Por lo tanto, la propuesta parece interferir con la aplicación del RGPD, el LED y el EUDPR. Si bien el CEPD y el SEPD acogen con satisfacción el intento de establecer garantías adecuadas, es necesario un enfoque regulador más coherente, ya que las disposiciones actuales no parecen suficientemente claras para crear una base jurídica para el tratamiento de categorías especiales de datos, y deben complementarse con medidas de protección adicionales que aún deben evaluarse. Además, cuando se hayan recopilado datos personales mediante el tratamiento en el ámbito de aplicación del LED, deberán tenerse en cuenta las posibles salvaguardias y limitaciones adicionales derivadas de las transposiciones nacionales del LED.» ([EDPS y EDPB, 2021, p. 28]).

70. Artículo 30 («Registro de actividades de tratamiento») RGPD.

Debe tenerse en cuenta, en cualquier caso, que este precepto también se encuentra al servicio del derecho fundamental a la protección de datos, y los principios en los que se asienta⁷¹.

Artículo 59 Reglamento, relativo al «tratamiento ulterior de datos personales para el desarrollo de determinados sistemas de inteligencia artificial en favor del interés público en el espacio controlado de pruebas de la inteligencia artificial».

El art. 59.1 RIA proporciona la base legítima, y levantamiento de la prohibición, necesarios para el tratamiento ulterior de datos personales para el desarrollo, entrenamiento y prueba de determinados SIAs, por razones de interés público, en el espacio controlado de pruebas para la IA, de conformidad con las previsiones del artículo 6, apartado 4, RGPD y el artículo 9.2.g) RGPD⁷².

Nos referimos a SIAs desarrollados para salvaguardar un interés público sustancial, en los ámbitos a que se refiere el 59.1.a) RIA, tales como la seguridad pública y salud pública, incluida la detección de enfermedades, el diagnóstico prevención, control y tratamiento y mejora de los sistemas sanitarios; un alto nivel de protección y mejora de la calidad del medio ambiente, la protección de la biodiversidad, la contaminación, así como la transición ecológica, la mitigación del cambio climático y la adaptación al mismo; sostenibilidad energética; seguridad y resiliencia de los sistemas de transporte y movilidad, infraestructuras y redes críticas; eficiencia y calidad de la administración pública y los servicios públicos.

En este punto debe tenerse en cuenta que la habilitación se asienta en el cumplimiento de requisitos de forma cumulativa, junto con la finalidad de interés público sustancial a que se refiere el art. 59.1.a), enumerados de forma exhaustiva en el artículo 59.1 RIA entre los que destacan, desde el punto de vista de la protección de los datos personales⁷³:

— Los datos tratados sean necesarios para cumplir uno o varios de los requisitos exigidos a los SIA de AR (Cap. III, Secc. 2), cuando no puedan cumplirse efectivamente mediante el tratamiento de datos anonimizados, sintéticos u otros datos no personales.

— Existen mecanismos de seguimiento eficaces para determinar la necesidad de una EIPD conforme al art. 35 RGPD.

71. *«A fin de garantizar un tratamiento justo y transparente con respecto al interesado, (...) el responsable del tratamiento debe utilizar procedimientos matemáticos o estadísticos adecuados para la elaboración de perfiles, aplicar medidas técnicas y organizativas adecuadas para garantizar, en particular, que se corrijan los factores que dan lugar a inexactitudes en los datos personales y que se minimice el riesgo de errores y se garantice la seguridad de los datos personales de manera que se tengan en cuenta los riesgos potenciales para los intereses y los derechos del interesado y se evite, entre otras cosas, los efectos discriminatorios para las personas físicas (...).» (FRA UE, 2018, p. 7). En el mismo sentido CONTINO HUESO 2023. p. 303.*

72. *«...el tratamiento es necesario por razones de un interés público esencial, sobre la base del Derecho de la Unión o de los Estados miembros, que debe ser proporcional al objetivo perseguido, respetar en lo esencial el derecho a la protección de datos y establecer medidas adecuadas y específicas para proteger los intereses y derechos fundamentales del interesado».*

73. La Enmienda 506 y ss. del Parlamento al artículo 54, incrementó las garantías jurídicas, técnicas y organizativas para la protección de los datos personales. En este sentido, EDPS y EDPB. 2021, apartados 64 y ss.

— Los datos personales están en un entorno de tratamiento de datos funcionalmente separado, aislado y protegido bajo el control del posible proveedor y con acceso solo de personas autorizadas.

— Los datos han sido recopilados conforme al RGPD, y no pueden compartirse fuera del entorno controlado de pruebas.

— El tratamiento de los datos no puede dar lugar a medidas o decisiones que afecten a los interesados o sus derechos conforme al RGPD, siendo así aclarado por el Considerando (140)⁷⁴ RIA.

— Los datos personales estarán protegidos mediante medidas técnicas y organizativas adecuadas y se suprimirán una vez que la participación haya terminado o los datos personales hayan llegado al final de su período de conservación.

— Los registros del tratamiento de datos personales se conservarán mientras dure la participación, a menos que la legislación de la Unión o nacional disponga otra cosa.

En cualquier caso, y desde el punto de vista del cumplimiento del RGPD, serán aplicables las obligaciones impuestas en el mismo a los responsables de tratamiento y los derechos de los interesados, resolviéndose la duda planteada durante la tramitación de la propuesta sobre si efectivamente nos encontrábamos en un marco de tratamiento de datos personales en el que alcance de las obligaciones de los responsables y de los derechos de los interesados estaban siendo limitadas⁷⁵.

Por otro lado, siendo este espacio un marco concreto y controlado establecido por una autoridad competente y ofrecido por dicha autoridad a proveedores o posibles proveedores de sistemas de IA para desarrollar, entrenar, validar y probar, el SIA conforme a un plan, temporal y bajo supervisión regulatoria⁷⁶, pueden suscitarse dudas sobre el marco de rendición de cuentas conforme al RGPD, en concreto sobre la identificación de responsable, o corresponsables, de los tratamientos⁷⁷.

Hay que decir, por último, que la base jurídica considerada no se prevé con relación al tratamiento de datos personales que pudiera tener lugar en el caso de «Pruebas de sistemas de IA de alto riesgo en condiciones del mundo real fuera de los espacios controlados de pruebas de IA», conforme a las previsiones del artículo 60. En este punto, debe considerarse que los requisitos para los tratamientos, las obligaciones de sus responsables, los derechos de los interesados previstos en el RGPD se superponen íntegramente a las condiciones, cumulativas, establecidas en el apartado 4, especialmente párrafos e) g) h) i) y k), y el apartado 5. Esta consideración nos permite descartar el *consentimiento informado del sujeto de pruebas* (Artículo 61 RIA) como el consentimiento a que se refieren los artículos 6 y 9.2 RGPD, base legítima del tratamiento o supuesto de levantamiento de la prohibición del tratamiento, siendo así confirmado por el Considerando (141) *in fine* RIA⁷⁸.

74. «En particular, el presente Reglamento no debe proporcionar una base jurídica en el sentido del artículo 22, apartado 2, letra b), del Reglamento (UE) 2016/679 y del artículo 24, apartado 2, letra b), del Reglamento (UE) 2018/1725.».

75. Considerando (140) RIA; y EDPS y EDPB. 2021 apartado 64.

76. Art. 3.55 RIA.

77. Así se puso de manifiesto por EDPS y EDPB. 2021 apartado 65.

78. «...El consentimiento de los interesados para participar en dichas pruebas en virtud del presente Reglamento es distinto y sin perjuicio del consentimiento de los interesados para el

5. AUTORIDADES INDEPENDIENTES DE CONTROL EN MATERIA DE PROTECCIÓN DE DATOS PERSONALES

Como indicamos en el apartado III.1, forma parte del marco jurídico de protección de los datos personales, como derecho fundamental, sus instituciones de garantía, destacando en este sentido la intervención de las *autoridades independientes de control en materia de protección de datos*, respecto de las cuales es exigible un marco de certeza y seguridad jurídica, en los términos indicados. Esta intervención en el caso de SIAs, por otro lado, debe producirse en un ámbito de «*cooperación estructurada e institucionalizada*» con otras autoridades competentes⁷⁹, como las autoridades de vigilancia del mercado, siempre respetando aquella independencia⁸⁰.

Siendo este el punto de partida, podemos señalar:

— El RIA debe entenderse sin perjuicio de la aplicación del RGPD, que incluye las competencias, funciones y potestades de las autoridades independientes de control (arts. 55 y ss. RGPD) posibilitando además acceso a cualquier documentación, entendiendo que se extiende al procedimiento de salvaguarda para garantizar una aplicación adecuada y oportuna de los SIAs que presenten un riesgo para los derechos fundamentales⁸¹.

De acuerdo con ello, el artículo 77 RIA relativo a los *poderes de las autoridades que protegen los derechos fundamentales*, las faculta para solicitar y acceder a cualquier documentación creada o mantenida en virtud del RIA cuando sea necesario «*para el cumplimiento efectivo de su mandato dentro de los límites de su jurisdicción*», informando a la autoridad de vigilancia del mercado. En los supuestos previstos en el apartado 3, podrán incluso solicitar motivadamente a la autoridad de vigilancia del mercado la organización de pruebas de SIA de AR.

Estas previsiones en todo caso deben ser interpretadas en el sentido más acorde con la independencia de las autoridades de control en materia de protección de datos y las normas que articulan su actuación conforme al RGPD.

— En el contexto de la consideración del RIA como *lex specialis* con relación a los *sistemas de identificación biométrica a distancia «en tiempo real» en espacios accesibles al público* con fines policiales, sin perjuicio de la previsión sobre la participación de una autoridad independiente de control, también establece la necesaria notificación a las autoridades de control en materia de protección de datos (Art. 5, apartado 4, 5 y 6 RIA), debiendo completarse con la obligaciones de los responsables del despliegue respecto del uso de sistemas de identificación biométrica a distancia, de puesta a disposición e informe, conforme al artículo 26, apartado 10, RIA⁸².

La exclusión de la comunicación de *datos operativos sensibles* debe ponerse en relación con el Considerando (159) *in fine* según el cual «*ninguna exclusión*

tratamiento de sus datos personales con arreglo a la legislación pertinente en materia de protección de datos...».

79. EDPS. 2022, p.14.

80. Brown, 2023, p. 69.

81. Considerandos (10) y (157).

82. Considerando (36).

de la divulgación de datos a las autoridades nacionales de protección de datos en virtud del presente Reglamento debe afectar a las competencias actuales o futuras de dichas autoridades más allá del ámbito de aplicación del presente Reglamento».

— Se hace referencia a las autoridades de control en materia de protección de datos en el Considerando (140), previendo la cooperación con las autoridades competentes en el espacio controlado de pruebas de la IA.

Esta previsión es desarrollada en el artículo 57, apartado 10 RIA, con relación a su eventual participación⁸³, y cuando hayan proporcionado orientaciones para el cumplimiento del RGPD (apartado 12).

En la medida en que se impide la imposición de multas administrativas a los responsables que hubiesen seguido las orientaciones de *buena fe* (Apartado 1, h) *in fine*) concepto este que puede generar incertidumbre, será necesario establecer una regulación completa y precisa para las orientaciones indicadas, y su marco de cumplimiento.

— El artículo 74 RIA, apartado 8, prevé la posibilidad de nombramiento de las Autoridades de control en materia de protección de datos como autoridades de vigilancia del mercado, respecto de los SIA de AR enumerados en el punto 1 del anexo III, en la medida en que los sistemas se utilicen a los efectos de la aplicación de la ley y para los fines enumerados en los puntos 6, 7 y 8 del mismo anexo.

Siendo este un escenario propuesto, lo cierto es que requiere en todo caso una adecuada delimitación de las competencias, funciones y potestades de estas autoridades en uno y otro ámbito, considerando las normas de aplicación.

— Por último, el artículo 85, se refiere al derecho a presentar una reclamación ante una autoridad de vigilancia del mercado, por quien considere que se ha producido una infracción del RIA⁸⁴.

Esta reclamación debe entenderse sin perjuicio del derecho a presentar reclamaciones ante las autoridades de control de protección de datos de acuerdo con el artículo 77 RGPD, que las tratarán conforme al artículo 57.1.f) RGPD y resolverán, en su caso, de acuerdo con las potestades prevista en el 58.2 RGPD

83. «Las autoridades nacionales competentes velarán por que, en la medida en que los sistemas innovadores de IA impliquen el tratamiento de datos personales o entren en el ámbito de control de otras autoridades nacionales o autoridades competentes que faciliten o apoyen el acceso a los datos, las autoridades nacionales de protección de datos y esas otras autoridades nacionales estén asociadas al funcionamiento del sandbox regulador de la IA y participen en la supervisión de dichos aspectos en la medida de sus respectivas funciones y competencias, según proceda».

84. En este punto debe recordarse que el artículo 110 (Capítulo XIII. Disposiciones Finales) modifica el Anexo I de la Directiva (UE) 2020/1828 del Parlamento Europeo y del Consejo, de 25 de noviembre de 2020, relativa a las acciones de representación para la protección de los intereses colectivos de los consumidores, y por la que se deroga la Directiva 2009/22/CE (DO L 409 de 4.12.2020, p. 1), incluyendo el apartado 68) que posibilita *acciones de representación* (para la protección de intereses colectivos de consumidores) en caso de infracción del RIA. Debemos recordar que también se posibilita el ejercicio de estas acciones en caso de infracción del RGPD (apartado 56) de la Directiva. Ver COTINO HUESO (2022) pp. 83 y 84.

—sancionadoras, correctivas, cautelares ...—. Así resulta del artículo 85.1 RIA, interpretado a la vista del Considerando (170)⁸⁵.

Debe recordarse, por otro lado, el alcance de la finalidad de esta reclamación del artículo 85, pues conforme a su párrafo segundo, se tendrán en cuenta para el objetivo de llevar a cabo las actividades de vigilancia del mercado y gestionarse de conformidad con los procedimientos específicos establecidos por las autoridades de vigilancia del mercado (Art. 11 Reglamento (UE) 2019/1020)⁸⁶.

6. VIGILANCIA HUMANA Y DECISIONES INDIVIDUALES AUTOMATIZADAS: ARTÍCULO 22 RGPD Y EL REGLAMENTO

El artículo 22 RGPD consagra el derecho de todo interesado a «*no ser objeto de una decisión basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzca efectos jurídicos en él o le afecte significativamente de modo similar*».

Si bien puede entenderse que el art. 22 RGPD es un supuesto restringido («... únicamente...») lo cierto es que las últimas interpretaciones jurisprudenciales analizan la relación entre el resultado del tratamiento automatizado y la decisión adoptada desde el punto de vista cualitativo⁸⁷, en función de la influencia real que puede que ser ejercida⁸⁸. Todo lo anterior sin perjuicio de la intervención humana a que se refiere el artículo 22, apartado 3, RGPD de carácter reactiva.

Si analizamos los requisitos de los SIAs de AR (Capítulo III, Sección 2, RIA) en concreto la *vigilancia humana* a que se refiere el artículo 14, podemos llegar a la conclusión que el supuesto de hecho del artículo 22 RGPD, en particular en los casos en que la decisión individual automatizada sería posible por incluirse en alguno de los casos de su apartado 2, no podría tener lugar.

El alcance del artículo 14 RIA se complementa con la información que ha de ser suministrada al implementador, conforme al artículo 13, apartado 3, relativo a las instrucciones de uso, letra d). También se informa al ANEXO IV, apartado 2.e), como parte de la documentación técnica del SIA de AR a que se refiere el artículo 11 RIA. y forma parte del contenido de la Evaluación de Impacto para los Derechos Fundamentales (Art. 27.1.e) RIA, debiendo incorporarse a la Evaluación de Impacto para la Protección de los Datos Personales (EIPD) conforme a las previsiones del art. 35.7.d) RGPD.

Efectivamente, podemos cuestionarnos si el efectivo cumplimiento de los requisitos impuestos a los SIAs de AR por el precitado artículo 14 RIA, en sus apartados 4 y 5, impiden la existencia de decisiones individuales en el sentido del artículo 22, que sí podrían estar permitidas con arreglo a su apartado 2.

85. «El Derecho nacional y de la Unión ya prevé vías de recurso efectivas para las personas físicas y jurídicas cuyos derechos y libertades se vean perjudicados por el uso de sistemas de IA. Sin perjuicio de dichas vías de recurso, toda persona física o jurídica que tenga motivos para considerar que se ha producido una infracción del presente Reglamento debe tener derecho a presentar una reclamación ante la correspondiente autoridad de vigilancia del mercado...».

86. Pueden resultar de interés las reflexiones sobre acciones colectivas.

87. Puede analizarse la Sentencia TJUE (Sala Primera) de 7 de septiembre de 2023, OQ y Land Hessen, con SCHUFA Holding AG, Asunto C-634/21, apartados 53 a 55 (TJCE 2023\146), y el análisis realizado por COTINO HUESO (2024).

88. En el mismo sentido GT art. 29 (2018) apartado 4.1.

7. EL DERECHO A UNA EXPLICACIÓN. DECISIONES INDIVIDUALES EN EL CONTEXTO DE DETERMINADOS SIA DE AR Y ARTÍCULO 22 RGPD

El artículo 86 («*Derecho a explicación de decisiones tomadas individualmente*») establece el derecho de toda persona afectada por una decisión adoptada por el responsable del despliegue basándose en los resultados de un SIA de AR (salvo Anexo III.2⁸⁹), y que produzca efectos jurídicos o le afecte considerablemente del mismo modo, a solicitarle explicaciones claras y significativas sobre el papel del sistema de IA en el procedimiento de toma de decisiones y los principales elementos de la decisión adoptada⁹⁰, si estima que afecta negativamente a su salud, seguridad y derechos fundamentales.

Este precepto se relaciona con el artículo 26.11 RIA, que impone a los responsables del despliegue⁹¹ que tomen decisiones o ayuden a tomar decisiones relacionadas con personas físicas, la obligación de informarles que están sujetos al uso del sistema de IA de alto riesgo⁹², y este, a su vez, con el artículo 13.1 y 3.b) iv) y v) RIA relativo a las obligaciones de transparencia y suministro de información a los responsables del despliegue, precisamente para el cumplimiento, entre otras, de aquellas obligaciones⁹³, todo ello sin perjuicio de las previsiones relativas a la Base de Datos de la UE para SIA de AR, conforme al artículo 71 RIA —en su remisión al ANEXO VIII, Sección A, apartado 6—⁹⁴.

Es importante establecer la relación del Artículo 86 RIA, y el derecho que establece, con los derechos del interesado en el contexto del Artículo 22 RGPD («*Decisiones individuales automatizadas, incluida la elaboración de perfiles*») y concordantes sobre el derecho a una explicación en estos supuestos, en la medida en que el derecho a que se refiere el artículo 86 «*solo se aplicará en la medida en que ...no sea ya previsto en la legislación de la Unión*» según establece su apartado 3.

89. «*Sistemas de IA destinados a utilizarse como componentes de seguridad en la gestión y el funcionamiento de infraestructuras digitales críticas, el tráfico por carretera y el suministro de agua, gas, calefacción y electricidad.*».

90. «*...salvo, y en la medida, en que se restrinja esta obligación por el Derecho de la Unión o nacional de conformidad con el Derecho de la Unión (apartado 2)*».

91. Sin perjuicio de las previsiones del Artículo 50, relativo a las «*Obligaciones de transparencia para los proveedores y responsables del despliegue de determinados sistemas de IA*», que interactúen con personas físicas; o de las previsiones específicas relativas a SIA de AR con fines policiales (Art. 13 DEP).

92. Para los sistemas de IA de alto riesgo utilizados con fines policiales, se aplicará el artículo 13 de la Directiva 2016/680.

93. En este punto el Parlamento Europeo propuso la introducción de un segundo párrafo, de contenido más explícito: «*De este modo, la transparencia significará que, en el momento en que se comercialice el sistema de IA de alto riesgo, se utilicen todos los medios técnicos disponibles de conformidad con los últimos avances tecnológicos generalmente reconocidos para garantizar que los resultados del sistema de IA sean interpretables por el proveedor y el usuario. Se permitirá al usuario comprender y utilizar el sistema de IA de forma adecuada sabiendo en general cómo funciona el sistema de IA y qué datos procesa, lo que le permitirá explicar las decisiones adoptadas por el sistema de IA a la persona afectada de conformidad con el artículo 68, letra c).*»

94. «*Con respecto a los sistemas de IA de alto riesgo que vayan a inscribirse en el registro de conformidad con el artículo 49, apartado 1, se facilitará y se actualizará debidamente la información siguiente:...6. Una descripción sencilla y concisa de la información que utiliza el sistema (datos, entradas) y su lógica de funcionamiento.*»

Debemos considerar que el artículo 86.1 RIA asume que la *vigilancia humana* del SIA de AR ha tenido lugar —de acuerdo con lo previsto en el artículo 14 RIA—, lo cual no ha impedido que el responsable del despliegue haya adoptado la decisión basándose en el resultado del RIA de AR.

Sobre la interpretación de la «*decisión que el responsable del despliegue adopte basándose en los resultados de un sistema de IA*», podríamos entender que incluye aquella en la que la información de salida tiene una *influencia significativa*⁹⁵. Esta interpretación resulta del Considerando (171)⁹⁶ («...cuando la decisión...se base principalmente...»), de la habilitación a la Comisión contenida en el Artículo 7.1 RIA, en relación con el artículo 7.2.g) RIA⁹⁷, siendo la más favorable al derecho de los afectados. Se corresponde, además, con la evolución de la doctrina del TJUE (Nota 61).

Por otro lado, ya hemos indiciado el debate sobre la dificultad de ocurrencia del supuesto de hecho del artículo 22 RGPD, en su apartado 2, en el caso de SIAs de AR, en la medida que la vigilancia humana es exigida para estos. Si considerásemos que la vigilancia humana exigida en el Art. 14 RIA no es un obstáculo a las decisiones automatizadas en el sentido del Artículo 22 RGPD, en los supuestos de su apartado 2, podríamos analizar:

— El artículo 86.1 RIA se refiere a decisiones adoptadas por el responsable del despliegue sobre la base de la salida de determinados SIAs de AR, en los términos amplios que hemos indicado, y el Art. 22 se refiere a «*decisión basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles*», siendo también objeto de una interpretación amplia.

— El derecho a la explicación previsto el artículo 86 RIA, se refiere exclusivamente a SIA de AR⁹⁸ (salvo Anexo III.2), siendo, por tanto y en cierta forma, más restrictivo que el supuesto hecho del art. 22 RGPD.

95. Ver propuesta del Parlamento Europeo, a los efectos de la categorización como de SIA de AR. p.e. Enmiendas 46 relativa al Considerando 36, Enmienda 47 al Considerando 37 y Enmienda 814 al ANEXO III.4.a) y en cierto modo Considerando (53) RIA.

96. «*Las personas afectadas deben tener derecho a obtener una explicación cuando la decisión de un responsable del despliegue se base principalmente en los resultados de determinados sistemas de alto riesgo que entran dentro del ámbito de aplicación del presente Reglamento y cuando dicha decisión produzca efectos jurídicos o afecte significativamente de modo similar a dichas personas, de manera que consideren que tiene un efecto negativo en su salud, su seguridad o sus derechos fundamentales. Dicha explicación debe ser clara y significativa y servir de base para que las personas afectadas puedan ejercer sus derechos. El derecho a obtener una explicación no debe aplicarse a la utilización de sistemas de IA para los que se deriven excepciones o restricciones con arreglo al Derecho nacional o de la Unión, y solo debe aplicarse en la medida en que este derecho no esté ya previsto en el Derecho de la Unión.*».

97. «*la medida en que las personas que podrían sufrir dicho perjuicio o dichas repercusiones negativas dependan del resultado generado por un sistema de IA, en particular porque, por motivos prácticos o jurídicos, no sea sensato poder renunciar a dicho resultado;*». También del contenido del Considerando (53) en el establecimiento de las condiciones para identificar el carácter sustancial de la influencia del SIA en la toma de decisiones, en una de las hipótesis humana.

98. Debe recordarse que el alcance de la vinculación de la información de la información de salida con la decisión potencialmente perjudicial es uno de los elementos considerados para determinación de los SIA de AR al amparo del Artículo 7.1 RIA,

— Sobre el alcance del derecho de los afectados, el artículo 86 se refiere a solicitar al implementador *«explicaciones claras y significativas sobre el papel del sistema de IA en el procedimiento de toma de decisiones y los principales elementos de la decisión adoptada»* de modo que proporciona una base para que «ejercen sus derechos» (Considerando 171). Debe completarse en cualquier caso con el artículo 26.11 RIA, relativo a la obligación de informar a los afectados que están sujetos al uso del sistema de IA de alto riesgo.

— Por su parte, el responsable de tratamiento asume la obligación de información sobre al interesado en forma concisa, transparente, inteligible y de fácil acceso, con un lenguaje claro y sencillo cualquier información dirigida específicamente a un niño (art. 12.1 RGPD). Específicamente en caso de decisiones automatizadas, incluida la elaboración de perfiles, a que se refieren los artículos 22.1 y 4 RGPD (arts. 13.2.f) y 14.2 g) RGPD, bien en el momento de obtención de los datos, en un plazo razonable, o antes de empezar el tratamiento, debe facilitar información sobre su existencia e información significativa sobre la lógica aplicada, así como la importancia y las consecuencias previstas de dicho tratamiento para el interesado. En lo que nos interesa, esta misma información ha de ser proporcionada, entre otra, en el caso de ejercicio del derecho de acceso a que se refiere el artículo 15.1.h) RGPD.

A la vista de todo lo anterior, podemos considerar que los derechos atribuidos a los interesados en el RGPD son más amplios en sus presupuestos, por lo que con carácter general se activará la excepción contenida en el apartado 3 del artículo 86 RIA, en el caso de decisiones automatizadas que afecte a las personas, sobre todo y en la medida en que ampara supuestos de SIA que no sean de AR.

Si se aceptase la posibilidad de aplicación del supuesto previsto en el art. 22.2 RGPD a los SIAs de AR, y se interpretara que alguna información de las previstas en el RIA no se proporciona al amparo del RGPD, se podría solicitar por el interesado en la parte no prevista. Ello tendría especial importancia para el caso de aplicación del supuesto contemplado en el artículo 22.2.b) RGPD, no amparado por el derecho a que se refiere el artículo 22.3 RGPD.

Sería deseable, en cualquier caso, que la información a que hace referencia el precitado art. 86 RIA sirva para construir un estándar de información a proporcionar al interesado conforme al RGPD, incorporándose a ésta en todo caso⁹⁹. Recordemos que, en ambos casos, debe ser información suficiente para el ejercicio de los derechos que les asisten (Considerando (171) RIA y art. 22.3 RGPD).

como establece el artículo 7.2.e) RIA, y que fue tenido en cuenta en su inclusión en el ANEXO III.

99. EDPB (2022): «119. El artículo 15, apartado 1, letra h), establece que todo interesado debe tener derecho a ser informado, de manera significativa, entre otras cosas, sobre la existencia y la lógica subyacente de la toma de decisiones automatizada, incluida la elaboración de perfiles sobre el interesado, y sobre la importancia y las consecuencias previstas que dicho tratamiento podría tener⁶⁹. Si es posible, la información con arreglo al artículo 15, apartado 1, letra h), debe ser más específica en relación con el razonamiento que conduzca a decisiones específicas relativas al interesado que solicitó el acceso.» pp. 39 y 40.

69 Véanse, en este nombre, las Directrices sobre transparencia en virtud del Reglamento 2016/679 (WP 260), apartado 41, con referencia a las Directrices sobre la toma automatizada de decisiones individuales y s

8. COLABORACIÓN DEL REGLAMENTO EN EL CUMPLIMIENTO DEL RGPD

El último inciso del Considerando (10) RIA declara que *«las normas armonizadas para la introducción en el mercado, la puesta en servicio y el uso de los sistemas de IA establecidos en virtud del presente Reglamento deben facilitar la aplicación efectiva y permitir el ejercicio de los derechos de los interesados y otras vías de recurso garantizadas en virtud del Derecho de la Unión en materia de protección de los datos personales así como de otros derechos fundamentales»*.

Este enunciado general permite aproximarnos a la medida en que las normas armonizadas contenidas en RIA facilitan la aplicación del RGPD a los SIAs cuando supongan, formen parte, sean empleados en o para un tratamiento de datos personales, en los términos y con el alcance que hemos anticipado.

Con este fin, podemos identificar preceptos del RIA que imponen a los operadores y agentes, principalmente proveedores y responsables del despliegue, determinadas obligaciones que faciliten el cumplimiento del RGPD. Así:

— En primer lugar, de forma natural, tendríamos la enumeración de SIAs que deben considerarse prohibidos, o limitados, por suponer una finalidad que también habría de considerarse prohibida conforme al RGPD. De este modo, el RIA, y sus mecanismos institucionales de control, y garantía del cumplimiento, refuerzan la aplicación del RGPD.

— El establecimiento de requisitos que han de cumplir los SIAs de AR (Sección 2 del Capítulo III) así como el establecimiento de obligaciones para los proveedores y responsables del despliegue de los SIAs de AR, puede facilitar igualmente el cumplimiento del RGPD —en los términos ya indicados, incluyendo el principio de transparencia— y, sobre todo, la demostración de su cumplimiento en el marco de la *responsabilidad proactiva* (art. 5.2 RGPD).

Nos estamos refiriendo al cumplimiento por proveedores (Art. 16) responsables del despliegue (Art. 26) y otros agentes (artículos 22 a 25 RIA) de las obligaciones establecidas en las Secciones 2 y 3 del Capítulo III, relativas al *sistema de gestión de riesgos* (art. 9 RIA), *datos y gobernanza de dato* (art. 10 RIA), a la *documentación técnica* —incluido el contenido del ANEXO IV al que se remite y al *mantenimiento de registros* (Arts. 11 y 12), sobre *transparencia y suministro de información a los responsables del despliegue* (Artículo 13 RIA), sobre *precisión, robustez y ciberseguridad* (Art. 15 RIA) sobre el *sistema de gestión de calidad* (art. 17 RIA), *conservación de la documentación* (art. 18 RIA), *archivos de registros generados automáticamente* (art. 19 RIA), *medidas correctoras y deber de información* (Art. 20), *responsabilidades a lo largo de la cadena de valor* (art. 25 RIA), sobre la *Evaluación de Impacto sobre los derechos fundamentales* (art. 27 RIA), e incluso el cumplimiento de las obligaciones de registro a que se refiere el artículo 49, en relación con el artículo 16.i) RIA. Lo mismo podemos decir en el caso de modelos de IA de uso general, conforme a los artículos 51 y siguientes RIA, así como el contenido en el ANEXO XI y siguientes.

— Sobre la evaluación de impacto en la protección de los datos, a que se refiere el artículo 35 RGPD, se contienen en el RIA normas que la ponen en valor.

Así, conforme al artículo 26, apartado 9, RIA los responsables del despliegue de los SIA de AR utilizarán la información que le sea suministrada conforme al art. 13 RIA para la elaboración de la evaluación de impacto en materia de protección

de datos (Art. 35 RGPD). Del mismo modo, el artículo 27 RIA, que impone a los responsables del despliegue de SIAs de AR la elaboración de una evaluación de impacto para los derechos fundamentales, en su apartado 4, que si ya se cumplen cualquiera de las obligaciones previstas en el precepto como resultado de la EIPD la evaluación de impacto relativa a los derechos fundamentales complementará dicha evaluación de impacto relativa a la protección de datos.

Por último, se incluye un resumen de la EIPD en relación con la información que ha de presentarse al Registro de SIA de AR, de conformidad con el art. 49, apartado 4 (ANEXO VIII, Sección B)) RIA, lo que facilitará su conocimiento, el ejercicio de derechos conforme al RGPD por los interesados, así como, en su caso, la reclamación ante las autoridades de control en materia de protección de datos.

Estas declaraciones, no obstante, pueden entenderse como la delimitación del espacio de diligencia exigible al responsable del despliegue en la elaboración de la EIPD, e incluso introducir confusión sobre el agente obligado, predeterminando la condición de responsables de los tratamientos en el contexto del SIA. En cualquier caso, la exigencia del EIPD, su contenido y el responsable de tratamiento, deberá determinarse en cada caso de acuerdo con el RGPD.

— Sobre el marco establecido en torno a la previsión de normas armonizadas, especificaciones comunes, evaluación de conformidad y certificados a que se refiere la Sección 5, Capítulo III, es evidente que al estar dirigidas a verificar el cumplimiento de las previsiones de la Sección 2, pueden facilitar la aplicación del RGPD, más aún cuando facilitará un marco de interacción normalizado entre los intervinientes, e incluso la identificación de los responsables y corresponsables de tratamiento de datos personales en la cadena de valor.

Aunque en el ANEXO V, apartado 5, se prevé como contenido de la Evaluación de Conformidad, por remisión del artículo 47, apartado 2 RIA, una declaración redactada y suscrita por el proveedor del SIA de AR, relativa a la existencia de tratamiento de datos personales y el cumplimiento del RGPD en ese caso, no se define como cometido específico de dicha evaluación de conformidad ni se documenta, posibilitando la confusión en torno al cumplimiento efectivo del derecho de la unión en lo que se refiere a la protección de datos personales en los SIAs de AR, que hayan pasado por el proceso. Esta previsión fue incorporada por Parlamento Europeo en su Enmienda 867.a. al ANEXO V y puede ponerse en relación con la declaración contenida en el Considerando (69), también introducido por el Parlamento Europeo.

Precisamente, hubiese sido deseable concretar de forma más precisa en la parte dispositiva la declaración contenida en el precitado Considerando (69)¹⁰⁰, pudiendo considerar idónea el artículo 10 RIA relativo a los datos y su gobernanza en este contexto, e incluso darle contenido en los ANEXOS IV y VII,

100. *«El derecho a la intimidad y a la protección de datos personales debe garantizarse a lo largo de todo el ciclo de vida del sistema de IA. A este respecto, los principios de minimización de datos y de protección de datos desde el diseño y por defecto, establecidos en el Derecho de la Unión en materia de protección de datos, son aplicables cuando se tratan datos personales. Las medidas adoptadas por los proveedores para garantizar el cumplimiento de estos principios podrán incluir no solo la anonimización y el cifrado, sino también el uso de una tecnología que permita llevar los algoritmos a los datos y el entrenamiento de los sistemas de IA sin que sea necesaria*

y disposiciones concordantes, en estos últimos supuestos, en su caso, mediante los Actos Delegados que pudiera adoptar la Comisión Europea a que se refiere el artículo 97 RIA (en relación con los artículos 11 y 43 RIA).

9. LIMITACIONES A LA COLABORACIÓN POR EL ESPACIO INICIAL DE REGULACIÓN

La mayoría de las previsiones que sirven para facilitar el cumplimiento del RGPD, se refieren o tienen como presupuesto que nos encontramos ante RIA de AR, incluyendo en algunos casos a los modelos de IA de uso general.

Si bien es cierto que la existencia de SIA de AR hace necesario aplicar el RGPD de modo acorde con el riesgo que se asume para el derecho a la protección de datos, también lo es que los SIAs, aunque no sean de alto riesgo, requieren asumir un marco de cumplimiento normativo desde la perspectiva RGPD, en los términos antes indicados. Se exigiría para estos SIAs no solo en relación al cumplimiento de requisitos y principios de tratamiento, sino también en la adopción de medidas técnicas y organizativas para la protección de datos personales desde el diseño y por defecto (art. 25 RGPD), con especial importancia de la EIPD.

Lo anterior debe entenderse sin perjuicio de la adopción de los códigos de conducta y mecanismos de gobernanza a que se refiere el artículo 95 RIA, para la aplicación voluntaria a sistemas de IA distintos de los SIA de AR, de algunos o todos los requisitos establecidos en el Capítulo III, Sección 2, que permitiría solventar total o parcialmente la limitación.

IV. REFLEXIONES FINALES

Era el objeto de estas líneas analizar, en la búsqueda de un marco jurídico cierto, las eventuales interacciones que se producen entre el RIA y el RGPD. Como decíamos, cómo se relacionan y complementan desde el punto de vista de la seguridad jurídica necesaria para la preservación del derecho a la protección de datos personales, en el entorno del imprescindible desarrollo tecnológico y de su regulación.

Se exponen como reflexiones finales, también a modo de conclusión:

— El RGPD es de plena aplicación a todo el ciclo de vida de los SIAs y a toda su cadena de valor, si en su contexto y con cualquier alcance se produce un tratamiento de datos personales o, a partir de ellos, el interesado resulte afectado por decisiones automatizadas, siendo el cumplimiento normativo, también en lo referido al RGPD, uno de los cimientos de una IA ética.

— En aquellos supuestos en los que el RIA prevé una norma como *lex specialis* (Apartado III.3) no puede considerarse que proporciona base legítima para el tratamiento de datos personales conforme al artículo 8 DEP (Considerando 23 RIA), debe garantizarse la aplicación del RGPD, y atenderse a los principios de necesidad, proporcionalidad, limitación de finalidad, entre otros, y garantizarse la intervención de una autoridad de independiente.

la transmisión entre las partes ni la copia de los datos brutos o estructurados, sin perjuicio de los requisitos en materia de gobernanza de datos establecidos en el presente Reglamento».

— En los supuestos de establecimiento de una base legítima de tratamiento, o en caso de levantamiento de la prohibición de tratamiento establecida en el art. 9.1 RGPD (Arts. 10.5. y 54 RIA), deben respetarse las normas previstas en garantías de los intereses y derechos fundamentales de los interesados.

— La intervención de las autoridades de protección de datos debe producirse en un ámbito de «*cooperación estructurada e institucionalizada*» con otras autoridades competentes, como las autoridades de vigilancia del mercado. Las normas que prevean la relación entre ambas en el RIA deben ser siempre interpretadas en el sentido más acorde con la independencia de aquellas y su marco de actuación conforme al RGPD.

— Las orientaciones de las autoridades de protección de datos en el ámbito de los *sandbox* han de formalizarse en un contexto de certeza jurídica, con una regulación completa y precisa también de su marco de cumplimiento.

— Si analizamos los requisitos de los SIAs de AR (Capítulo III, Sección 2, RIA) en concreto la vigilancia humana a que se refiere el artículo 14, apartados 4 y 5, podemos llegar a la conclusión que el supuesto de hecho del artículo 22 RGPD, en particular en los casos en que la decisión individual automatizada sería posible por incluirse en alguno de los casos de su apartado 2, no podría tener lugar.

— Sobre el espacio establecido en torno a la previsión de normas armonizadas, especificaciones comunes, evaluación de conformidad y certificados a que se refiere el Sección 5 del Capítulo III, es evidente que en la medida que están dirigidas a verificar el cumplimiento de las previsiones de la Sección 2, pueden facilitar la aplicación del RGPD, más aún cuando facilitará un marco de interacción normalizado entre los intervinientes, e incluso la identificación de los responsables y corresponsables de tratamiento de datos personales en la cadena de valor.

— En este contexto, podría darse contenido y posibilitar la verificación de los principios, requisitos y medidas relativas a los tratamientos de datos personales (RGPD) contenidos, como declaración, en el Considerando (69) antes citado y reproducido, en su caso mediante la modificación de los ANEXOS V, apartado 5 (relativo a la declaración de existencia de tratamientos de datos personales y el cumplimiento del RGPD), IV («*Documentación técnica a que se refiere el artículo 11, apartado 1*») y VII («*Conformidad fundamentada en la evaluación del sistema de gestión de la calidad y la evaluación de la documentación técnica*») en su caso mediante los Actos Delegados que pudiera adoptar la Comisión Europea conforme al artículo 97 RIA (en relación con los artículos 11 y 43 RIA).

— Deben promoverse la adopción de los códigos de conducta y mecanismos de gobernanza a que se refiere el artículo 95 RIA, para la aplicación voluntaria a sistemas de IA distintos de los SIA de AR, de las previsiones del Capítulo III, Sección 2, permitiendo contribuir al marco de seguridad jurídica en la aplicación del RGPD.

— En la medida en que pueden existir dificultades en la identificación de los responsables de tratamiento en el contexto de *ciclo de vida* y *cadena de valor* del SIA, y no pudiendo renunciarse al cumplimiento del RGPD, todos los agentes que en ella participan deben desplegar una especial diligencia, adoptando las medidas, también contractuales, necesarias, o decidiendo incluso la no integración del

servicio, para asegurar el cumplimiento del RGPD. Se enerva así el riesgo que supone la integración en el mismo un elemento no conforme con la norma, no sujeto a rendición de cuentas, o carentes de la documentación exigida.

A esta solución se aproxima la obligación de declarar el cumplimiento del RGPD, como contenido de la evaluación de conformidad (ANEXO V, apartado 4^a, RIA) cuando el SIA de AR a que se refiere suponga el tratamiento de datos personales.

LA INTELIGENCIA ARTIFICIAL PROHIBIDA
O INACEPTABLE PARA EL REGLAMENTO
(ARTÍCULO 5)

El reconocimiento biométrico en el Reglamento de inteligencia artificial: exenciones, prohibiciones y especialidades de alto riesgo

LEIRE ESCAJEDO SAN-EPIFANIO¹

Profesora Titular de Derecho Constitucional
en la Universidad del País Vasco/ Euskal Herriko Unibertsitatea

I. EL RECONOCIMIENTO BIOMÉTRICO EN EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL: ASPECTOS CLAVE DE SU REGULACIÓN

1. EL ENFOQUE DESDE EL QUE SE REGULA EL RECONOCIMIENTO BIOMÉTRICO EN EL REGLAMENTO

En las últimas dos décadas, las expresiones *biometrías* o *reconocimiento biométrico* vienen asociándose con sistemas de reconocimiento automatizado que tienen su base en características anatómico-físicas, fisiológicas o comportamentales de las personas². Esa asociación reciente, unida al hecho de que con frecuencia las definiciones de datos biométricos contienen ejemplos como el escaneo de huellas, de rostros o del iris, explica que al tratar el reconocimiento biométrico automatizado tiendan a pasarse por alto aspectos muy relevantes de la disciplina de la Biometría, de sus utilidades y de su estado del arte. Etimológicamente, la voz *Biometría* proviene de las voces griegas *bios* (vida) y *metron* (medir)³, y su primera definición formal, que no el primer uso del término, se atribuye a Francis Galton co-fundador en 1901 de la revista *Biometrika*⁴.

1. Este trabajo forma parte del Proyecto coordinado *AI-Biosuro — Biovigilancia mediante IA en la Era Post-Covid. Corporeidad, Identidad y Ciberseguridad* (MICINN, Proyectos de Transición Ecológica y digital, TED2021-129975B-C21), IP L., Leire Escajedo San-Epifanio; y de dos Infraestructuras, *Infraestructura Experimental mejorada para Investigación en 5G y 6G y servicios avanzados SN4E+* (Enhanced Smart Networks for Everything), e *Infraestructura Experimental para Investigación en 5G y 6G y servicios avanzados SmartNets4E* (Smart Networks for Everything), IP E.Jacob Taquet.
2. Véase Busch, C., «Biometrische Verfahren — Chancen, Stolpersteine und Perspektiven», en P. Schaar (ed.), cit., 2007, 29; Lassman, G., *Bewertungskriterien zur Vergleichbarkeit Biometrischer Verfahren*, TeleTrust Deutschland, 2002.
3. Escajedo San-Epifanio, L., *Tecnologías biométricas, Identidad y Derechos Fundamentales*, Thomson Reuters Aranzadi, 2017, 44-45.
4. Entre los primeros usos de la expresión suele citarse también a Christoph Bernoulli, quien en 1841 empleó la expresión *biometría* para referirse a tomas de medidas en se-

El término agrupa un amplísimo conjunto de métodos que permiten estudiar de forma mensurativa todo tipo de fenómenos o procesos biológicos que acontecen en los organismos vivos — sean humanos o no⁵. Por su parte, la Sociedad internacional de Biometría, fundada en 1947 y con presencia en más de 60 países, se describe a sí misma como promotora del «desarrollo y la aplicación de la teoría y de los métodos matemáticos y estadísticos a las Biociencias, incluyendo la agricultura, las ciencias biomédicas y la salud pública, la ecología, las ciencias ambientales forestales y disciplinas afines»⁶. En su haber se incluyen métodos más antiguos, desde luego, que la Inteligencia Artificial o la computación, e incluso el que la voz *Biometría*. Se ha llegado a decir que el encuentro entre las Ciencias de la Vida y la métrica representa un inmenso capítulo de la Historia de la Ciencia⁷.

Desde esa referencia amplia, y por cuanto se refiere al ser humano, en un primer acercamiento la expresión puede emplearse para referirse a *cualquier dato* obtenido a partir de propiedades biológicas, aspectos comportamentales, características fisiológicas, hábitos o acciones que han sido obtenidos mediante algún tipo de método o técnica mensurativa⁸. Y es precisamente este último detalle, el procesamiento mensurativo, el que, por encima de su vinculación de la corporeidad, caracteriza a los datos biométricos. Respecto a otros conjuntos de datos que se obtienen del cuerpo de una persona, los biométricos, no se distinguen tanto por describir con precisión «propiedades naturales» —o atributos—⁹, sino por ser una expresión del tratamiento mensurativo de éstas¹⁰.

Hasta la reciente aprobación del RIA, los datos biométricos que más atención jurídica habían recibido son aquellos que describe el Reglamento General de Protección de Datos (RGPD) en su art. 4.14. Es más, algunos textos de la literatura jurídica tienden a considerar que únicamente son datos biométricos los «*datos personales obtenidos a partir de un tratamiento técnico específico, relativos a las características físicas, fisiológicas o conductuales de una persona física que permitan o confirmen la identificación única de dicha persona [...]»* (art. 4.14 RGPD). **El caso es, sin embargo, que no todos los datos biométricos tienen esa potencialidad de permitir o confirmar la identificación única de una persona**¹².

res humanos con fines estadísticos. Vid. Saborowksi, M., «Die Pluripotenz der Biodaten. Beobachtungen zu einem Verwertungsgeschehen», en Potthast, T./ Herrmann, B./ Müller, U. (eds.), *Wem gehört der menschliche Körper?*, Mentis, Paderborn, 2010, 380.

5. Escajedo San-Epifanio, L., *Tecnologías biométricas*, cit., 2017, 27-28.

6. *Ibidem*.

7. Albrizio, A., «Biometry and Anthropometry», *Journal of Anthropological Sciences*, vol. 85, 2007, 101-123, 102-106; Abs, M., «Biometrik», 1971, 945-946; Ghilardi, G./ Keller, F., «Epistemological Foundations of Biometrics», en *Second Generation*, 24-25.

8. Se aborda esta noción infra, en II.1.

9. Saborowksi, M., «Die Pluripotenz der Biodaten», cit., 2010, 367-368.

10. Mordini, E./ Tzovaras, D./ Ashton, H., en Mordini, E./ Tzovaras, D. (eds.), *Second Generation Biometrics: The Ethical, Legal and Social Context*, Springer, 2012, 7-8.

11. Deliberadamente se excluye aquí la mención del inciso del art. 4.14 RGPD, en el que, como ejemplos de datos biométricos, se mencionan las imágenes faciales y los datos dactiloscópicos. Sobre ello se volverá en II.2.

12. Kindt, E., «Having yes, using no? About the new legal regime for biometric data», *Computer Law & Security Review*, 34 (3), 2018, 523-538.

La noción del art. 4.14 RGPD es coherente con un tiempo en el que las tecnologías de reconocimiento de identidad, en especial la autenticación biométrica, dominaban el panorama de tecnologías implementadas en la vida real. En la actualidad, sin embargo, son muchas las tecnologías de reconocimiento que ofrecen utilidades de reconocimiento no singulares; así, por ejemplo, a los efectos de determinar la edad de una persona, su estado de salud o nivel de estrés o de precisar las emociones por las que transita. En tanto estos últimos sistemas no tienen potencialidad para servir de base a una identificación única, pueden encajar en el concepto de datos personales, pero no, propiamente, en la noción del art. 4.14 RGPD¹³.

En ausencia de un estatuto jurídico claro respecto de estos datos biométricos no identificantes, los legisladores se vieron en la necesidad de acometer esta tarea en el proceso de elaboración del RIA, que junto con las utilidades biométricas de identificación se referirá a las de reconocimiento de emociones y a la categorización. Comisión¹⁴, Parlamento¹⁵ y Consejo¹⁶ ofrecieron diferentes alternativas para completar la limitada noción del art. 4.14 RGPD a los efectos del RIA, pero, como se verá en el apartado II.2, en algún momento de los trílogos terminó imponiéndose una nueva propuesta.

Además de precisar esta noción de datos biométricos, los legisladores vieron dar respuesta a otras cuestiones relativas a los sistemas de reconocimiento biométrico que no tenían precedentes regulatorios adecuados. Sin ánimo de exhaustividad, escogiendo las más relevantes a los efectos de esta exposición, los legisladores debían decidir si el RIA comprendería todos los reconocimientos biométricos automatizados imaginables (desde la verificación e identificación, hasta el reconocimiento de emociones, las categorizaciones en sentido amplio y los cribados) o si, por el contrario, sería descartada esa ambición holística.

En segundo lugar, y atendiendo a que el RIA es, en esencia un reglamento de evaluación de riesgos previa al mercado, existía la incógnita de qué tipos de reconocimiento biométrico quedarían en todo caso excluidos del acceso al mercado y cuáles serían tratados como modalidades de alto riesgo.

13. Relativo al tratamiento de las categorías especiales de datos.

14. La principal referencia manejada para la interpretación de la toma de postura de la Comisión, salvo que se indique otra cosa, ha sido la *Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de Inteligencia Artificial*, {SEC(2021) 167 final} —{SWD(2021) 84 final}— {SWD(2021) 85 final}.

15. Como referencia para la postura del Parlamento Europeo se han manejado las *Enmiendas aprobadas por el Parlamento Europeo el 14 de junio de 2023 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial)* y se modifican determinados actos legislativos de la Unión (COM (2021)0206 —C9-0146/2021— 2021/0106(COD)).

16. Como texto principal para la postura del consejo se ha manejado *el texto de compromiso de la Cuarta Presidencia sobre la Propuesta de Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas sobre la inteligencia artificial (Ley de Inteligencia Artificial)* y se modifican determinados actos legislativos de la Unión, Bruselas, 19 de octubre de 2022. Expediente interinstitucional 2021/ 0106 (COD).

Por último, en tercer lugar, quedaba por precisar lo que habría de suceder con los sistemas de reconocimiento biométrico (la mayor parte de gestión de identidades nacionales o de identificaciones en frontera) que, antes de la entrada en vigor del RIA —cosa que no ha sucedido aún a la entrega de este trabajo, ya eran manejados por los Estados Miembros en sus territorios o, incluso, a nivel comunitario como parte del *espacio de libertad, seguridad y justicia* (como el sistema SIS o EURODAC)¹⁷.

2. ANTECEDENTES RELEVANTES: LA DISTANCIA ENTRE EL LIBRO BLANCO Y LAS RESOLUCIONES DEL PE SOBRE ALGUNAS MODALIDADES DE RECONOCIMIENTO BIOMÉTRICO

Por cuanto se refiere al reconocimiento biométrico, la distancia entre el Libro Blanco sobre la Inteligencia Artificial (2020)¹⁸ presentado por la Comisión y algunas posturas previas del Parlamento sobre reconocimiento biométrico, entre ellas dos resoluciones de 2021, hacían prever una complicada búsqueda de acuerdos en la tramitación del RIA.

La Comisión, en el Libro Blanco, señalaba la identificación biométrica remota como un ejemplo de aplicaciones de IA que, independientemente del sector en que se aplicasen, cabría pensar que «son de un nivel de riesgo elevado»¹⁹. En coherencia, su propuesta de Reglamento recogerá la mayoría de los sistemas de identificación biométrica remota en las modalidades de alto riesgo. No debe llevar a confusión el hecho de que esa misma propuesta recogiera la prohibición del «uso» —que no de introducción en el mercado— de algunas modalidades de identificación biométrica remota entre las prácticas a prohibir de su propuesta de artículo 5.1.d). El amplio número de supuestos que podían acogerse a las salvedades que la Comisión admitía respecto esa prohibición prácticamente vaciaba de contenido a esta última²⁰.

Se producirá así una importante paradoja, porque el RGPD prohíbe con pocas excepciones el tratamiento de datos biométricos identificantes —imprescindibles en los sistemas de identificación biométrica remota—, y la propuesta de RIA parece abrirlas la «posibilidad de uso» como modalidades de alto riesgo. La confusión aumenta si tenemos en cuenta que la Comisión, en el Libro Blanco —y posteriormente en la Exposición de Motivos de la propuesta— recordará que, de conformidad con las normas con las normas vigentes en materia de protección de datos y *con la Carta de Derechos Fundamentales de la UE*, la IA solo puede utilizarse con fines de identificación biométrica remota cuando dicho uso esté debidamente justificado, sea proporcionado y esté sujeto a garantías adecuadas²¹. A la vista de tal reconocimiento, parece inoportuno que RIA y RGPD, respetando el objetivo de cada una de tales normas, no se coordinen adecuadamente en el tratamiento del «uso» de sistemas de identificación biométrica remota.

17. Véase *infra*, VI.

18. Comisión Europea, *Libro Blanco sobre la inteligencia artificial — un enfoque europeo orientado a la excelencia y la confianza*, Bruselas, 19.2.2020, COM (2020) 65 final.

19. Comisión Europea, *Libro Blanco*, cit., 2020,22.

20. Véase *infra*, III y IV, sobre los sistemas de reconocimiento comprendidos entre las prácticas prohibidas del art. 5.

21. Comisión Europea, *Libro Blanco*, cit., 2020,26-27.

En cualquier caso, y en coherencia con su posición sobre el RIA, los esfuerzos de la Comisión respecto a las tecnologías de reconocimiento biométrico en general y las modalidades prohibidas en particular, se centrarán en clasificarlas por nivel de riesgo, delimitar con claridad cuáles pueden ser los supuestos de uso justificado y proporcionado, y recoger algunas de las garantías aplicables. Es oportuno apuntar, además, que el Libro Blanco propondrá una distinción entre, de una parte, la identificación biométrica remota y, de otra, la autenticación. Este hecho resulta insólito porque el RGPD, ha de notarse, ya prohibía el uso de datos biométricos con fines de identificación única, exceptuando casos que reunieran condiciones muy específicas (art. 9.1 y 9.2. RGPD), con independencia de que fueran empleados en funcionalidades de verificación o «uno-a-uno» (1 a 1) o en identificaciones «uno-a-muchos» (1 a n). La Comisión, sin embargo, únicamente avanza en el Libro Blanco que las utilidades de identificación recibirán un tratamiento diferente a las de identificación, sin precisar más. La distinción resultará siendo muy relevante dado que, en el texto final del RIA, como se verá, las verificaciones biométricas —interpretadas en una forma muy amplia— terminarán siendo excluidas, incluso, de las modalidades de alto riesgo²², quedando como mucho al albur de posibles Códigos de Conducta voluntarios²³.

Por cuanto se refiere al Parlamento Europeo, este publicó dos importantes resoluciones previas al RIA en las que se recogió un amplio rechazo a la mayor parte de las modalidades de reconocimiento biométrico, en especial cuando fueran empleadas por parte de las autoridades policiales.

En la Resolución del Parlamento Europeo de 6 de octubre 2021²⁴, relativa al uso de IA por las autoridades policiales y judiciales en asuntos penales. El Parlamento considera, en primer lugar, que el despliegue de sistemas de reconocimiento facial debe limitarse a fines claramente justificados y hacerse con pleno respeto a de los principios de proporcionalidad y necesidad y de la legislación aplicable. Asimismo, y en segundo lugar, pide una prohibición permanente del uso de análisis automatizados o el reconocimiento en espacios accesibles al público de características humanas como los andares, las huellas dactilares, el ADN, la voz y otras señales biométricas y comportamentales. Y por último, en tercer lugar, reclama una moratoria en el uso del reconocimiento facial hasta que se den las siguientes circunstancias: que las normas técnicas puedan considerarse plenamente acordes con los derechos fundamentales; que los resultados no estén sesgados ni sean discriminatorios; que el marco regulador sea estricto; y que existan pruebas empíricas de la necesidad y proporcional del despliegue de estas tecnologías, con la única excepción del caso en el que se usen estrictamente para la identificación víctimas de delitos²⁵.

Ha de tenerse en cuenta, además, que antes de esa fecha el Parlamento Europeo ya había recomendado una prohibición del uso de aplicaciones de reconocimiento

22. Vid. infra. II.3. La verificación biométrica.

23. Vid. infra. V. Modalidades de alto riesgo.

24. Resolución del Parlamento Europeo, de 6 de octubre de 2021, *sobre la inteligencia artificial en el Derecho penal y su utilización por las autoridades policiales y judiciales en asuntos penales*, (2020/2016(INI)), publicado en el Diario Oficial C 132/ 17, de 23 de marzo de 2022.

25. Rostalski, F./ Weiss, E., «Verbotene KI-Praktiken», en Hilgendorf, E./ Roth-Isigkeit, D. (eds), *Die neue Verordnung der EU*, cit., 2023, 47-48.

biométrico automatizado, como el reconocimiento facial con fines educativos y culturales, en especial respecto a los menores de edad, a menos que su uso estuviera expresamente autorizado por Ley²⁶. Esta reivindicación de restricciones y moratorias en relación con el reconocimiento biométrico automatizado caracterizará la postura del Parlamento frente al Consejo hasta el último suspiro de la aprobación del RIA.

3. ESQUEMA REGULATORIO DE LAS TECNOLOGÍAS DE RECONOCIMIENTO BIOMÉTRICO EN EL REGLAMENTO

El reconocimiento biométrico automatizado se recoge a lo largo del RIA en un importante número de considerandos, apartados de las definiciones y artículos sustantivos. En ese último conjunto destacan especialmente cinco estatutos jurídicos diferentes, aplicables a determinados conjuntos de tecnologías de reconocimiento biométrico.

Encontramos en primer lugar, una serie de prácticas biométricas prohibidas en el único artículo del capítulo II (el art. 5. RIA). Un segundo estatuto es el de las prácticas biométricas consideradas de alto riesgo (arts. 6 y siguientes, con el complemento del Anexo III). En tercer lugar, han de señalarse una serie de prácticas biométricas que, como consecuencia de la previsión del art.2.3, quedan excluidas del ámbito de aplicación de RIA. En cuarto lugar, atendiendo al artículo 111 RIA con el complemento del Anexo X, se prevé un estatuto específico para un conjunto de prácticas de reconocimiento biométrico que se emplean en el ámbito de grandes sistemas informáticos instaurados mediante actos legislativos de la UE en asuntos policiales y de control de fronteras. Por último, y en quinto lugar, una interpretación sistemática del RIA hace aflorar un conjunto de sistemas de reconocimiento biométrico que, por la escasa o nula atención que reciben en el RIA, parece que quedan fuera de su ámbito de aplicación o, al menos, en duda.

Respecto al primero de los conjuntos, el de las modalidades de reconocimiento biométrico afectadas por las prohibiciones del RIA, son seis los grupos de prácticas que aparecen en el listado del art. 5²⁷:

1. algunos de los sistemas biométricos que puedan utilizarse para *evaluar o clasificar* a personas físicas o a colectivos de personas *atendiendo a su comportamiento social o a características personales o de su personalidad* conocidas, inferidas o predichas (art. 5.1.c);
2. algunos de los sistemas biométricos que, mediante el *perfilado o la evaluación de la personalidad*, puedan emplearse para realizar evaluaciones de riesgos de personas físicas *con el fin de valorar o predecir el riesgo de comisión de delitos* (art.5.1. d);
3. algunos sistemas de reconocimiento biométrico que pueden ser empleados en la *creación o ampliación de determinadas bases* de reconocimiento facial (art.5.1. e);
4. algunos sistemas de reconocimiento que permitan *inferir las emociones* de una persona física *en los lugares de trabajo y en los centros educativos* (art. 5.1. f);
5. algunos sistemas de categorización biométrica que clasifiquen individualmente a las personas físicas sobre la base de sus datos biométricos para *deducir o inferir su*

26. Punto 45 de la Resolución del Parlamento Europeo, de 19 de mayo de 2021, sobre la inteligencia artificial en los sectores educativo, cultural y audiovisual (2020/2017(INI)).

27. Véanse infra, III y IV.

raza, opiniones políticas, afiliación sindical, convicciones religiosas o filosóficas, vida sexual u orientación sexual (art. 5.1. g).

y 6. algunos de *identificación biométrica remota «en tiempo real»* en espacios de acceso público con fines de garantía del cumplimiento del Derecho.

Un segundo conjunto de tecnologías biométricas, al tenor literal del artículo 6 y, en especial, del Anexo III del Reglamento, es clasificado como *sistemas de alto riesgo*²⁸. Este conjunto comprende, como primer subgrupo, los sistemas de identificación biométrica remota, los sistemas de categorización biométrica basados en atributos o características sensibles y determinados sistemas de reconocimiento de emociones, todos ellos con independencia del escenario operativo en que se apliquen, siempre y cuando no estén entre las categorías prohibidas. Se excluyen, en cualquier caso: los sistemas de reconocimiento abarcados en las definiciones de IA prohibida; los sistemas que los Estados Miembros utilicen con fines militares, de defensa o de seguridad nacional (art. 2.3 RIA); y, según indica el propio Anexo III, los sistemas de reconocimiento que ofrezcan funcionalidades de autenticación o verificación.

El segundo subgrupo de reconocimientos biométricos de alto riesgo es el de los sistemas que cabe considerar incluidos en los apartados 2 a 8 del anexo III RIA. En esos apartados del anexo se ofrece un listado de veintiuna modalidades de IA consideradas de alto riesgo, agrupadas en seis escenarios operativos: ámbito de la educación y formación profesional; ámbito laboral; servicios y prestaciones esenciales; garantía de cumplimiento del Derecho; tránsito transfronterizo; y administración de justicia²⁹. Los enunciados de estas modalidades de IA de alto riesgo se han formulado empleando con insistencia la expresión «*sistemas de IA destinados a ser utilizados para*» acciones tales como evaluar (riesgos, resultados, niveles de aprendizaje, fiabilidad), realizar seguimientos, detectar comportamientos prohibidos, clasificar o tomar decisiones, y, como se detallará, pueden abarcar potencialmente algunas funcionalidades de reconocimiento biométrico no recogidas en el punto 1 del anexo III.

Desde el punto de vista de la perspectiva legislativa, este enunciado, genera inseguridad jurídica y de algún modo, contradice en apariencia el modelo regulatorio del RIA, horizontal y centrado en el riesgo. Así, en relación con esta última cuestión, se ha de señalar que: algunas de las previsiones de estos listados, organizados por escenarios operativos, se solapan con el subgrupo del apartado 1º del anexo III, haciendo innecesaria la reiteración; y en otras previsiones, por su parte, se plantea un solapamiento con la necesaria evaluación de proporcionalidad que, respecto a los datos identificantes, establece el RGPD.

El tercer conjunto importante de tecnologías biométricas queda afectado por las previsiones de los artículos 111 y siguientes RIA, en conexión con el Anexo X. Se trata de sistemas informáticos de gran magnitud que han sido puestos en servicio o lo estarán antes de que transcurran 36 meses de la fecha prevista para la entrada en vigor de la Ley de IA³⁰. Como se detallará en el apartado VI, se trata de sistemas

28. Véase infra, en V.

29. Detalladamente sobre los sistemas de alto riesgo, véase en esta obra el comentario a los artículos 6 y siguientes a cargo de L. Cotino Hueso.

30. En las disposiciones finales se establece que el Reglamento se aplicará plenamente a los tres años de su publicación en el Diario Oficial, publicación esta que a la fecha de redacción de este trabajo aún no se ha producido.

de reconocimiento de gran magnitud regulados por actos legislativos de la UE que, además, que se emplean ya o están en proceso de implantación para la gestión del espacio de libertad, seguridad y justicia. La previsión es que, con excepción de la aplicación las prohibiciones previstas en el artículo 5.1 RIA —todavía por determinar en qué forma—, estos grandes sistemas disfruten de una moratoria temporal de la aplicación del RIA, extensible como se verá hasta prácticamente enero de 2031. Pasado ese plazo, además, deberá afrontarse el hecho de que otros preceptos del RIA dejan potencialmente fuera de su ámbito de aplicación algunas utilidades de verificación y de identificación no remota que son ofrecidas por estos sistemas.

En quito y último lugar, tal y como se ha advertido, una interpretación sistemática del RIA hace aflorar un quinto estatuto aplicable a determinadas modalidades de reconocimiento biométrico: el de las excluidas del RIA. Además de la citada situación de la verificación y la identificación no remota, que se abordarán en el apartado II, el texto final del art.2.3 RIA, relativo al ámbito de aplicación, indica que el RIA no se aplicará a los sistemas de IA que, y en la medida en que, se introduzcan en el mercado, se pongan en servicio o se utilicen, con o sin modificaciones, exclusivamente con fines militares, de defensa o de seguridad nacional, independientemente del tipo de entidad que lleve a cabo estas actividades. Estos sistemas, no obstante, sí están sometidos al RGPD y sus actos derivados.

II. ALGUNOS CONCEPTOS CLAVE EN LA TIPIFICACIÓN DE LAS MODALIDADES DE RECONOCIMIENTO BIOMÉTRICO EN EL ENUNCIADO DEL REGLAMENTO: DATOS BIOMÉTRICOS Y VERIFICACIÓN BIOMÉTRICA

El RIA recoge en su artículo 3 un largo listado de definiciones. Algunas de ellas, como identificación biométrica remota (en tiempo real y en diferido), o categorización y perfilado serán objeto de atención como parte de la tipificación de supuestos de hecho que ofrece el RIA. Pero existen otras nociones que tienen una relevancia transversal, sobre el conjunto de artículos que se refieren al reconocimiento biométrico, y que, por tal motivo, merecen recibir una atención en estos primeros apartados. Es el caso de la noción de «datos biométricos» que emplea el RIA —apartándose del art. 4.14 RGPD—, y de la noción de «verificación biométrica», como categoría que con insistencia el RIA trata de distanciar del concepto de identificación biométrica, en especial de la modalidad remota.

1. ¿UNA NUEVA NOCIÓN DE «DATOS BIOMÉTRICOS» PARA ABARCAR CON CLARIDAD LAS BIOMETRÍAS NO SINGULARIZANTES?

1.1. La situación previa a la aprobación del Reglamento: la noción de datos biométricos del RGPD versus la noción científico-técnica

Se advertía en la introducción que, hasta la aprobación del RIA, la noción jurídica más relevante de «datos biométricos» era la recogida en el número 14, artículo 4, del RGPD. El último inciso del art. 4.14 RGPD será excluido de nuestra reflexión, porque resulta un tanto desafortunado. En él, los legisladores presentan como ejemplos de

datos biométricos las imágenes faciales y los datos dactiloscópicos³¹. No obstante, por sí mismos, ni la foto de un rostro ni la información dactiloscópica —incluyendo una impresión de huellas³²— son en puridad datos biométricos. Son, cierto es, posibles fuentes de datos biométricos, pero será un tratamiento mensurativo lo que determinará si se obtienen o no de ellas datos biométricos singularizantes³³.

Excluyendo, por tanto, ese inciso, cabe decir que, conforme al RGPD son datos biométricos los «*datos personales obtenidos a partir de un tratamiento técnico específico, relativos a las características físicas, fisiológicas o conductuales de una persona física que permitan o confirmen la identificación única de dicha persona*» (art. 4.14 RGPD). **Y procede ahora realizar una comparación entre esa noción, a partir de la cual el RGPD establece una categoría de datos de especial protección, y la definición de datos biométricos coherente con la entidad de la disciplina de la Biometría, presentada en el apartado I.1 de este trabajo.**

Antes incluso de la aprobación del RGPD, diferentes expertos venían advirtiendo ya sobre la distancia entre, de una parte, las nociones de datos biométricos recogidas en diferentes documentos jurídicos, fueran vinculantes o no, y, de otra parte, el concepto de datos biométricos en perspectiva científica³⁴. Como señalaron Kindt y Jasserand, existía una notable desalineación entre las posibilidades tecnológicas y la definición legal³⁵, problema que vio agudizado con la aprobación del RGPD.

Se propone aquí la siguiente definición para la noción científico-técnica de datos biométricos postulando, además, que es deseable que progresivamente vaya siendo tenido en cuenta al diseñar los marcos reguladores de las técnicas de reconocimiento biométrico³⁶:

Son datos biométricos aquellos que se obtienen del cuerpo de las personas como expresión de algún tipo de estudio mensurativo y, en su vasto conjunto, resulta

31. Sumer, B. «When do the images of biometric characteristics qualify as special categories of data under the GDPR: a systematic approach to biometric data processing», BIOSIG 2922 — International Conference of the Biometrics Special Interest Group, publicado en acceso abierto en IEE Xplore; Romeo Casabona, C., «Datos biométricos (Comentario al artículo 4.13 RGPD)», en *Comentario al Reglamento General de Protección de Datos y a la Ley Orgánica de Protección de Datos personales y Garantía de los Derechos Digitales*, A. Troncoso Reigada (dir.), Vol. 1, 2021, 709-714.
32. Véase sobre este concepto de fuente de datos biométricas infra, II.2.
33. Kindt, E., Having yes, using no? About the new legal regime for biometric data, *Computer Law & Security Review*, Volume 34, Issue 3, 2018, 523-5388.
34. Jasserand, C. A., «Avoiding Terminological Confusion between the Notions of “biometrics” and “biometric Data”: An Investigation into the Meanings of the Terms from a European Data Protection and a Scientific Perspective», *International Data Privacy Law* 6 (1), 2015.
35. Jasserand, C. A., «Avoiding Terminological Confusion», cit., 2015; Kindt, E., «Having yes, using no?», cit., 2018, 523-538.
36. Kindt, por su parte, propone definir los datos biométricos como «*todos los datos personales que (a) se relacionan directa o indirectamente con características biológicas o de comportamiento únicas o distintivas de los seres humanos y (b) se utilizan o son aptos para ser utilizados por medios automatizados (c) para fines de identificación, verificación de identidad o verificación de una reclamación de personas físicas vivas*», si bien por «personales» se refiere a procedentes de personas y, no por tanto, a datos que encajen en la definición de datos personales del RGPD. Vid. Kindt, E. *Privacy and Data Protection Issues of Biometric Applications —A Comparative Legal Analysis*, Springer, 2013, 11.

*relevante la distinción de dos grandes grupos, según se trate de datos que sido obtenidos de biometrías estáticas o de biometrías dinámicas*³⁷.

Las biometrías estáticas agrupan aquellos métodos que captan información métrica puntual a partir de las características anatómico-físicas de un cuerpo humano. Las biometrías dinámicas, por su parte, comprenden los métodos que se aplican para captar de un cuerpo humano información secuencial o cíclica de las habilidades motoras, así como de las señas y parámetros corporales en sentido amplio, ya sea en sus dimensiones externa o interna, voluntaria o involuntaria.

Nótese, porque la distinción tendrá relevancia, que las biometrías estáticas y las dinámicas se diferencian por dos aspectos clave: en primer lugar, por el tipo de fuentes de las que obtienen información (características anatómico-físicas versus, de algún modo, el cuerpo en funcionamiento); y en segundo lugar por el hecho de que la captación de las biometrías estáticas puede realizarse en un instante, mientras que la captación de las biometrías dinámicas requiere un mayor o menor lapso de tiempo. Dicho de otro modo, las biometrías que captan el patrón de una huella dactilar o un iris, así como la geometría de un rostro, actúan sobre un dato en crudo puntual. Se extrae, por ejemplo, una impresión dactilar de los dermatoglifos de un dedo y esa impresión es suficiente como para dar comienzo al proceso de biometrización de los atributos que singularizan esa huella respecto a otras. Las tecnologías basadas en biometrías dinámicas, en cambio, como es caso de las que analizan el espectro de la voz de una persona, el patrón de su marcha al caminar o la velocidad de sus latidos, son tecnologías que necesitan poder captar información de la fuente durante un lapso más o menos largo de tiempo.

1.2. La necesidad de una noción funcional que abarque los datos biométricos con y sin potencial identificante, sean personales o no

Según la Exposición de Motivos de la Propuesta del RIA, este reglamento viene a completar —sin desplazar— el marco de garantías que ya recogen, entre otros, textos jurídicos como el RGPD. Respecto a este último, dice también la Exposición de Motivos, el RIA viene a ser una *lex specialis*, pero con un carácter fundamentalmente complementario. A pesar de esa afirmación, sin embargo, en el caso concreto del uso que se da en el RIA a la noción de datos biométricos, con una definición propia, la diferencia respecto al RGPD es muy relevante³⁸. Se describirá tal diferencia en este apartado, dejando para el apartado II.1.3 tanto el modo en que Comisión, Parlamento y Consejo propusieron respectivamente atender a esta circunstancia, como la opción que se incorporó finalmente al RIA.

Aunque repite el criticado inciso final del art. 4.14 RGPD al que nos hemos referido ya en II.1.1., la noción de datos biométricos que emplea el RIA se distingue de aquella en que se ha eliminado la exigencia de que se trate de datos que permitan o confirmen «la identificación única de una persona». **En consecuencia, en el conjunto de datos que pueden emplearse en los sistemas de reconocimiento biométrico habrá datos sujetos**

37. Escajedo San-Epifanio, L., *Tecnologías biométricas*, cit., 2017,100-101.

38. Czarnocki, J., «Will new definitions of emotion recognition and biometric data hamper the objectives of the proposed AI Act?», *International Conference of the Biometrics Special Interest Group (BIOSIG)*, Darmstadt, Germany, 2021, 1-4.

al ámbito de aplicación de RGPD en tanto puedan considerarse personales, siendo algunos de ellos, además, pertenecientes a la categoría de especial protección (art. 9 RGPD). Por cuanto respecta al RIA, por su parte, algunas dimensiones del tratamiento de datos biométricos quedarán sujetas a su aplicación —con independencia de que se trate o no de datos singularizantes e, incluso, sin necesidad de que se trate de datos personales—, y, al mismo tiempo, los sistemas de IA de verificación que emplean datos biométricos de categoría especial (arts. 4.14 y 9 del RGPD) quedarán excluidos del RIA³⁹. En términos de garantías, ha de decirse, esta sistemática —con el añadido de la complicada redacción de algunos artículos del RIA— resulta bastante preocupante.

En este contexto se hace necesario realizar una serie de anotaciones sobre las características que deben reunir los sistemas de reconocimiento biométrico que pretenden ofrecer una funcionalidad de identificación (o singularización de identidad), frente a las características de aquellos sistemas que captan otro tipo de informaciones biométricas no singularizantes:

a) Características de los sistemas de reconocimiento biométrico con potencial singularizante

Los sistemas de reconocimiento con potencial de identificación única sólo pueden asentarse sobre conjuntos de datos biométricos que reúnen una serie de características⁴⁰, entre las que destacan la universalidad, el potencial singularizante, la inherencia o la permanencia. Un sistema de reconocimiento biométrico con potencial singularizante debe asentar su base en una característica o atributo que sea *universal*, en el sentido de que la fuente de información biométrica esté, en principio, disponible en todos los seres humanos (o con muy pocas excepciones).

A esa universalidad ha de añadirse, además, que la fuente de información debe permitir captar en ella la *presencia de elementos singularizantes*. Los dedos de las manos tienen una presencia prácticamente universal y la mayoría de las personas tienen en ellos dermatoglifos, o pliegues de la piel, que dejan huellas dactilares en su contacto con las superficies. Un sistema de reconocimiento biométrico que cuantifique el número de dedos en la mano derecha no resulta lo suficientemente singularizante, por la elevada coincidencia de la cantidad de cinco dedos en la mayoría de las manos humanas. El caso de los dermatoglifos, en cambio, es diferente. Se trata de una fuente de información en la que podemos captar una singularidad tan elevada que, estadísticamente, se considera improbable que existan dos dedos —incluso en la misma persona— con una huella idéntica. Eso hace que la biometría dactilar sea de un elevado potencial singularizante, potencial que además puede captarse. Esa circunstancia contrasta, por ejemplo, con el ADN, que siendo muy singularizante en los humanos, en la actualidad no puede captarse y procesarse de forma automatizada.

Para servir de soporte a un sistema de identificación, por último, es también muy relevante que la información biométrica de referencia sea una información *inherente* al cuerpo de la persona —o al menos razonablemente difícil de modificar o suplantar

39. Sobre ello, vid. II.3.

40. Análisis detallado de dichas características y de planteamientos como el de los llamados *Seven Pillar sor Biometrics Wisdom*, con detalladas referencias, en Escajedo San-Epifanio, L., *Tecnologías biométricas*, cit., 2017, 88-98.

a voluntad— y que *permanezca* a pesar de que pase el tiempo⁴¹. La permanencia no se exige en términos absolutos, sino más bien en el sentido de que tenga una estabilidad suficiente como para que pueda ser útil para re-identificar a la persona transcurrido un período razonable de tiempo. El patrón de pabellón auricular, por ejemplo, permanece a pesar de que el tamaño de la oreja varíe con la edad. Puede ponerse también el ejemplo de la imagen del DNI, que no es biométrica pero sí fuente de datos. Esta imagen, por los requisitos con los que es tomada, se estima que razonablemente puede servir para re-identificar a alguien dentro de un período de diez años desde su emisión o, incluso, durante más tiempo cuando la persona ha alcanzado cierta edad; de ahí la emisión de un DNI permanente a partir de los 70 años.

Cabe decir, en este sentido, que la permanencia permite garantizar que el esfuerzo de desplegar un sistema de identificación biométrica en sentido estricto se verá compensado con la posibilidad de emplearlo de forma dilatada en el tiempo. Respecto a otros atributos biometrizable, la universalidad y facilidad de captación contrastan con la escasa permanencia de los datos obtenidos. Condiciones ambientales y/o temporales impactan de forma relevante en esos datos⁴². Por ejemplo, el peso de una persona o la longitud de su cabello (salvo que carezca de éste) son fáciles de medir, pero en una misma persona están sometidos a una importante variabilidad a lo largo de su vida. No tendría sentido construir un sistema de identificación única sobre la base de datos de esas características. Los dermatoglifos, sin embargo, terminan de formarse hacia el tercer o cuarto mes del desarrollo fetal y se mantienen —salvo lesiones externas— a lo largo de toda la vida de la persona, pudiendo incluso ser captadas en algunos estadios post-mortem. De hecho, la disciplina científica de la necropapiloscopia dispone de técnicas para la captación que pueden, incluso, aplicarse a dedos petrificados o momificados⁴³.

b) Características de los datos que no persiguen una identificación única, sino la captación de informaciones de otra utilidad

Por cuanto se refiere a las categorizaciones no identificantes y, en especial, el reconocimiento de emociones, los sistemas de reconocimiento que aplican este tipo de modalidades tienden a emplear biometrías débiles (*soft biometrics*), con bajo rango de permanencia y potencial identificante⁴⁴. Las fuentes, eso sí, son fuentes universales, y tienen capacidad para captar los atributos en una forma categorizante, en el sentido de asociar, de una parte, los atributos captados en fresco respecto a una persona con, de otra parte, patrones promedio que se han elaborado como rango de referencia para cada una de las categorías en las que se pretende clasificar a los humanos (por ejemplo, clasificación por rangos de edad, interpretación de emociones básicas o detección de comportamientos sospechosos).

41. European Commission/ DG JRC/ Institute of Prospective Technological Studies (IPTS), *Biometrics at the Frontiers: Assessing the Impact on Society*, European Commission, 2005, p.37; Mordini, E./ Massari, S., «Body, Biometrics and Identity», loc. cit., 2012; ZHANG, D./ LU, G. *3D Biometrics. Systems and Applications*, Springer, 2013, 9-12.

42. Escajedo San-Epifanio, L., *Tecnologías biométricas*, cit., 2017, 93-96.

43. Alegretti, J. C./ Brandimarti de Pini, N. M., «Necropapilospía. Identificación de cadáveres y restos humanos», en *Tratado de papiloscopia*, La Rocca, Buenos Aires, 2007, 245-263.

44. Escajedo San-Epifanio, L., *Tecnologías biométricas*, cit., 2017, 105-106.

El caso de las emociones es, sin duda, uno de los más sencillos de explicar. El sistema ha sido adiestrado, por ejemplo, para reconocer en los rostros un rango de movimientos en ojos, cejas, frente y barbilla que, en amplios conjuntos de personas, suele asociarse con una alta probabilidad con emociones como la ira, la alegría o el miedo. Cada vez que se sitúe un rostro frente al sistema, éste buscará en él una emoción, tratando de hallar una coincidencia razonable respecto a alguna de las categorías biométricamente predefinidas que tiene disponibles. Una información de este tipo puede, en su caso, servir de guía para persuadir a alguien sobre la adquisición de determinados productos o servicios.

1.3. Evolución de la noción de datos biométricos en el proceso de elaboración del Reglamento

Por si en algún momento se revisa la descoordinación actual entre las nociones de datos biométricos respectivamente previstas en el RIA y el RGPD, resulta de interés señalar cómo se fraguó la noción finalmente incorporada el RIA. Comisión, Parlamento y Consejo ofrecieron diferentes alternativas, que se analizarán, aunque, sorpresivamente, en algún momento de los trólogos terminó imponiéndose una noción que dice inspirarse en el RGPD, pero, como se ha avanzado ya, en realidad se aleja de éste.

El considerando 7º de la Propuesta de la Comisión recogía que su noción de «datos biométricos» coincidía con la noción de «datos biométricos» definida en el artículo 4, punto 14, del Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo⁴⁵; en el artículo 3, punto 18, del Reglamento (UE) 2018/1725 del Parlamento Europeo y del Consejo⁴⁶; y en el artículo 3, punto 13, de la Directiva (UE) 2016/680 del Parlamento Europeo y del Consejo⁴⁷. Se indicaba, en coherencia, que tal noción debía interpretarse en consonancia con la del RGPD. En sí misma la noción no planteaba dudas. Resultaba problemático, no obstante, que ni esta ni otras nociones atendieran al hecho de que mediante la IA era posible el tratamiento de datos biométricos no identificantes.

45. Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos) (DO L 119 de 4.5.2016, p. 1).

46. Reglamento (UE) 2018/1725 del Parlamento Europeo y del Consejo, de 23 de octubre de 2018, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por las instituciones, órganos y organismos de la Unión, y a la libre circulación de esos datos, y por el que se derogan el Reglamento (CE) n.º 45/2001 y la Decisión n.º 1247/2002/CE (DO L 295 de 21.11.2018, p. 39).

47. Directiva (UE) 2016/680 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por parte de las autoridades competentes para fines de prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de sanciones penales, y a la libre circulación de dichos datos y por la que se deroga la Decisión Marco 2008/977/JAI del Consejo (Directiva sobre protección de datos en el ámbito penal) (DO L 119 de 4.5.2016, p. 89).

El Parlamento, por su parte, sí atendió esta última necesidad. La noción de datos biométricos que adoptó en su postura coincidía literalmente con la de la Comisión, pero propuso añadir un nuevo concepto al listado de definiciones del art. 3 RIA: el concepto de *datos de base biométrica*. Esta segunda noción, diferenciada de la noción de *datos biométricos*, se propone como fórmula para completar a los efectos del RIA las limitaciones funcionales del art. 4.14 RGPD. Los datos de base biométrico, conforme a la enmienda del Parlamento Europeo que propone su inclusión como número 33 bis del art. 3, son definidos como «*datos obtenidos a partir de un tratamiento técnico específico relativos a las señales físicas, fisiológicas o conductuales de una persona física*»⁴⁸.

Este segundo concepto es, sin embargo, descartado en la última versión del RIA, estimándose preferible utilizar una única noción de datos biométricos para todo el texto. Con el objetivo de reducir la inseguridad jurídica, en cualquier caso, se procede a modificar la propuesta inicial de noción de datos personales realizada por la Comisión, respaldada en las tomas de postura del Consejo y el Parlamento. Así, y según el punto 34 del artículo 3 RIA, a los efectos de dicho texto serán datos biométricos «*los datos personales obtenidos a partir de un tratamiento técnico específico, relativos a las características físicas, fisiológicas o conductuales de una persona física, como imágenes faciales o datos dactiloscópicos*». Dicen los considerandos que esa noción debe interpretarse a la luz del concepto de datos biométricos del artículo 4.14 RGPD, pero las diferencias entre ambas nociones son obvias.

2. MÍNIMO COMÚN DE TODOS LOS SISTEMAS DE RECONOCIMIENTO BIOMÉTRICO

2.1. Sistemas de reconocimiento biométrico

Sea identificante o no, para que un dato biométrico pueda servir de base a un sistema de reconocimiento biométrico, es necesario construir un sistema. En concreto, un sistema con capacidad de examinar el cuerpo de las personas y «reconocer» en él, sea de forma puntual sea de forma secuencial, algún atributo que pueda vincularse con la información que, como referencia, ha sido almacenada previamente — en una base de datos o en dispositivos de almacenamiento externo, como los chips que portan algunos documentos identificativos. Hubo sistemas de reconocimiento biométrico no automatizados, como el mítico *Bertillonage* al que nos referiremos brevemente *infra*, pero en la actualidad el término sistema de reconocimiento biométrico se aplica por lo común a estructuras que combinan *Hardware* y *Software*, cada vez con más tendencia a ser impulsados por IA.

En su arquitectura básica los sistemas automatizados de reconocimiento biométrico necesitan un mínimo de módulos que suele operar de forma secuencial⁴⁹: interfaces que incorporan lectores o sensores, módulos de mejora de la captación y de extracción de atributos, espacios de almacenamiento y, por supuesto, módulos de cotejo de información captada en fresco respecto a la almacenada.

Por otra parte, y por cuanto se refiere a sus fases de funcionamiento, los sistemas de reconocimiento biométrico comprende una serie de procesos o fases que de forma

48. Enmiendas aprobadas por el Parlamento Europeo el 14 de junio de 2023, cit. supra.

49. Escajedo San-Epifanio, L., *Tecnologías biométricas*, cit., 2017, 177-178.

genérica pueden agruparse en cuatro⁵⁰: a) fase de enrolamiento, sea de una persona en singular, sea de una categoría como las de «emoción miedo», «franja de edad de mayores de 65», o categorías de sexo; b) una fase de elaboración del patrón (que puede comprender a su vez varias subfases); c) un tercer momento de captación del patrón en fresco, es decir, un momento en el que colocamos a una persona ante el sistema, para que reconozca en ese cuerpo una información captable y susceptible de ser cotejada; y, d), por último, el cotejo del patrón captado en fresco con el patrón almacenado.

2.2. Fuentes de datos, datos en crudo y patrones biométricos

En términos absolutos no existe una Biometría que pueda señalarse la más óptima para el reconocimiento. Aquello que se desea reconocer en un cuerpo puede ser muy diverso, como también han de tenerse en cuenta las posibilidades de los escenarios operativos en los que se aplicará el reconocimiento. Por poner un ejemplo, las huellas dactilares son muy singularizantes, pero en la entrada y salida de una excavación arqueológica o una obra en construcción, o en la entrada y salida de un comedor escolar de infantil, es poco previsible que el promedio de personas tenga las manos lo suficientemente limpias como para poder captarlas bien. Si el grupo de sujetos enrolados no es muy elevado, daría mejor servicio un sistema de reconocimiento basado en la geometría de la mano, aunque en términos netos resulte menos singularizante.

Desde el punto de vista científico-técnico se procede a una importante distinción entre tres conceptos. Como primer concepto, ha de decirse que son *fuentes de datos biométricos* aquellas características anatómico-físicas o dinámicas del cuerpo humano de las que se capta información para el funcionamiento de un sistema de reconocimiento biométrico, con independencia de que éste sea identificante o no. En segundo lugar, son *datos en crudo* los atributos biometrizable que están presentes en esa fuente y pueden captarse como tales. En tercer y último lugar, son patrones o plantillas biométricas (*biometric template* o *empreinte numérique*), las versiones transformadas o tecnificadas de la información captada. Estos patrones son susceptibles de almacenamiento digital y posterior comparación y, en puridad, coinciden con las nociones de datos biométricos que ofrecen el art. 3.34 RIA y el art. 4.14 RGPD, en el sentido de que son el resultante de un tratamiento técnico de características físicas, fisiológicas o conductuales de una persona física.

Cada sistema de reconocimiento, además de basarse en una fuente concreta de información, está diseñado para procesar en un modo concreto los atributos singularizantes presentes en el dato en crudo. Así, por ejemplo, los dermatoglifos o dibujos que los pliegues cutáneos, las crestas y los surcos dérmicos forman sobre las palmas de las manos, las plantas de los pies y los pulpejos de los dedos, se consideran desde hace dos siglos fuentes de datos muy singularizantes⁵¹; estadísticamente se piensa que la posibilidad de que dos dedos —incluso de la misma persona— tengan una huella idéntica es algo que raya lo imposible. Eso sí, para obtener un patrón biométrico, necesitamos captar algún tipo de imagen de la impresión dactilar,

50. *Ibidem*, pp. 178-180.

51. *Ibidem* pp. 109-111.

aplicarle una serie de filtros y, a partir de ahí, cada sistema de reconocimiento procederá a captar información biométrica. Algunos sistemas de reconocimiento de huellas, por ejemplo, proceden a señalar la posición o geolocalización en la impresión de una serie de puntos o minucias que consideran más llamativos. En el procesamiento no se almacenarán las particularidades presentes en esos puntos (alguna bifurcación llamativa, por ejemplo), sino únicamente la posición o posiciones en los que dentro de esa huella está presente alguna particularidad. Esta información relativa a la localización de caracteres singulares no es tan singularizante como un cotejo completo de las huellas pero, estadísticamente, se considera lo suficientemente singularizante como para poder servir de base a un sistema identificación.

A la vista queda, nótese, que un patrón biométrico nunca almacena la totalidad de atributos característicos que pueden estar presentes en un dato en crudo. Puede decirse que ese patrón pretende ser siempre la mejor síntesis posible de lo característico de un dato en crudo. Se trata, con todo de una pretensión y no todos los sistemas de reconocimiento resultan igual de fiables. Si el sistema no es de calidad y son muchas las personas enroladas, crecerá el riesgo de que aparezcan dos patrones digitales prácticamente idénticos para dos dermatoglifos diferentes en la fuente⁵² o de que el sistema no sea capaz de reconocer de forma positiva una coincidencia entre el dato en crudo de un sujeto enrolado y el dato que fue almacenado en su enrolamiento.

2.3. Utilidades biométricas que identifican, utilidades que no

La existencia de un enrolamiento singularizado previo es un elemento que, a la luz de lo que se ha descrito en los apartados anteriores, distingue claramente los sistemas que ofrecen utilidades de identificación única de aquellos que no la ofrecen.

Las utilidades de identificación que, en sentido amplio, ofrecen los sistemas de reconocimiento mediante biometría pueden agruparse en tres grandes conjuntos⁵³: utilidades de verificación (o comprobación de la identidad alegada, uno-a-uno); utilidades de identificación en sentido estricto (o de determinación de la identidad sin alegación previa de ésta, uno-a-muchos); y utilidades de *cribado biométrico* o de búsqueda y localización de determinadas personas singulares en entornos —físicos o virtuales— no delimitados, buscando, de algún modo, un sujeto x , es decir, no enrolado, en una masa infinita.

Desde el punto de vista jurídico, estas utilidades de identificación, en especial las dos primeras, tienden a definirse de forma separada a otras —como la categorización— debido a que implican el tratamiento de datos biométricos que, sin duda, encajan en la noción del artículo 4.14 RGPD. Son datos que *«permiten o confirman la identificación única de dicha persona»*. Para ofrecer tal funcionalidad, este tipo de sistemas necesita que exista, con carácter previo, un enrolamiento biométrico de la persona a re-identificar, en el sentido de que en un momento previo al de la identificación o verificación se procedió a captar y almacenar la información de la persona. Ese almacenamiento puede realizarse en diferentes formas: 1) en algunos casos existe una base de datos de referencia que almacena la información de los sujetos enrolados, ya sea una base exclusiva que opera como parte del sistema

52. *Ibidem* pp. 109-111.

53. Wayman, J./ Jain, A.K./ Maio, D., «An Introduction», 2005, 4-5.

informático concreto, ya sea como una base a la que se accede en línea; 2) en otros casos, como el del pasaporte biométrico, la información biométrica queda recogida en un chip insertado en dicho documento, de forma que, en cualquier lugar del mundo, y siempre que se cuente dispositivo que combine lectura mecánica del pasaporte y captación en fresco de los rasgos de la persona, es posible proceder a verificar si un pasaporte biométrico pertenece —o no— al cuerpo que lo lleva consigo.

La categorización, el perfilado o el reconocimiento de emociones, entre otros, no pretenden una identificación única. Lo que persiguen es reconocer en el cuerpo de una persona —sea en su dimensión estática, sea en la dinámica— atributos biométricos que, previamente, se han asociado a una serie de categorías. Así, por ejemplo, si con el objeto de distinguir emociones se han asociado emociones a determinadas franjas de movimiento de la barbilla o del arqueamiento de las cejas, el sistema tratará de vincular los atributos que capta en fresco en un rostro con una de las categorías disponibles.

Los historiadores destacan la pionera contribución de Alphonse Bertillon (su Bertillonage o *signalément anthropométrique*) como el hito a partir del cual comienza a desarrollarse una rama de la Biometría específicamente orientada a la identificación inequívoca de las personas. El 31 de agosto de 1832 se abolió en Francia la marca de fuego de los condenados y, desde el punto de vista práctico, determinar en qué casos un delincuente merecía una sanción agravada, en tanto reincidente, se hizo más complejo. La única referencia eran los archivos documentales —entradas en papel de más de 5 millones de personas— y no resultaban de utilidad cuando los sujetos ocultaban su identidad o la falseaban, con la intención evitar que se les aplicase la agravante por reincidencia⁵⁴.

Esa y otras circunstancias propiciaron la aparición de un cuerpo especializado dentro de la policía: la llamada policía técnica⁵⁵. Incorporado a la Prefectura de la Policía París en 1879, Alphonse Bertillon asumió la tarea de elaborar una serie de fichas «signaléticas» (en español vendría a ser algo así como *señaléticas*)⁵⁶ de determinados individuos, basadas en los trabajos antropobiológicos de expertos como Lambert Adolphe Quetelet (1796-1874), o Paul Broca (1825-1880) y sus colaboradores en el Laboratorio de Antropología de la *École Pratique des Hautes Études* de París. El «*signalément anthropométrique*» ofrecía novedoso sistema de rastreo de las fichas, que ya no estaban ordenadas alfabéticamente sino en base a descripciones métricas detalladas de los cuerpos (como la envergadura, el tamaño del cráneo o del codo, los pies o las orejas)⁵⁷. De este modo, la toma de medidas a un detenido reincidente hacia potencialmente posible localizar su ficha aún en el caso de que no hubiera facilitado su nombre real.

54. Auger, D., *Biométrie: l'équilibre entre «liberté individuelle» et promesse sécuritaire serait-il impossible?*, 2005, 26-27.

55. Días, C., *La police technique et scientifique*, PUF, Paris, 2000, 12.

56. Madureira, N., «Policía sin ciencia: la investigación criminal en Portugal: 1880-1936», *Política y Sociedad*, 2005, Vol. 42, n.º 3, 45-62. Vid. Bertillon, A., *Signalitic instructions including the theory and practice of anthropometrical identification*, Werner, Chicago, 1896.

57. McCarthy, P., voz «Biometric Technologies», en *Encyclopedia of applied ethics*, 2ª ed., 2102, Elsevier; vid. también Sutrop, M./ Lass-Mikko, K., «From Identity Verification to Behavior Prediction», *Review of Policy Research*, vol. 29, n.º 1, 21 y ss.

El desarrollo de técnicas de procesamiento de señales digitales (DSP — *Digital Signal Processing*) y sus proyecciones en posibles sistemas reconocimiento singularizante a través de la voz⁵⁸ o las huellas dactilares⁵⁹, llevó a reconocer ya a comienzos de los años 60 el importante potencial de estas tecnologías de cara a garantizar elevados niveles de seguridad en los controles de acceso, uso de claves personales o transacciones financieras⁶⁰. Con el desarrollo e implementación de los primeros sistemas se dará ya por comenzada *la primera generación* de sistemas de reconocimiento automatizado mediante Biometría, desarrollados fundamentalmente sobre fuentes de datos biométricos estáticos y aplicables a pequeños grupos de personas. En los años 70 se desarrollaron e implementaron sistemas de reconocimiento basados en la geometría manual⁶¹, y comenzó, además, el testado de los sistemas sobre grupos más numerosos⁶², tratando de mejorar la celeridad y eficiencia de las formas de reconocimiento.

La literatura considera que fue a partir de ese momento cuando comenzó a evidenciarse un creciente interés por los posibles usos gubernamentales de las tecnologías de identificación automatizada. En los años 80 se desarrollaron sistemas de reconocimiento por Biometría de la retina y mediante firma dinámica, a los que siguieron sistemas de reconocimiento facial. Las tecnologías que tienen su base en el patrón del iris, por su parte, fueron propuestas a mediados de los años 80, pero no se hicieron realidad hasta que se desarrolló un algoritmo lo suficientemente fiable como para la captación de la singularidad presente en ese rasgo humano⁶³.

En aquel tiempo no se preveía, precisamente por su nivel de invasividad, una progresiva expansión del uso de las tecnologías de reconocimiento biométrico al conjunto de la sociedad, pero esa situación cambió radicalmente con los atentados del

-
58. Pruzansky, S., «Pattern-matching», *Journal of the Acoustical Society of America*, 1963, 35, 354-358; Li, K. P. / Dammann, J. E. / Chapman, W.D., «Experimental studies in speaker verification», *J. Acoust. Soc. Am.*, 1966, 40, 966-978; Luck, J., «Automatic speaker verification using spectral measurements», *J. Acoust. Soc. Am.*, 1969, 46, 1026-1031; Stevens, K. / Williams, C. / Carbonell, J. / Woods, B., «Speaker authentication and identification: a comparison of spectrographic and auditory presentation of speech material», *J. Acoust. Soc. Am.*, 1968, 44, 596-607; Atal, B., «Automatic recognition of», *Proc. IEEE*, 1976, 64(4), 460-474; Rosenberg, A., «Automatic speaker recognition», *Proc. IEEE*, 1976, 64(4), 475-487.
59. Trauring, M., «Automatic comparison of finger-ridge patterns», *Nature*, 1963, 197, 938-940.
60. Trauring, M., «On the automatic comparison of finger-ridge patterns», *Hughes Laboratory Research Report* 1961, n. 190.
61. El primer sistema totalmente automatizado de reconocimiento biométrico fue el sistema basado en geometría manual que patentó Robert P. Miller en 1971; referencia de Zunkel, R., «Hand geometry based verifications», en A. Jain, et al. (eds.) *Biometrics: Personal Identification in Networked Society*, 1999.
62. Fejfar, A. / Myers, J., «The testing of 3 automatic ID verification techniques for entry control», *2nd Int. Conf. on Crime Countermeasures*, Oxford, 25-29 de julio de 1977.
63. Wildes, R. P., «Iris recognition: an emerging biometric technology», *Proc. IEEE*, 85(9), 1348-1364, 1997; Jain, A. / Bolle, R. / Pankati, S. *Introduction to biometrics*, en Jain, A. / et al. (eds.) *Biometrics: Personal Identification in Networked Society*. Kluwer Academic Press, 1999. Vid. NSTC *Biometrics History*, 2006.

11-S⁶⁴ y la aprobación del pasaporte biométrico europeo⁶⁵. Este último, por exigencias de los EE.UU., incorporó con carácter general un chip de lectura automatizada que almacena las biometrías del rostro y las huellas de los ciudadanos de la UE.

En el contexto de las reflexiones originadas por este pasaporte, numerosas instituciones reflexionaron sobre lo que implicaba esta inclusión. El Comité Consultatif National d'Éthique pour les Sciences de la vie et de la santé (CCNE) francés, por ejemplo, se refirió en 2007 al *riesgo de biometrización* del ser humano⁶⁶, mientras otras instituciones hablaban de hipervigilancia, códigos de barras⁶⁷ para humanos o tatuajes biopolíticos⁶⁸. El Supervisor Europeo para la Protección de Datos, el Grupo del Artículo 29, la Comisión de libertades del Parlamento Europeo y los expertos que durante meses expusieron sus tesis ante la Cámara de los Lores del Reino Unido, por su parte, también advirtieron de los riesgos de permitir recurrir a identificaciones basadas en la corporeidad⁶⁹. Bajo la amenaza de que los ciudadanos de la UE fueran excluidos del sistema de exención de visados para acceder a los EEUU, no obstante, los legisladores europeos no tuvieron más remedio que aceptar la imposición del pasaporte biométrico.

2.4. Importantes limitaciones de los sistemas de reconocimiento biométrico: científico-técnicas, de arquitectura y ético sociales

A mediados de los años 90 un informe del Tesoro de los Estados Unidos, que se hizo público tiempo después⁷⁰, destacaba por qué razones el uso de las tecnologías biométricas se había reservado hasta prácticamente el siglo XXI a parcelas muy exclusivas de la realidad social (como la seguridad de las altas operaciones financieras, o el control de instalaciones militares o instancias de alta seguridad nacional). Uno de los hitos que contribuyó a un cambio de este paradigma, fue Ley patriota, aprobada en los EEUU entre las respuestas a los atentados del 11-S. Y sus consecuencias impulsaron un cambio de paradigma. Como previeron muchos expertos, la inclusión de información biométrica en los pasaportes europeos contribuyó a normalizar el uso de estos sistemas de reconocimiento, y se expandieron de forma acrítica a contextos tan cotidianos como el desbloqueo de un móvil o el acceso a un lugar de ocio.

Ello no debería hacer olvidar, no obstante, que los sistemas de reconocimiento biométrico, en especial los que usan biometrías débiles, siguen teniendo en la actualidad importantes limitaciones en su base científica, en su arquitectura informática y tienen importantes impactos desde la perspectiva ético-social y jurídica.

En relación con su base científica, hemos de tener en cuenta que, por más que el cuerpo humano ofrezca un sinnúmero de posibilidades de captación de biometrías,

64. *Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act*, Pub L., n.º 107-56, 115 Stat. 272, 2001.

65. Lyon, D., *La società sorvegliata*, cit., 2002, 96; Rule, J. B., *Privacy in Peril*, cit., 43-39.

66. CCNE, *Biométrie, dones identifiants et droits de de l'homme*, 2007, cit., 3.

67. Crews jr., C. W., «Human Bar Code. Monitoring Biometric Technologies in a Free Society», *Policy Analysis*, n.º 452, 2002,1 y ss.

68. Agambe, G., «Bio-political tattooing», *Le Monde*, 11 de enero de 2004.

69. Detalladamente, Escajedo San-Epifanio, L., *Tecnologías biométricas*, cit., 2017, 71-75, 110 y ss.

70. Jain, A. K./ Flynn, P./ Ross, A. A.: *Handbook on Biometrics*, Springer, 2008, 1.

los sistemas de reconocimiento biométrico son siempre imperfectos⁷¹. Imperfectos, para empezar, porque no hay ninguna biometría universal, permanente o estable al cien por cien⁷², ni que ofrezca una base indubitada para una categorización consistente de los cuerpos humanos respecto a parámetros como las emociones. En segundo lugar, ha de tenerse en cuenta que la variabilidad biológica entre las personas no se distribuye de forma homogénea y que esa circunstancia hace que todo sistema de reconocimiento sea mucho más eficiente respecto a las personas que están en el promedio de su rango de identificación que respecto a las personas que resultan biométricamente atípicas. Por último, y no menos importante, ha de tenerse en cuenta que, si bien en hipótesis pueden medirse muchos atributos en el cuerpo humano, el actual estado del arte no ofrece métodos mensurativos sólidos aplicables a todas las fuentes de información potencial. El cuerpo humano no es fácilmente biometrizable⁷³.

Por cuanto se refiere a los sistemas informáticos, ha de tenerse en cuenta que la arquitectura de los sistemas de reconocimiento resulta determinante respecto a parámetros como la precisión, el rendimiento y la ciberseguridad⁷⁴. Incluso escogiendo los algoritmos más óptimos (esto es, algoritmos que den una respuesta adecuada a la necesidad de captar la singularidad de la muestra)⁷⁵, acciones como la mejora de las imágenes, la extracción de los atributos, la comparación y la toma de decisiones constituyen elementos críticos en todos los sistemas de reconocimiento biométrico. Por una conjunción entre las limitaciones de base científica y las de la arquitectura del sistema, todos los sistemas de reconocimiento biométrico tienen además un umbral de tolerancia que hace previsible cierto rango de falsos positivos y falsos negativos. Sólo el ojo del experto humano es capaz de reducir significativamente tales umbrales.

A esto ha de añadirse una reflexión sobre los impactos ético-sociales y jurídicos de los sistemas de reconocimiento biométrico. Aunque a primera vista las posibilidades de hipervigilancia intencionada parecen las más problemáticas en perspectiva de derechos fundamentales, no deben pasarse por alto otras consecuencias controvertidas, como es el caso de: el impacto colateral en personas no sospechosas; y la probabilidad de que la información biométrica pudiera revelar información sobre la salud de una persona, su origen étnico o racial o características corporales en las que se expresan determinadas convicciones religiosas (como las tonsuras

71. De Hert, P./ Scheuers, W./ Brouwer, E., «Machine-readable identity documents with biometric data in the EU —part III— Overview of the legal Framework», *Keesing Journal of Documents and Identity*, 2007/ 22,23-26; Kindt, E., «Biometric applications and the data protection legislation (the legal review and the proportionality test)», *Datenschutz und Datensicherheit (DuD)*, 31/ 2007,166-170; Brouwer, E.R. *Digital borders and real rights: effective remedies for third-country nationals in the Schengen Information System*. Brill, 2008, 137.

72. Lanitis, A., «A survey of the effects of ageing on biometrical identity verification», *International Journal of Biometrics*, 2 (1), 2010, 34-52.

73. Magnet, S.A., *When Biometrics Fail. Gender, Race and Technology of Identity*, Duke University Press, 2011, 2.

74. Maltoni/ Maio/ Jain/ Prabhakar, *Handbook Fingerprint*, 2009, 11-22.

75. Pfaffenberger, B., *Que's Computer and Internet Dictionary*, Que, 1995, 15; Preneel, B., «An Introduction to Modern Cryptology», en *Criptology best practices*, KU Leuven, 2018, 19-25.

en el cabello, las cabezas rapadas de los creyentes budistas o las barbas de judíos ortodoxos, entre muchas otras).

A ello se suma el hecho de que muchas fuentes de información biométrica están expuestas y pueden ser recogidas en la vida diaria de una persona con poco esfuerzo (huellas en los objetos que toca, imágenes no consentidas del rostro, etc.). Sin perjuicio de lo delicado de datos biométricos relativos a la identidad sexual y la orientación sexual, ha de advertirse también sobre la vulnerable situación de las personas cuyo cuerpo es atípico para un sistema y de las personas que se ven ante el riesgo de que los sistemas capten de forma colateral datos de salud asociados a sus características biométricas estáticas o dinámicas.

Por el azar de la naturaleza o por acontecimientos posteriores al nacimiento hay personas cuyos cuerpos carecen de los rasgos supuestamente universales que dan soporte al sistema o presentan atributos fuera de rango —manos muy grandes o muy pequeñas, por ejemplo—, así como personas en las que, temporalmente —por ejemplo, por una enfermedad— no es posible captar de forma adecuada la información que el sistema utiliza como base para el reconocimiento. Una retinitis, por ejemplo, es una de las muchas patologías oculares que dificulta la captación fiable del patrón del iris. Ese tipo de circunstancias son, *per sé*, altamente sensibles y estigmatizantes, y su impacto puede verse agudizado si las personas se ven expuestas cotidianamente a sistemas de reconocimiento biométrico ante los que necesitan explicar que, al menos por hoy, su cuerpo no está en condiciones de presentarse como dato en fresco. Otro grupo de personas en alto riesgo de estigmatización son las que en aplicaciones de cribado o categorizantes tienen, siendo totalmente inocentes, patrones biométricos cercanos a lo que se hubiera determinado como perfiles sospechosos.

Por cuanto se refiere a la salud, la literatura distingue entre, de una parte, las implicaciones médicas directas y, de otra, las implicaciones indirectas de las tecnologías de reconocimiento biométrico. Así, son implicaciones directas aquellos impactos que pueden generarse en la salud por el uso de los componentes de un sistema de reconocimiento biométrico. Es caso del riesgo de transmisión de enfermedades mediante los sistemas que requieren un contacto físico, o los riesgos de exposición prolongada a sistemas de captación —por ejemplo, los que usan infrarrojos— en especial cuando la fuente de información biométrica es el iris o la retina.

Las implicaciones indirectas, por su parte, se refieren a la posibilidad de que en el funcionamiento del sistema quede expuesta información médica de la persona. Sin ánimo de exhaustividad, algunos síndromes genéticos tienen expresiones en los dermatoglifos, en las características del rostro o en el color de la piel. La información con más riesgo de quedar expuesta de forma colateral es, en general, la de aquellas condiciones de salud que, temporalmente o de forma definitiva, impiden captar la información biométrica que sirve de base a un sistema. Es el caso, por ejemplo, de dolencias oculares que impiden ver el iris (degeneración macular, retinitis, retinoblastoma, entre otros), dolencias que han dañado las huellas dactilares y enfermedades que han alterado sustancialmente la morfología facial (como la hinchazón por paperas, o un flemón dental).

3. LA VERIFICACIÓN BIOMÉTRICA, ¿FUERA DEL REGLAMENTO?

Para destacar el elevado riesgo de la identificación remota en tiempo real, el Libro Blanco opta por compararlo con el riesgo que percibe en la autenticación biométrica, a juicio de la Comisión de mucho menor riesgo. Ese contraste, recogido en el Libro Blanco, se mantendrá en las diferentes versiones del RIA y ha terminado por imponerse en el texto final. Tal circunstancia, no obstante, no debe hacer olvidar que a efectos del RGPD, tanto los sistemas de identificación en sentido estricto (remotos o no), como los de autenticación, emplean datos biométricos con encaje en el artículo 4.14 y, por tanto, sometidos a la prohibición que recoge el art. 9.1⁷⁶ respecto al tratamiento de categorías especiales de datos. Recientemente el CEPD⁷⁷ y, siguiendo su estela, las APDs insisten de forma rotunda en que los datos biométricos empleados en los sistemas de autenticación (1 a 1) son datos que encajan en la definición del 4.14 del RGPD, y que recaen sobre ellos la especial protección y las limitaciones que se recogen en el artículo 9, porque la autenticación persigue las más de las veces una identificación de la persona, aunque lo sea «uno-a-uno». Salvo concurrencia de las circunstancias y los requisitos de garantía del apartado 9.2. RGPD, el tratamiento de datos biométricos identificantes está prohibido, por más que el RIA pueda paradójicamente estimar que algunos sistemas de identificación afectados por esa prohibición puedan certificarse como aptos para ser ofrecidos en el mercado.

Cabe señalar, por tanto, que lo que aquí se planteará en relación con la verificación no trae causa en las tensiones que en la elaboración del RIA se han planteado respecto a la noción de datos biométricos del RGPD, sino en la necesidad de determinar qué modalidades de reconocimiento biométrico serían acogidas en la regulación del RIA y cuáles no.

Una enmienda del Parlamento Europeo propuso incorporar al RIA la definición de verificación biométrica, así como dos enmiendas en los considerandos 8 y 8bis. Así, y dado que la noción de identificación remota del considerando n.º 8 de la propuesta de Reglamento incluía todo tipo de identificaciones a distancia «*independientemente de la tecnología, los procesos o los tipos de datos*», el Parlamento propone matizar que se excluirán «*los sistemas de verificación que se limitan a comparar los datos biométricos de una persona con sus datos biométricos facilitados previamente (“uno respecto a uno”)*». La razón por la que la verificación es excluida de la identificación a distancia resulta llamativa. De algún modo, aunque no lo explica, el Parlamento parece presumir, que el inciso relativo a «todo tipo de procesos», puede llevar a entender que la identificación comprende tanto comparaciones uno-a-muchos, como comparaciones uno-a-uno, pero que ni unas ni otras resultan preocupantes cuando no se realizan en forma «remota».

En el 8 bis, por su parte, el Parlamento propone incluir una delimitación entre a la identificación a distancia y verificación, explicando, en una frase algo confusa, que

76. Art. 9.1. RGPD. Quedan prohibidos el tratamiento de datos personales que revelen el origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas, o la afiliación sindical, y el tratamiento de datos genéticos, datos biométricos dirigidos a identificar de manera unívoca a una persona física, datos relativos a la salud o datos relativos a la vida sexual o las orientaciones sexuales de una persona física.

77. CEPD, *Guidelines 05/022 on the use of facial recognition technology in the area of Law enforcement*, versión 1.0 de 2022 y versión 2.0 de 2023.

los sistemas de identificación remota se distinguen de los sistemas de verificación individual «de cerca» por la finalidad con la que se utilizan. Así, entiende que los sistemas de verificación únicamente persiguen «confirmar si una persona física concreta que se presenta para su identificación» está, por ejemplo, autorizada para acceder a un servicio, un dispositivo a un local.

La definición de verificación que propone el parlamento resulta, sin embargo, bastante confusa. Nótese que se dice que la persona «se presenta para su identificación», dando a entender que hay una participación activa en su identificación, pero no precisa si como parte de ese presentarse la persona alegará una identidad (que llevaría a hacer una comparación uno-a-uno) o si el sistema deberá comprobar uno-a-muchos si la persona está o no efectivamente enrolada. A efectos de la normativa de protección de datos, la comparación uno-a-uno y la comparación uno-a-muchos tienen un impacto muy diferente, en especial porque el uno-a-uno impide cotejos amplios e, incluso, permite operar sin la necesidad de una base de datos de referencia. Será relevante, por otra parte, que el Parlamento abra el debate sobre los elementos que delimitarán verificación e identificación remota. En esa distinción entre cerca y lejos, la toma de postura inicial apunta —sin precisar— a una cuestión de distancia física, aunque aparecen de forma velada la cuestión de la participación del sujeto a identificar y el hecho de si el identificador conoce o no si tal sujeto había sido previamente enrolado en el sistema.

En coherencia con esos considerandos, el Parlamento propuso modificar el punto 36 del artículo 3 propuesto por la Comisión, incluyendo un inciso entre paréntesis destinado a que la verificación quedase sin lugar a dudas excluida de la identificación biométrica remota⁷⁸. Así, y conforme a la postura del Parlamento Europeo, la noción sistema de identificación remota pasaría a describirse, en el punto 36, como un *«sistema de IA (con excepción de los sistemas de verificación) destinado a identificar a personas físicas a distancia comparando sus datos biométricos con los que figuran en una base de datos de referencia, y sin que el implementador del sistema de IA sepa de antemano si la persona en cuestión se encontrará en dicha base de datos y podrá ser identificada»*.

Tanto la toma de posición del Consejo como el texto finalmente aprobado (art. 3.41 RIA) vuelven sobre la base del concepto inicial de identificación biométrica, sin aceptar el inciso propuesto por el Parlamento. Con el mismo fin, y de forma más clara, lo que se hace es incluir en el listado del artículo 3, punto 36, la noción de verificación biométrica, en estos términos *«la verificación automatizada y uno-a-uno, incluida la autenticación, de la identidad de las personas físicas mediante la comparación de sus datos biométricos con los datos biométricos facilitados previamente»*. Nótese que verificación automatizada y uno-a-uno aparecen en una expresión copulativa, no quedando claro si se trata de sinónimos, como sucede con el inciso *«incluida la autenticación»*. De lo que no hay duda es que, en todos los casos, es imprescindible que existan datos biométricos facilitados en un enrolamiento previo, aunque no se dice que estos deban estar recogidos *«en una base de referencia»*. Este último inciso, en cambio, sí aparecerá en el articulado al hablar de identificación biométrica remota.

Esta noción de verificación cobra especial relevancia toda vez que, en diferentes puntos del articulado, tanto de las prohibiciones del art. 5. como de las categorías de

78. Enmienda n.º 193 del Parlamento Europeo.

alto riesgo del Anexo III, se insiste en que quedan excluidos, según el caso de que se trate, los sistemas de IA destinados a la verificación. Así, por ejemplo, a los efectos de la prohibición del art.5.1. letra h) se dice expresamente en el considerando 17, que, del concepto de sistemas de identificación biométrica remota «*Quedan excluidos los sistemas de IA destinados a la verificación biométrica, que comprende la autenticación, cuyo único propósito es confirmar que una persona física concreta es la persona que dice ser, así como la identidad de una persona física con la finalidad exclusiva de que tenga acceso a un servicio, desbloquee un dispositivo o tenga acceso de seguridad a un local*». Asimismo, el punto 1 del Anexo III, en su redacción final —promovida por el Parlamento en sus enmiendas— indica que los sistemas de verificación no se entenderán incluidos en los sistemas de identificación a los efectos del Anexo III⁷⁹.

Sobre la interpretación de estas exclusiones, de entenderse que verificación incluye identificaciones uno-a-uno e identificaciones uno-a-muchos, en este segundo caso siempre y cuando no se trate de identificaciones remotas, las tecnologías de identificación excluidas del RIA serían numerosísimas, algo que se aborda en el apartado siguiente.

4. LA IDENTIFICACIÓN NO REMOTA, ¿EN EL LIMBO?

El esfuerzo por excluir la verificación, distanciándose de la identificación remota, contrasta con la nula atención que se ha prestado a definir adecuadamente la identificación no remota (el uno-a-muchos) en sentido estricto. El concepto de «identificación biométrica», remota o no, se define en el considerando 15 como «*el reconocimiento automatizado de características humanas de tipo físico, fisiológico o conductual, como la cara, el movimiento ocular, la forma del cuerpo, la voz, la entonación, el modo de andar, la postura, la frecuencia cardíaca, la presión arterial, el olor o las características de las pulsaciones de tecla, a fin de determinar la identidad de una persona comparando sus datos biométricos con los datos biométricos de personas almacenados en una base de datos de referencia, independientemente de que la persona haya dado o no su consentimiento*». La existencia de una base de datos de referencia, es decir, un muchos con el que comparar los unos, parece ser el elemento que diferencia identificación de verificación. Asimismo, y como ya se ha visto en II.3, parece que la ausencia de participación activa es el elemento diferencial entre la identificación biométrica «remota» y la identificación no remota, con independencia de la distancia a la que se realice.

El problema se plantea, como se ha explicado en el apartado anterior, por el hecho de que, al margen de los considerandos, el articulado sitúa la identificación no remota como parte de la verificación y, por lo tanto, bajo el mismo estatuto a los efectos del RIA. Consideran los legisladores que la verificación biométrica «*probablemente tenga una repercusión menor en los derechos de las personas*», aunque no indican de qué depende tal probabilidad. Agrupar identificación biométrica no remota en la categoría de verificación, de forma además poco clara, es una pésima decisión desde la perspectiva de la defensa de los derechos y libertades de los ciudadanos. Tres razones, por otra parte, nos llevan a pensar que esta no ha sido una decisión muy meditada.

79. Enmiendas n.º 710 y siguientes del Parlamento Europeo.

La decisión no parece muy meditada, en primer lugar, porque el RIA, además de no explicar la razón de esa exclusión cuando se aplica a la identificación uno-a-muchos con la participación del sujeto, tampoco delimita con claridad en qué consiste esa *participación activa*. Ello resulta preocupante porque la ausencia de participación activa determinaría que algunos supuestos de identificación no remota uno-a-muchos queden bajo la prohibición del art. 5.1.h o, cuando menos, al abrigo de las garantías previstas para las categorías de alto riesgo. No parece que los legisladores hayan tenido esto en cuenta.

Tampoco parece, en segundo lugar, que los legisladores se hayan detenido en las implicaciones que tiene el uso de datos biométricos singularizantes. Es cierto que el RIA dice recordar, en más de una ocasión, que, conforme al RGPD todos los datos biométricos singularizantes pertenecen a la categoría de datos personales sensibles⁸⁰. Pero no parece actuar en consecuencia. La repercusión de la identificación singularizante en los derechos de las personas no depende sólo de la supuesta participación activa de las personas, sino también de circunstancias como que esa participación pueda ser contraria a su voluntad, de las particularidades del escenario operativo en el que se apliquen, de la calidad y eficiencia de los sistemas o, entre otros, del tipo de decisiones o consecuencias que puedan derivarse del uso de estos. Al excluirse estos sistemas respecto al ámbito del RIA nos encontramos, sin embargo, que la evaluación de su calidad y fiabilidad no disfrutará de las garantías que se aplican a las modalidades clasificadas como de alto riesgo. Tampoco se tiene en cuenta el hecho de que la identificación uno-a-muchos requiere siempre el manejo de bases de datos, cosa no imprescindible en la autenticación, dado que es posible que las personas porten consigo (en el chip del pasaporte, por ejemplo) los datos almacenados que serán empleados para el cotejo con los datos en fresco.

Con todo, la mayor incongruencia se produce respecto al art. 111 y los grandes sistemas informáticos listados en el Anexo X del RIA. Sistemas como SIS, Eurodac y otros, recogidos en el Anexo X, son sistemas que ofrecen, según el caso, utilidades de identificación uno-a-uno o uno-a-muchos, por el momento además con la participación activa de los sujetos. Ninguno de ellos funciona, a día de hoy, como sistema de identificación remota, a menos que quepa entenderse como identificación remota las búsquedas utilizando huellas no captadas del cuerpo (sino presentes en objetos físicos en las que están latentes) o, en su caso, imágenes de la persona que hubieran sido captadas con fines diferentes a un enrolamiento en sentido estricto. En la mayoría de los casos, en especial en el tránsito de fronteras, huellas y rostros de las personas son captadas con su participación activa, en el sentido de que la captación de biometrías de calidad en fresco requiere que la persona se exponga al sistema en formas muy determinadas. Así, por ejemplo, en muchos casos es imprescindible que las personas se presten a hacer rodar sus dedos o pulsar el escáner para la obtención de huellas rodadas o planas, o colocar el rostro en un ángulo determinado para la captación de la geometría del rostro, habiendo retirado de él lentes o mechones de pelo que puedan ocultarlo parcialmente. Sólo en las bases de uso policial se realizan

80. Circunstancia esta que reitera el considerando 54 del texto final, si bien a renglón seguido indica que el hecho de que constituyan una categoría de datos sensibles lleva a clasificar de alto riesgo «varios», que no todos, los casos de «uso crítico» de los sistemas de identificación biométrica remota.

en ocasiones búsquedas desde imágenes o captaciones de huellas tomadas fuera del escenario de enrolamiento, si bien hemos de tener en cuenta que el art. 2.3 excluye de su aplicación los reconocimientos biométricos que los Estados Miembros realizan con fines militares, de defensa o de seguridad nacional, si bien el RGPD es muy estricto en las garantías que les exige. La interpretación del conjunto, por tanto, está lejos de ser clara. Ciertamente es, en cualquier caso, que si una interpretación sistemática excluye de facto la posibilidad de aplicar el RIA a los sistemas biométricos de anexo X, cabría preguntarse qué es lo que ha motivado su inclusión.

5. ¿QUEDA ABARCADO EL CRIBADO BIOMÉTRICO?

Los sistemas de cribado biométrico son sistemas que reconocer «algo», alguna característica —identificante o no— en un conjunto de personas captadas en fresco, sin que haya existido ningún proceso previo de enrolamiento singularizante.

Puede tratarse de sistemas categorizantes, que permiten calcular, por ejemplo, cuántas personas de cada rango de edad y sexo están presentes en un espacio, así como sistemas que pretenden localizar en una multitud bien comportamientos sospechosos, bien posibles similitudes entre las personas presentes y patrones biométricos artificiales, creados a modo robot como cercanos a los de personas buscadas por la comisión de actos delictivos. Cercanos a, nótese, implica que tales patrones artificiales bajo ninguna circunstancia se han confeccionado mediante enrolamiento.

En tanto no funcionan sobre identidades enroladas, es claro que no cabe entenderlos comprendidos en los conceptos de identificación (remota o no) ni de autenticación o verificación. Sí resulta posible entender que los cribados, con independencia de la referencia de la criba, pueden encajar en la idea de categorización biométrica. Desde esa base, y defendiendo una interpretación garantista del RIA, parece posible defender que, según las categorías empleadas, los sistemas de IA de cribado, en la mayor de los casos, encajarán en las prohibiciones del artículo 5 o, en su caso, en las modalidades de alto riesgo del anexo III.

III. SISTEMAS DE RECONOCIMIENTO BIOMÉTRICO AFECTADOS POR LAS PROHIBICIONES Y RESTRICCIONES DEL ARTÍCULO 5 DE REGLAMENTO: EVALUACIÓN SOCIAL, PREDICCIÓN DE PELIGROSIDAD CRIMINAL, AMPLIACIÓN DE BASES DE RECONOCIMIENTO FACIAL, INFERENCIA DE EMOCIONES EN ALGUNOS CONTEXTOS Y CATEGORIZACIONES SOBRE DATOS ESPECIALMENTE PROTEGIDOS EN EL ART. 9 RGPD

1. LA SISTEMÁTICA DEL ARTÍCULO 5 EN LO QUE SE REFIERE AL RECONOCIMIENTO BIOMÉTRICO

Las propuestas de Comisión, Parlamento y Consejo distribuían de forma diferente las modalidades de reconocimiento biométrico que, respectivamente, debían ser objeto de prohibición o, en su caso, quedar recogidas en la categoría de modalidades de alto riesgo. La versión final del texto cede en algunos aspectos a las demandas del

Parlamento, pero resultan también relevantes las concesiones que se han hecho a los intereses de los Estados Miembros.

1.1. La propuesta inicial de la Comisión

El artículo 5 de la propuesta de la Comisión recogía en su apartado 1º la prohibición de cuatro prácticas de IA, relativas al uso de determinadas técnicas subliminales (letra a), la alteración sustancial del comportamiento de personas vulnerables (letra b), la evaluación de la fiabilidad de las personas físicas, conforme a su conducta social o a sus características personales conocidas o predichas, y (letra d) el uso de sistemas de identificación biométrica remota en tiempo real en espacios de acceso público con fines de aplicación de la Ley. Dado que algunos de los supuestos mencionados no guardan relación directa con el reconocimiento biométrico (especialmente las letras a y b) y que han sido abordados por otros autores de este Tratado, no se entrará a analizar el alcance de todas estas propuestas de prohibición, máxime teniendo en cuenta que la propuesta de la Comisión dista mucho del texto finalmente aprobado.

El objetivo de este apartado del comentario es situar el debate durante el período de elaboración del texto, para entrar después en el detalle de la regulación finalmente aprobada.

1.2. Toma de postura del Parlamento

La propuesta de la Comisión fue criticada⁸¹ por desatender relevantes tomas de posturas previas, incluyendo las del Consejo de Europa (2021)⁸², la Agencia de Derechos Fundamentales de la UE (2019)⁸³ y el Parlamento⁸⁴. En esos documentos se había expresado con mucha contundencia que la vigilancia biométrica, el reconocimiento de emociones o la categorización impactan de forma muy negativa sobre una amplia gama de derechos fundamentales, además de los principios del Estado de derecho y los valores democráticos⁸⁵. Conocida la propuesta, además, se añadieron nuevas críticas por el hecho de que el supuesto enfoque de derechos humanos aplicado a la misma no se había traducido en mecanismos efectivos frente a los supuestos más amenazantes⁸⁶.

81. Barkane, I., «Questioning the EU Proposal for an Artificial Intelligence Act: The Need for Prohibitions and a Stricter Approach to Biometric Surveillance», *Information Polity* 27, 2022: 147-162.

82. Council of Europe, *Guidelines on Facial Recognition*, 2021.

83. FRA, *Facial recognition technology: fundamental rights considerations in the context of law enforcement*, 2019.

84. Madiaga, T./ Mildebrath, H., *Regulating facial recognition in the EU*, European Parliament, 2021.

85. Véanse las críticas, entre otros, de Veale, M/ Zuiderveen Borgesius, F., «Demystifying the Draft EU Artificial Intelligence Act», *Computer Law Review International*, 2021, 22: (4), 97-112; Malgieri, G/ Ienca, M., «The EU regulates AI but forgets to protect our mind», *European Law Blog*; EDRI, *New ECI calls Europeans to stand together for a future free from harmful biometric mass surveillance*, 2021; EDPB, EDPS, *Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*, 2021.

86. Smuha, N.A., «Beyond a Human Rights-Based Approach to AI Governance: Promise, Pitfalls, Plea», *Philosophy & Technology*, 2021, 34: 91-104; Mantelero, A./ Esposito, M. S.

En esa línea, y en sintonía con Resoluciones que se han mencionado con anterioridad, el Parlamento Europeo se mostró partidario a prohibir o, en su caso, restringir al máximo el uso de tecnologías de reconocimiento biométrico. *A priori*, el escenario operativo en el que se pretendieran utilizar estas tecnologías o la finalidad perseguida con el reconocimiento no matizaban en exceso el rechazo del Parlamento, mostrando en su postura muy poca intención de aceptar salvedades a las restricciones generales. Así, su propuesta de enmiendas incluye nuevas prohibiciones que, claramente, implican el uso de modalidades de reconocimiento biométrico y propone que la letra d) propuesta por la Comisión sea redactada en unos términos más amplios. En concreto, en su enmienda 220 el Parlamento Europeo propone prohibir⁸⁷ «d) El uso de sistemas de identificación biométrica remota “en tiempo real” en espacios de acceso público» con muy pocas excepciones.

Por otra parte, en su enmienda 224 y las siguientes, el Parlamento propuso también que determinados supuestos descritos por la Comisión como de alto riesgo, fueran reubicados entre las prácticas de IA prohibida. Es el caso, por ejemplo, de las evaluaciones que sobre la base de la personalidad pretendan predecir la peligrosidad criminal o el riesgo de reincidencia, de algunas actividades relacionadas con los sistemas de ampliación de bases de datos de reconocimiento facial, o del reconocimiento de emociones en los lugares de trabajo, los centros educativos y las fronteras.

1.3. Toma de postura del Consejo y texto finalmente aprobado

Algunos Estados Miembros, en especial Alemania, Francia e Italia, eran contrarios a que el RIA pudiera limitar de algún modo el uso de sistemas de identificación biométrica remota con fines de seguridad nacional. Ciertamente el RGPD limitaba mucho la posibilidad de emplear este tipo de sistemas, pero con el respaldo de los parlamentos nacionales, en especial en forma de Ley, era posible acogerse a las salvedades del art. 9.2. RGPD.

Esa visión de los Estados Miembros, no obstante, se alejaba mucho de la postura previa del Parlamento, refrendada en su toma de posición sobre el RIA. Con la intención de evitar que este desacuerdo supusiera un obstáculo insalvable en la tramitación del RIA, el Consejo optó⁸⁸ por una primera toma de postura alineada con la Comisión, dejando para el tiempo de los trilogos su presión por reducir a la mínima expresión aquellas prohibiciones que pudieran afectar a los intereses de defensa y seguridad de los Estados miembros. La amplitud de la noción de verificación en el RIA y el sorprendente nuevo apartado del art. 2 (el art. 2.3 RIA)⁸⁹ supusieron un importante avance en esa dirección, a lo que se sumó una nueva propuesta de redacción de las salvedades aplicables a las prácticas de IA prohibidas del artículo 5.

Este conjunto de prohibiciones, en especial la relativa a algunos usos de los sistemas de identificación biométrica remota, se ha señalado con frecuencia como símbolo del pretendido garantismo del RIA, aunque la realidad dista mucho de

«An evidence-based methodology for human rights impact assessment (HRIA) in the development of AI data-intensive systems» *Computer Law & Security Review*, 2021: 41.

87. Enmiendas aprobadas por el Parlamento Europeo el 14 de junio de 2023, cit. supra.

88. Véase el texto de compromiso de la Cuarta Presidencia, 19 de octubre de 2022, cit. supra.

89. Vid. supra el texto completo, en I.1.

esa afirmación⁹⁰. Unido a lo ambiguo de algunas exclusiones, que son muchas, la calidad técnica de algunas de las normas aprobadas no hace previsible una eficacia instrumental significativa respecto al ordenamiento jurídico que ya está en vigor. Es llamativo, por ello, que la referencia a las prácticas prohibidas siga siendo un tópico insistente a la hora de comunicar a los ciudadanos que los legisladores europeos han reflejado en el RIA un elevado compromiso con la justicia, la seguridad y otros relevantes valores⁹¹.

En ese conjunto de «prácticas a prohibir» —que no exactamente prohibidas— ha destacado especialmente el interés suscitado por el reconocimiento biométrico remoto en tiempo real y en espacios de acceso público, realizado con la finalidad de garantizar la aplicación del Derecho. Las excepciones y necesidades de precisión respecto a la prohibición fueron notables —y divergentes— en las tomas de postura de la Comisión, el Parlamento y el Consejo. También ocupan un lugar considerable en el texto final, repartidas en los considerandos, las definiciones del artículo 3 y buena parte del artículo 5, incluyendo sus apartados 2 a 8. Por tal motivo, se ha visto oportuno desglosar en dos epígrafes diferenciados el tratamiento de los sistemas de reconocimiento biométrico afectados por las prohibiciones del capítulo II. En el presente epígrafe se abordan los sistemas de reconocimiento biométrico abarcados por las letras c) a g) del artículo 5.1. RIA, dejando para el siguiente el tratamiento singular de la identificación biométrica remota regulada en la letra h) el art. 5.1, junto con los apartados 5.2. a 5.8.

2. BIOMETRÍAS AFECTADAS POR LAS PROHIBICIONES DE LAS LETRAS C, D Y E DEL ART. 5.1. REGLAMENTO: EVALUACIONES DE LA PERSONALIDAD CON FINALIDADES DE PUNTUACIÓN CIUDADANA, PREDICCIÓN DE RIESGO DE COMETER DELITOS Y AMPLIACIÓN DE LAS BASES DE RECONOCIMIENTO FACIAL

Las letras c), d) y e) del texto definitivo de RIA recogen supuestos de hecho en los que pueden encajar tecnologías biométricas, aunque con estas últimas no se agote la totalidad de supuestos posible. La letra c) comprendería las tecnologías biométricas que puedan emplearse para evaluar el comportamiento o características de la personalidad inferidas o predichas. La letra d), por su parte, se refiere a sistemas de IA, incluyendo los de base biométrica, que puedan ser utilizados para realizar evaluaciones de riesgos de personas físicas con el fin de valorar o predecir el riesgo de que una persona física cometa un delito, siempre y cuando la forma de elaborar los perfiles de referencia sea una evaluación de rasgos y características de la personalidad. Algunas biometrías blandas son empleadas, no sin polémica, con este tipo de objetivos y merecen consideraciones desde la perspectiva de los principios del Estado democrático de Derecho y de la política criminal que son atendidos con más detalle en otro capítulo de este

90. Cotino Hueso, L., «Sistemas de inteligencia artificial con reconocimiento facial y datos biométricos, Mejor regular bien que prohibir mal», *El Cronista del Estado Social y Democrático de Derecho*, n.º 100, 68-79, 73.

91. García-Villegas M., «The Symbolic Uses of Law: At the Heart of a Political Sociology of Law», en *The Powers of Law: A Comparative Analysis of Sociopolitical Legal Studies*. Cambridge University Press, 2018, 19-37.

Tratado⁹². El riesgo de hacer resurgir *biologizaciones de la criminalidad*, como las míticas las tesis del *uomo delinquentis* de Lombroso y sus discípulos⁹³, disimuladas u ocultas bajo el halo de la aparente infalibilidad las tecnologías digitales, es una circunstancia que no debe ser perdida de vista.

Se establece una salvedad para el caso en el que la evaluación no sea meramente predictiva, sino que se realice —con el apoyo, si es caso de la IA— a partir de la implicación de una persona en una actividad delictiva y con «base en hechos objetivos y verificables directamente relacionados con una actividad delictiva».

En el caso de la letra e), por su parte, es claro que la creación o ampliación de bases de datos de reconocimiento facial requiere —sea mediante enrolamiento de personas, sea mediante captaciones no consentidas de imágenes faciales— el uso de tecnologías biométricas. La prohibición, no obstante, se extiende a dos supuestos muy diferentes, dado que en un caso se refiere a imágenes que están supuestamente difundidas en abierto, con un régimen jurídico muy diferente según su titular haya consentido o no esa difusión, y en el segundo de los casos se habla de circuitos cerrados, apuntando a grabaciones privadas (que no de vigilancia). Nótese además que en ambos casos lo que se prohíbe es la extracción «no selectiva» de imágenes, concepto que no se precisa pero que deja claramente fuera de la prohibición la extracción selectiva.

3. RECONOCIMIENTO BIOMÉTRICO DE EMOCIONES EN LOS LUGARES DE TRABAJO Y EN LOS CENTROS EDUCATIVOS, EXCEPTUANDO LOS CASOS EN QUE SE PERSIGAN MOTIVOS MÉDICOS O DE SEGURIDAD (ART. 5.1. F). CONCEPTO DE RECONOCIMIENTO DE EMOCIONES EN EL REGLAMENTO

Durante siglos, las fuerzas del orden han utilizado rostros no sólo para identificar sino también para intentar leer estados mentales e inferir comportamientos sospechosos⁹⁴. A pesar de la impresión que puede dar la reciente atención que está recibiendo el reconocimiento facial en el espacio político, la literatura académica e incluso la prensa, se trata de un área de desarrollo tecnológico de larga data —al menos desde los años 70⁹⁵,⁹⁶—. Es cierto, no obstante, que los avances en aprendizaje automático y la mejora de las técnicas de visión por computación, combinados con las biometrías, han hecho aumentar significativamente las capacidades de vigilancia de

92. Véase, en esta obra, el trabajo de F. Miró Llinares sobre los sistemas policiales predictivos y los sistemas de reconocimiento de emociones.

93. Wechsler, H., «Biometric Security and Privacy Using Smart Identity», *Review of Policy Research*, vol. 29, 1/ 2012, 78-79; Strasser, P., «Biometrie - ein Schritt in die Überwachungsdemokratie?», Schaar, P. (ed), *Biometrie Und Datenschutz - Die vermessene Mensch*, 2007, 14-15.

94. Miranda, D., «Identifying Suspicious Bodies? Historically Tracing the Trajectory of Criminal Identification Technologies in Portugal», *Surveillance & Society* 2020, 18 (1): 30-47.

95. Gray, M. «Urban Surveillance and Panopticism: Will We Recognize the Facial Recognition Society?», *Surveillance & Society*, 2003, 1(3), 314-30; Introna, L., Wood, D. «Picturing Algorithmic Surveillance: The Politics of Facial Recognition Systems» *Surveillance & Society* 2004 (2), 177-98.

96. Urquhart, L./ Miranda, D., «Policing faces: the present and future of intelligent facial surveillance», *Information & Communications Technology Law*, 2021, 31(2), 194-219.

las fuerzas policiales⁹⁷, con el riesgo de perpetuar formas inadecuadas de elaboración de perfiles y de refuerzo de las categorías de sospecha sobre determinados grupos de sujetos⁹⁸.

El SEPD, antes de la aprobación final del RIA, solicitó una regulación más estricta de las tecnologías de reconocimiento facial y del uso de sistemas biométricos de reconocimiento en un sentido amplio⁹⁹, incluyendo «*la marcha, las huellas dactilares, el ADN, la voz, pulsaciones de teclas y otras señales biométricas o de comportamiento*» cuando fueran empleadas por las fuerzas del orden, ya fuera en espacios públicos ya en otro tipo de espacios¹⁰⁰. La atención al uso que de este tipo de sistemas hacen las empresas ha pasado más desapercibido, aunque se emplea de forma creciente en los contextos publicitarios y comerciales¹⁰¹. Se acostumbra a agregar datos de comportamiento a los de preferencias o compras¹⁰², así como datos que pueden ayudar a la prevención de delitos¹⁰³. Esos usos se están extendiendo también a los espacios de trabajo, por el momento de forma bastante alegal. A diferencia de la videovigilancia, el monitoreo digital se traduce en una vigilancia a menudo inapreciable pero omnipresente¹⁰⁴ y, en realidad, sabemos muy poco del modo en que se utilizan los *big data* en este tipo de vigilancias¹⁰⁵.

El reconocimiento de emociones es un campo de investigación interdisciplinar que abarca, entre otras, conocimientos de disciplinas de la psicología, las ciencias cognitivas y la informática.¹⁰⁶ Persigue que las computadoras puedan captar las

-
97. Kotsoglou, K. / Oswald, M. «The Long Arm of the Algorithm? Automated Facial Recognition as Evidence and Trigger for Police Intervention» (2020) 2 *Forensic Science International: Synergy* 86-89; Venema, R. «How to Govern Visibility?: Legitimizations and Contestations of Visual Data Practices after the 2017 G20 Summit in Hamburg», *Surveillance & Society* 2020, 18 (4) 522-39; Purshouse, J. / Campbell, L., «Privacy, Crime Control and Police Use of Automated Facial Recognition Technology», *Criminal Law Review* 2019 (3), 188-204.
98. Garvie, C. / Bedoya, A. / Frankle, J., *The Perpetual Line-up: Unregulated Police Face Recognition in America*, Georgetown Law, Center on Privacy & Technology, 2016; Williams, D., «Fitting the Description: Historical and Sociotechnical Elements of Facial Recognition and Anti-Black Surveillance», *Journal of Responsible Innovation*, 2020, 7 (1), 74-83.
99. European Data Protection Board and European Data Protection Supervisor, (EDPB-EDPS), *Joint Opinion 5/2021 on the Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*, 2021.
100. European Parliament, *Motion for a European Parliament Resolution on Artificial Intelligence in Criminal Law and its Use by the Police and Judicial Authorities in Criminal Matters*, 2021.
101. McStay, A. *Emotional AI*, Sage, 2018; Stark, L., / Huey, J. «The Ethics of Emotion in AI Systems», *FAccT 21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 2021, 782-93.
102. Lyon, D., *Surveillance after Snowden*, John Wiley & Sons, 2015, 76-77.
103. Brayne, S., «Big data surveillance: The case of policing». *American Sociological Review* 82(5), 2017, 977.
104. Van Oort, M., «The Emotional Labor of Surveillance: Digital Control in Fast Fashion Retail», *Critical Sociology*, 2019, 45(7-8), 1167-1179.
105. Brayne, S., «Big data surveillance», cit., 2017, 977.
106. De Gregorio, G., «The Rise of Digital Constitutionalism in the European Union», *International Journal of Constitutional Law* 19.1, 41-70, 2021.

emociones e intenciones humanas y gestionarla con diferentes utilidades¹⁰⁷. Algunos sistemas automatizados de reconocimiento facial, por ejemplo, están adiestrados para detectar seis emociones básicas (ira, alegría, miedo, sorpresa, disgusto/asco y tristeza¹⁰⁸) ya sea en forma estática (arrugas en la frente, posición de la barbilla o de las cejas)¹⁰⁹, ya sea mediante la captación de información biométrica de diferentes micro-expresiones en un lapso de tiempo (velocidad de parpadeo, de los movimientos de la barbilla o las cejas, fijación de la mirada, etc.).

Según la propuesta de la Comisión, un sistema de reconocimiento de emociones es «*un sistema de IA cuyo objetivo es identificar o inferir emociones o intenciones de personas físicas a partir de sus datos biométricos*». Este tipo de sistemas puede utilizar datos biométricos singularizantes con encaje en el art. 4.14 RGPD, pero lo común es que se recurra a datos no singularizantes, como se ha explicado ya en II.1. En este sentido, la interpretación del n.º 39 del art. 3 RIA debe hacerse en consonancia con la noción de datos biométricos del RIA y no, por tanto, con la del RGPD. Así, se define como «sistema de reconocimiento de emociones» un «*sistema de IA destinado a distinguir o inferir las emociones o las intenciones de las personas físicas a partir de sus datos biométricos (art.3.39 RIA)*», entendiendo por estos últimos los datos biométricos descritos en el número 34 del mismo artículo.

La prohibición, nótese, sólo alcanza la introducción en el mercado, la puesta en servicio para este fin específico o el uso de *sistemas de IA para inferir las emociones de una persona física en los lugares de trabajo y en los centros educativos*, excepto cuando el sistema de IA esté destinado a ser instalado o introducido en el mercado por motivos médicos o de seguridad. De la propuesta del Parlamento, se ha excluido en este texto final la referencia al uso de estos sistemas en el control de fronteras (afectado, según el caso, por los artículos 111 y siguientes o la exclusión del art. 2.3 RIA). Otros supuestos, no abarcados por la prohibición ni excluidos en los artículos que acaban de citarse, quedan en principio en el conjunto de modalidades de IA de alto riesgo del Anexo III.

Conviene llamar la atención, no obstante, sobre el hecho de que, aún reconocida por los legisladores su preocupación por el importante margen de imprecisión que tiene el reconocimiento de emociones (e intenciones), no se ha previsto ninguna moratoria respecto a la posibilidad de introducir en el mercado este tipo de sistemas. Así, el considerando n.º 44 se refiere a una «gran preocupación» sobre la base científica de estos sistemas, señalando como deficiencias principales *la fiabilidad limitada, la falta de especificidad y la limitada posibilidad de generalizar*. En contextos como el laboral o el educativo, alcanzados por la prohibición, esa baja fiabilidad agudiza situaciones de desequilibrio de poder características de estos espacios, pero en general, los legisladores deberían haber tenido en cuenta que, por su poca o nula potencialidad de identificar individuos únicos, la captación de estos datos biométricos está excluida del art. 4.14 RGPD (y las garantías del art. 9) y necesitaba obtener del RIA garantías complementarias.

107. Picard, R. W., «Affective Computing: Challenges», *International Journal of Human-computer Studies* 59.1, 2003, 55-64.

108. Lewinski, P./ Trzaskowski, J./ Luzak, J., «Face and Emotion Recognition on Commercial Property under EU Data Protection Law», *Psychology & Marketing* 33 (9), 2016, 729-46.

109. Eckman, P., *Emotions Revealed: Understanding Faces and Feelings*, HB, 2003, p.17-19.

4. CATEGORIZACIÓN BIOMÉTRICA CON EL FIN DEDUCIR O INFERIR SU RAZA, OPINIONES POLÍTICAS, AFILIACIÓN SINDICAL, CONVICCIONES RELIGIOSAS O FILOSÓFICAS, VIDA SEXUAL U ORIENTACIÓN SEXUAL (ART.5.1 G)

La letra g) del artículo 5.1. prohíbe la introducción en el mercado, la puesta en servicio para este fin específico o el uso de *sistemas de categorización biométrica que clasifiquen individualmente a las personas físicas sobre la base de sus datos biométricos para deducir o inferir su raza, opiniones políticas, afiliación sindical, convicciones religiosas o filosóficas, vida sexual u orientación sexual*. Esta prohibición, indica en su inciso final, no incluye el etiquetado o filtrado de conjuntos de datos biométricos adquiridos lícitamente, como imágenes, ni la categorización de datos biométricos en el ámbito de la garantía del cumplimiento del Derecho. Respecto a la primera de las exclusiones, la del etiquetado de datos adquiridos lícitamente, se generan respecto a su interpretación una serie de dudas. La categorización biométrica no se basa en principios en datos singularizantes, comprendidos en la noción del art. 4.14 RGPD, por lo que es de esperar que la exclusión a la prohibición se mantenga dentro de las modalidades de categorización. La referencia al etiquetado, sin embargo, menciona datos adquiridos previamente, de forma lícita, e incluye las imágenes, por lo que no está del todo claro descartar la aplicabilidad de las garantías del RGPD.

No es claro, la verdad, en que supuesto se ha podido obtener lícitamente una imagen que permita etiquetar a una persona para inferir su opinión política, su afiliación sindical o sus convicciones religiosas. ¿Quizá una foto que, con motivo de alguna celebración, ha publicado una asociación política, o imágenes obtenidas en las celebraciones de manifestaciones públicas de diferente tipo? Sinceramente, es dudosa la base legal para el almacenamiento y el etiquetado de esas informaciones, como también lo es la posibilidad de contar con una base legal, y de justificar en un Estado democrático la necesidad y utilidad de tratar datos tan sensibles. A ello se añade, claro está, e, escaso margen de que la categorización biométrica —que es de lo que trata el artículo— permita realmente captar algunas de esas informaciones.

Así, y conforme al actual estado del arte, es llamativa la referencia a la posibilidad de inferir mediante categorías biométricas las opiniones políticas, la afiliación sindical o las convicciones filosóficas o religiosas a partir de los datos biométricos de las personas, y algo dudosa la fiabilidad de los sistemas que se han desarrollado en lo que se refiere a la orientación sexual y la vida sexual. El color de piel o algunos atributos claramente presentes en el cuerpo, como el tipo de pelo o algunos rasgos faciales llamativos tampoco permiten categorizar con la calidad que algunas voces parecen presuponer, porque el mestizaje en las sociedades contemporáneas ha crecido considerablemente. Llamativamente, en cambio, del listado de datos de categorías especiales del art. 9.1. RGPD se ha excluido de la prohibición a las categorizaciones biométricas que pretendan deducir o inferir las condiciones de salud de las personas. A día de hoy, en buena medida porque son utilizadas como referencia en el ámbito sanitario, son las categorías respecto a las que más esfuerzo científico-tecnológico se está realizando.

Como nota final, conviene advertir que el texto finalmente aprobado, recogido en el art. 3.35 RIA, define como «sistema de categorización biométrica» «*un sistema de IA destinado a incluir a las personas físicas en categorías específicas en función de sus*

datos biométricos», si bien ha de criticarse que se haya excluido de tal definición a los sistemas de este tipo que accesorios a un servicio comercial, y estrictamente necesarios por razones técnicas objetivas. Pudiera ser, por ejemplo, un sistema que infiera la edad de las personas a los efectos de controlar la exclusión de venta de tabaco o alcohol a menores de edad, pero ello no explica por qué se excluyen de la definición de sistema de categorización biométrica. Deberían haberse excluido, según corresponda, del tenor literal de la prohibición o del estatuto que corresponde, en su caso, a las categorías de alto riesgo.

IV. SISTEMAS DE RECONOCIMIENTO BIOMÉTRICO AFECTADOS POR LAS PROHIBICIONES Y RESTRICCIONES DEL ARTÍCULO 5 REGLAMENTO (Y II): LA IDENTIFICACIÓN BIOMÉTRICA REMOTA «EN TIEMPO REAL» EN ESPACIOS DE ACCESO PÚBLICO CON FINES DE GARANTÍA DEL CUMPLIMIENTO DEL DERECHO

1. PREOCUPACIÓN GLOBAL POR LA IDENTIFICACIÓN BIOMÉTRICA REMOTA EN ESPACIOS DE ACCESO PÚBLICO

El uso de la tecnología de reconocimiento facial se está extendiendo en todo el planeta¹¹⁰. África, América Latina o Asia (en especial China)¹¹¹ son citados con insistencia, pero lo cierto es que también los Estados Miembros de la UE recurren a reconocimientos biométricos a distancia, aunque sea de forma puntual en el abordaje de emergencias o en escenarios sensibles de acontecimientos masivos¹¹². A nivel global están aumentando las reivindicaciones de organizaciones civiles que demandan una regulación de las condiciones en las que la identificación biométrica impulsada por IA, en especial la remota, puede desplegarse¹¹³. En lugares con un Estado de Derecho inconsistente (o inexistente) el impacto potencial de usos policiales de este tipo en los derechos de las personas es mayor¹¹⁴, pero eso hace que en ocasiones pasen desapercibidos los usos que realizan las fuerzas corporativas y empresariales, explotando las lagunas del Derecho¹¹⁵.

110. Commission Nationale de l'Informatique et des Libertés (CNIL), *Reconnaissance Faciale — Pour Un Debat À la Hauteur des Enjeux*, 2019, 3; Urquhart, L./ Miranda, D., «Policing faces: the present and future of intelligent facial surveillance», *Information & Communications Technology Law*, 31(2), 2021, 194-219.

111. Dauvergne, P. «Facial recognition technology for policing and surveillance in the Global South: a call for bans», *Third World Quarterly*, 43(9), 2022, 2325-2335.

112. European Data Protection Board, *Guidelines 05/2022 on the Use of Facial Recognition Technology in the Area of Law Enforcement*, 2022, 7 y ss.

113. Ala-Pietilä, P./ Smuha, Nathalie A., «A Framework for Global Cooperation on Artificial Intelligence and its Governance», en *Reflections on Artificial Intelligence for Humanity*, B. Braunschweig/ M. Ghallab (eds.), Springer, 2021, 253.254.

114. Zalnieriute, M., «Facial recognition surveillance and public space: protecting protest movements», *International Review of Law, Computers & Technology*, 2024, 1-20; O'Flaherty, M., «Opinions, Facial Recognition Technology and Fundamental Rights», *European Data Protection Law Review*, 2020, 6 (2), 170 y ss.

115. Dushi, D., «The Use of Facial Recognition Technology in EU Law Enforcement: Fundamental Rights Implications», *Global Campus South East Europe*, 2020, 4; Raposo, V.

Este tipo de identificaciones usan, en un porcentaje elevado, biometrías faciales¹¹⁶ y, con ellas, pretenden identificar —o al menos a localizar— personas sospechosas. No obstante, durante la pandemia de la COVID-19 se utilizaron también con fines de salud pública¹¹⁷, destacando sus usos en países como Rusia o China¹¹⁸. La situación jurídica de su uso en Europa, incluso tras la aprobación del RIA resulta, cuando menos, ambigua¹¹⁹, a la vista de enorme mosaico de leyes primeras y secundarias de la UE o los Estados Miembros que regulan algunos aspectos de este tipo de reconocimientos, además de las resoluciones y directrices de diferentes instituciones.

El Garante Italiano se pronunció en 2021 sobre el sistema móvil SARI Real-time, diseñado para ser instalado específicamente en un lugar y analizar en tiempo real los rostros —capacidad máxima 10.000— filmados en una zona geográfica delimitada y dotada de una serie de cámaras interconectadas¹²⁰. Si, a través de un algoritmo de reconocimiento facial, SARI encuentra una coincidencia entre un rostro presente en la lista de vigilancia y un rostro filmado por una de las cámaras, el sistema es capaz de generar una alerta que atrae la atención de los operadores, aunque entre tanto es capaz de grabar flujos de video —como los sistemas tradicionales de videovigilancia—.

En su resolución, el Garante no sólo muestra su preocupación por las personas incluidas en listas de vigilancia, sino por la vigilancia colateral que este tipo de sistemas genera respecto a personas presentes en los espacios públicos o participantes en manifestaciones políticas o sociales que, *a priori*, no hayan sido listadas por las fuerzas policiales como personas objeto de atención. Considera el Garante que la base jurídica para aplicar un sistema de estas características es inexistente, tras analizar tanto el RGPD como preceptos del Código Procesal Penal italiano, entre otros.

En España, por citar otro ejemplo, un sistema piloto instalado en Mercadona —sin consulta previa a la AEPD ni evaluación de impacto en materia de protección de datos, preceptiva conforme al art. 35.1. RGPD— recibió una sanción por vulneración de licitud del tratamiento, agravada por el hecho de tratarse de categorías especiales de datos¹²¹.

L., «(Do Not) Remember My Face: Uses of Facial Recognition Technology in Light of the General Data Protection Regulation», *Information & Communications Technology Law* 45, 2022, 32 (1).

116. Negri, P./ Hupont, I./ Gomez, E., «A Framework for Assessing Proportionate Intervention with Face Recognition Systems in Real-Life Scenarios», *Computers and Society*, 2024 (2), 12.
117. Raposo, V. L., / Du, L., «Facial recognition technology: is it ready to be used in public health surveillance?», *International Data Privacy Law*, 14 (1), 2024, 66-86; Raposo, V. L., «Can China's "Standard of Care" for COVID-19 Be Replicated in Europe?», *Journal of Medical Ethics* 46, 2020, 451.
118. Article 19, «Emotional Entanglement: China's Emotion Recognition Market and its Implications for Human Rights», 2021.
119. Raposo, V. L., «Look at the camera and say cheese: the existing European legal framework for facial recognition technology in criminal investigations», *Information & Communications Technology Law*, 33(1), 2023, 1-20.
120. Garante per la Protezione dei Dati Personali, *Parere sul Sistema Sari Real Time*, doc. 9575877, n.º 127 de 25 de marzo de 2021.
121. Un análisis detallado de esa resolución en Simón Castellano, P./ Dorado Ferrer, X., «Límites y garantías constitucionales frente a la identificación biométrica», *Revista de Internet, Derecho y Política — IDP*, n.º 35, marzo de 2022, 1-13.

La prohibición que se recoge en el artículo 5 RIA y las modalidades de identificación remota clasificadas como de alto riesgo no vienen a paliar esta situación de falta de claridad, dado que, como se ha dicho, el artículo 2 en su apartado tercero excluye del ámbito de aplicación del RIA tanto los sistemas de uso militar como aquellos utilizados en el contexto de la seguridad nacional, además de aquellos que tengan como única finalidad la investigación y el desarrollo científico.

2. INTERPRETACIÓN DE LOS CONCEPTOS CLAVE DE LA PROHIBICIÓN DEL ART. 5.1.H)

2.1. Identificación biométrica remota en tiempo real y diferido

Como ya se ha dicho, la noción de identificación biométrica remota no se discutió en exceso. Se quiso adoptar un concepto funcional —*vid.* considerando n.º 8— por lo que el tipo de tecnología y los procesos o tipos de datos concretos que se empleen al efecto no son protagonistas de la definición. Lo relevante es que no sea necesaria la participación activa de las personas (art. 3.34). Queda en duda, no obstante, la referencia que se hace en el considerando 8º al hecho de que el cotejo con la base de datos de referencia se realice «*sin saber de antemano si la persona en cuestión se encontrará en dicha base de datos y podrá ser identificada*», porque en el caso de que el identificador «sepa» de antemano —no se sabe en qué términos— que la persona estará en dicha base, un inciso de esas características llevaría a considerar que la identificación no es remota. En tanto el texto articulado no recoge esa previsión y conocido el valor jurídico que corresponde a los considerandos —en comparación con el texto articulado—, es conveniente que la identificación remota sea delimitada ciñéndose en exclusiva al art. 3.34 RIA.

Mucho más controvertida fue la cuestión de distinguir entre sistemas que identifican en tiempo real y sistemas que lo hacen en diferido. Los considerandos indican que ese matiz se traduce en diferencias tanto en lo que se refiere a las características como a las formas de uso y los riesgos que entrañan, razón por la cual se hace el esfuerzo de proceder a delimitar cada uno de esos conceptos. En el texto final se indica que son sistemas de identificación en tiempo real aquellos en los que las tres fases de este tipo de modelos, esto es, la recogida de los datos biométricos, la comparación y la identificación, «se producen de manera instantánea, casi instantánea o, en cualquier caso, sin una demora significativa». En caso de duda, en aplicación conjunta de los puntos 42 y 43, que respectivamente recogen los sistemas de identificación remota en tiempo real y en diferido, por defecto cualquier sistema de identificación biométrica remota que no pueda ser considerado sistema que actúa en tiempo real, será considerado a los efectos del RIA sistema en diferido y, por tanto, sujeto a un régimen mucho más flexible.

Respecto a la videovigilancia tradicional, estaríamos, por tanto, ante sistemas que, en tiempo real, están captando bajo su rango de vigilancia y, al tiempo, cotejando patrones biométricos captados en fresco con patrones almacenados en una base información. En cambio, en los sistemas «en diferido» los datos en fresco serían datos ya recabados, y el cotejo habría de producirse con una demora significativa, al menos hasta el punto de no poder considerar que se está actuando «casi en directo». Esa delimitación, la verdad, no resulta muy garantista. Hubiera sido mejor, por poner un

ejemplo, requerir que la captación en fresco y el cotejo estuvieran, de algún modo, balcanizados, sin posibilidad de que el sistema empiece autónomamente, y sin una mínima intervención humana, el proceso cotejo.

2.2. Escenario operativo y misión: espacios de acceso público y fines de garantía del cumplimiento del Derecho

La prohibición de la letra h) del artículo 5.1. se aplica únicamente en los casos en que el sistema de identificación, además de ser remoto y operar en tiempo real, esté implantado en un espacio de acceso público. El texto final del RIA describe como «espacio de acceso público» «cualquier lugar físico, de propiedad privada o pública, al que pueda acceder un número indeterminado de personas físicas» (punto 44, art. 3) y no considera relevante que deban cumplirse, o no, determinadas condiciones de acceso —como tener una entrada en el caso de un museo o un teatro—, o que el espacio tenga posibles restricciones de capacidad o de seguridad, incluyendo las limitaciones por razón de edad. No hay duda, eso sí, de que el espacio debe ser «físico», por lo que los espacios en línea, independientemente de cómo se pueda acceder a ellos, no son objeto de atención a los efectos de este precepto.

La propuesta de considerandos de Comisión, Parlamento y Consejo ha ido aumentando de forma considerable la lista de espacios que pueden reunir estas características¹²², recogiendo además algunas observaciones para clarificar supuestos que queden en duda. Se reconoce, no obstante, que se deberá determinar caso por caso si un espacio es o no de acceso público, atendiendo a las particularidades de cada situación en concreto. A modo orientativo, se explica que hecho de que físicamente se pueda acceder (porque una verja está abierta) no implica, por sí solo, que el espacio sea de acceso público, pudiendo existir indicios —como una señal— de que el acceso está restringido. Tampoco son de acceso público los locales de empresas y fábricas o lugares a los que sólo se pretenda que accedan empleados o proveedores de servicios, o las zonas de acceso público de las prisiones. Se indica, asimismo, que algunos espacios pueden contener, de una parte, zonas de acceso público —como vestíbulos—, junto con otros espacios que no sean de acceso público.

La expresión «con fines de garantía del cumplimiento del Derecho», por su parte, fue incorporada en el proceso final de elaboración, para sustituir la referencia previa a la aplicación de la Ley. Con carácter previo, Comisión, Parlamento y Consejo habían mantenido en sus tomas de postura la referencia a los «fines de aplicación de la Ley», pero probablemente ese término dejó de tener sentido cuando el art. 2.3 RIA excluyó del ámbito de aplicación los usos de sistemas de IA por parte de los Estados miembros, con fines militares, de defensa y de seguridad nacional.

La expresión garantía de cumplimiento del Derecho es desarrollada en el considerando n.º 24, además en dos definiciones del artículo 3. Una de ellas

122. Sin ánimo de exhaustividad se listan, por ejemplo: tiendas, restaurantes, cafeterías; de prestación de servicios, por ejemplo, bancos, actividades profesionales, hostelería; deportivas, por ejemplo, piscinas, gimnasios, estadios; de transporte, por ejemplo, estaciones de autobús, metro y ferrocarril, aeropuertos, medios de transporte; de entretenimiento, por ejemplo, cines, teatros, museos, salas de conciertos, salas de conferencias; de ocio o de otro tipo, por ejemplo, vías y plazas públicas, parques, bosques, parques infantiles. Véase el considerando 19 del texto aprobado.

es la recogida en el art. 3.46 RIA, según la cual se entienden como garantía del cumplimiento del Derecho las actividades realizadas por las autoridades garantes del cumplimiento del derecho, o en su nombre, *«para la prevención, la investigación, la detección o el enjuiciamiento de delitos o la ejecución de sanciones penales, incluidas la protección frente a amenazas para la seguridad pública y la prevención de dichas amenazas»*. La segunda definición está recogida en el art. 3.45 RIA, que define como autoridades garantes del cumplimiento del Derecho aquellas autoridades públicas competentes para las actividades que acaban de listarse como parte del art. 3.46 RIA, incluyendo cualquier organismo o entidad a quien el Derecho de un Estado miembro hubiere confiado el ejercicio de la autoridad y las competencias necesarias para ejecutar ese tipo de actividades.

2.3. Excepcionable en la medida que el uso sea estrictamente necesario para alcanzar uno o varios de los siguientes objetivos

La inclusión de algunas modalidades de identificación biométrica remota entre las prohibiciones del artículo 5 parece obedecer más a motivos políticos que de técnica legislativa, dado que las excepciones a esta prohibición de tal relevancia desde la propuesta de la Comisión, que la forma de articulación del texto, acosado por tantas excepciones, resultaba muy poco clara. Es el número 21 de los considerandos el que, de forma más honesta, expresa cuál es la previsión real —que no prohibición— respecto a estos sistemas: *«Todo uso de un sistema de identificación biométrica remota “en tiempo real” en espacios de acceso público con fines de aplicación de la ley debe estar autorizado de manera expresa y específica por una autoridad judicial o por una autoridad administrativa independiente de un Estado miembro»*. Dicho de otro modo, serán los usos no autorizados aquellos a los que prestará atención el RIA, si bien cabe cuestionar si un reglamento orientado a evaluar los riesgos de cara a permitir la introducción en el mercado de sistemas de IA es, o no, el lugar idóneo para referirse a estos supuestos. Por sentido común, en un Estado de derecho las vigilancias de este tipo que no tengan base legal ni autorización adecuada son contrarias a los sistemas constitucionales de los Estados miembros, resultando de todo punto innecesario dedicarles tanto espacio en el RIA.

El considerando continúa afirmando que *«en principio, dicha autorización debe obtenerse con anterioridad al uso, excepto en situaciones de urgencia debidamente justificadas, es decir, aquellas en las que la necesidad de utilizar los sistemas en cuestión sea tan imperiosa que imposibilite, de manera efectiva y objetiva, obtener una autorización antes de iniciar el uso. En tales situaciones de urgencia, el uso debe limitarse al mínimo imprescindible y cumplir las salvaguardias y las condiciones oportunas, conforme a lo estipulado en el Derecho interno y según corresponda en cada caso concreto de uso urgente por parte de las fuerzas o cuerpos de seguridad»*. Ello aporta poca novedad respecto a lo ya dicho de las provisiones constitucionales. Nótese que no estamos hablando de la comercialización de los sistemas sino de su uso, por lo que adentrarse a precisar en un Reglamento como este —cual si fueran excepciones a una prohibición que no es tal— en qué circunstancias pueden o no los Estados miembros otorgar conforme a su Derecho interno dichas autorizaciones con el objetivo de aplicación de la Ley se preveía, como así ha sido, un campo abonado a discusiones interminables.

Por si fuera poco, recuerda el considerando 22, que *«en el marco exhaustivo que establece este Reglamento, que dicho uso en el territorio de un Estado miembro conforme a lo dispuesto en el presente Reglamento solo debe ser posible cuando el Estado miembro en cuestión haya decidido contemplar expresamente la posibilidad de autorizarlo en las normas detalladas de su Derecho interno, y en la medida en que lo haya contemplado. En consecuencia, con el presente Reglamento los Estados miembros siguen siendo libres de no ofrecer esta posibilidad en absoluto o de ofrecerla únicamente en relación con algunos de los objetivos que pueden justificar un uso autorizado conforme al presente Reglamento»*.

Todas estas explicaciones adicionales han sido necesarias porque el texto finalmente aprobado se aleja de la propuesta de enmienda del Parlamento Europeo de prohibir de forma amplia los sistemas de identificación biométrica, al menos en su aplicación por parte de los poderes públicos. Ese texto hubiera sido más sencillo de redactar, si bien cuestionable también en términos de necesidad y oportunidad de su inclusión en el RIA. El parlamento, en su enmienda 330, propuso prohibir todo uso de sistemas de identificación biométrica remota —entonces en la letra d del artículo 5.1.—, eliminando —y haciendo innecesaria la interpretación de— las tres situaciones que en la actualidad recoge la letra h) del art. 5.1.

La enmienda del Parlamento no prosperó y el texto actual admite excepciones a la prohibición de la letra h) cuando los usos de sistemas biométricos remotos en tiempo real que persigan, en el espacio público, la persecución de determinados objetivos:

- a) la búsqueda de posibles víctimas de un delito, incluidos menores desaparecidos;
- b) la atención a determinadas amenazas para la vida o la seguridad física de las personas físicas, o amenazas de atentado terrorista;
- y c) la detección, la localización, la identificación o el enjuiciamiento de los autores o sospechosos de los delitos listados en el Anexo II ¹²³, coincidentes con los de la a Decisión Marco 2002/584/JAI del Consejo, siempre y cuando la pena prevista para tales delitos en el Estado miembro que haya solicitado la Euroorden supere un umbral mínimo de tres años de prisión.

3. LOS APARTADOS 2 A 8

Los apartados 2 a 8 del artículo 5º se refieren, en exclusiva, a completar las excepciones a la prohibición contenida en la letra h) del apartado 1 del artículo 5. Como opción política se vio preferible incluir simbólicamente la prohibición del reconocimiento biométrico remoto en el listado del art. 5.1, aunque después fuera necesario un esfuerzo ingente para precisar el mínimo —o nulo— margen de aplicación real de esa prohibición.

123. Anexo II: *Lista de los delitos a los que se refiere el artículo 5, apartado 1, párrafo primero, letra h, inciso iii):* terrorismo; trata de seres humanos; explotación sexual de menores y pornografía infantil; tráfico ilícito de estupefacientes o sustancias psicotrópicas; tráfico ilícito de armas, municiones y explosivos; homicidio voluntario, agresión con lesiones graves; tráfico ilícito de órganos o tejidos humanos; tráfico ilícito de materiales nucleares o radiactivos; secuestro, detención ilegal o toma de rehenes; delitos que son competencia de la Corte Penal Internacional; secuestro de aeronaves o buques; violación; delitos contra el medio ambiente; robo organizado o a mano armada; sabotaje; participación en una organización delictiva implicada en uno o varios de los delitos enumerados en esta lista.

Antes de entrar en el detalle conviene hacer una observación. Seguridad nacional y defensa son los ámbitos más propicios para utilizar las biometrías de identificación remota, pero el art. 2.3 excluye potencialmente las nacionales y el art. 111 RIA abre la vía para eximir las que se refieran al de control de fronteras. Hemos dicho además, que algunos sistemas que se emplean en seguridad nacional y control de fronteras son sistemas de identificación no remota (uno-a-muchos) y, atendiendo al modo en que se define el término verificación, parece que se ha previsto para ellos el mismo estatuto jurídico que para la identificación uno-a-uno.

Habida cuenta de ello, ¿era realmente necesaria una redacción tan desordenada de la letra h), que se inicia de forma algo rotunda para quedar progresivamente vaciada de contenido en sus subapartados, y después, en los apartados 2 a 8 del artículo 5? A ello ha de añadirse, además, que algunas previsiones no son sino reiteraciones de preceptos ya previstos en otras normas. En este sentido, recuerda el apartado 8º (art. 5.8 RIA), que *«el presente artículo no afectará a las prohibiciones aplicables cuando una práctica de IA infrinja otras disposiciones de Derecho de la Unión»*.

El apartado 2 del artículo 5 señala la necesidad de tener en cuenta una serie de aspectos a la hora de desplegar un sistema de identificación biométrica remota para los fines que, excepcionalmente, la letra h) del art. 5.1. ha dicho que permite hacerlo. Los aspectos son, en primer lugar, a) la naturaleza de la situación que dé lugar al posible uso, y en particular la gravedad, probabilidad y magnitud del perjuicio que se produciría de no utilizarse el sistema; y, en segundo lugar, b) las consecuencias que tendría el uso del sistema en los derechos y las libertades de las personas implicadas, y en particular la gravedad, probabilidad y magnitud de dichas consecuencias. Sin olvidar que, como se ha dicho, importantes sistemas están excluidos de la aplicación de esta previsión, otra duda que se plantea es la de qué autoridad y en qué términos supervisará estos aspectos.

Por cuanto se refiere a los apartados tercero y quinto, tampoco parece ser el artículo 5 el lugar oportuno para recordar que *«todo uso de un sistema de identificación biométrica remota “en tiempo real” en espacios de acceso público con fines de garantía del cumplimiento del Derecho estará supeditado a la concesión de una autorización previa por parte de una autoridad judicial o una autoridad administrativa independiente cuya decisión sea vinculante del Estado miembro en el que vaya a utilizarse dicho sistema, que se expedirá previa solicitud motivada y de conformidad con las normas detalladas del Derecho nacional mencionadas en el apartado 5»* (art. 5.3) o que, dado que serán medidas restrictivas de derechos fundamentales, los Estados deberán detallar las condiciones en las que se puede solicitar y obtener este tipo de autorizaciones, y ponerlas en conocimiento de la comisión a más tardar 30 días después de su adopción (art. 5.5.), además de realizar respecto a ellas tareas de seguimiento de las que se reportará anualmente a la Comisión.

A día de hoy, se ha dicho ya con insistencia, los sistemas constitucionales de los Estados Miembros exigen explícitamente ese tipo de autorizaciones —además de una base jurídica sólida para las mismas—, como también se prevén las situaciones en que por especial urgencia puedan flexibilizarse las autorizaciones previas. A ello se añade que los datos que emplean estos sistemas son datos biométricos comprendidos en el art. 4.14 RGPD (y afectados por el art. 9 RGPD). Reiterarlo en este artículo 5 no resta vigencia a esas otras previsiones normativas, pero, la verdad, tampoco aporta.

Realmente, la única novedad a este respecto es la que se recoge en el art. 5.4, que exige que «todo uso de un sistema de identificación biométrica remota “en tiempo real” en espacios de acceso público con fines de garantía del cumplimiento del Derecho se notificará a la autoridad de vigilancia del mercado pertinente y a la autoridad nacional de protección de datos de conformidad con las normas nacionales a que se refiere el apartado 5». Esa obligación de información es interesante, aunque nuevamente deberá recordarse que los sistemas empleados con fines militares, de defensa o de seguridad nacional empleados por los Estados Miembros, según reza el art. 2.3. RIA, no serán parte de ese todo. Dado que el art. 9.2 RGPD deja pocas posibilidades para que operadores privados utilicen este tipo de tecnologías y, más importante, en tanto el artículo 5 RIA no persigue la prohibición de esos usos por operadores privados, estos apartados del 2 al 8 del artículo 5 RIA rozan, la verdad, la irrelevancia.

V. Modalidades de reconocimiento biométrico clasificadas como de alto riesgo

La regulación del tratamiento que merecen los sistemas de IA de alto riesgo es una cuestión a la que el RIA dedica gran parte de su articulado, tal y como se aborda en detalle en otro capítulo del presente tratado, a cargo del profesor Cotino Hueso. Además de los sistemas enumerados en el Anexo III del Reglamento, listado sobre el que se trabajará en este epígrafe, hemos de tener en cuenta que se habilita a la Comisión a actualizar esta lista mediante actos delegados.

Que un sistema de reconocimiento biométrico quede abarcado en este conjunto de modalidades tiene una gran trascendencia, vistos los requisitos que el RIA establece respecto a la gestión de sus riesgos, los mecanismos destinados a evitar sesgos negativos, la obligación de contar con registros automáticos de actividad de los sistemas, la supervisión humana o, entre otros, el adecuado nivel de precisión, robustez o ciberseguridad. Ciertamente es que el título IX abre la posibilidad de que la Comisión y los Estados Miembros apoyen la redacción de *Códigos de Conducta voluntarios* para su aplicación en sistemas que no sean de alto riesgo, pero no hay margen de comparación entre las medidas recogidas como obligatorias en un Reglamento y las medidas de autocontrol. Se esperaba del RIA que fuese una pieza clave del nuevo constitucionalismo digital de la UE¹²⁴, pero ha quedado muy lejos de esa previsión.

Como ya se ha dicho, atendiendo al tenor literal del artículo 6 y, en especial, del Anexo III del Reglamento, encontramos que algunos sistemas de reconocimiento biométrico son clasificados como *sistemas de alto riesgo* dos subgrupos.

En primer lugar, y con independencia del escenario operativo en el que se empleen —punto 1 del Anexo III— son de alto riesgo: a) los sistemas de identificación biométrica remota; b) los sistemas de categorización biométrica basados en atributos o características sensibles; y, c) determinados sistemas de IA destinados al reconocimiento biométrico de emociones. Se excluyen, en cualquier caso, los sistemas de reconocimiento que ofrezcan funcionalidades de autenticación o verificación.

Un segundo subgrupo de modalidades de alto riesgo se detalla en el anexo III, puntos 2 a 8. En concreto se ofrece un listado de veintiuna modalidades de IA

124. De Gregorio, G., «The Rise of Digital Constitutionalism in the European Union», *International Journal of Constitutional Law* 19.1, 41-70, 2021.

consideradas de alto riesgo, agrupadas en seis escenarios operativos: ámbito de la educación y formación profesional; ámbito laboral; servicios y prestaciones esenciales; garantía de cumplimiento del Derecho; tránsito transfronterizo; y administración de justicia. Los enunciados de estas modalidades de IA de alto riesgo se han formulado empleando con insistencia la expresión «*sistemas de IA destinados a ser utilizados para*» acciones tales como evaluar (riesgos, resultados, niveles de aprendizaje, fiabilidad), realizar seguimientos, detectar comportamientos prohibidos, clasificar o tomar decisiones, y cabe la posibilidad de que en alguno de esos casos se trate de: sistemas de reconocimiento biométrico, que no puedan entenderse comprendidos en las tres categorías que describe el punto primero, ni en la excluida categoría de autenticación.

La sistemática empleada para organizar la presentación de las modalidades de alto riesgo no ha sido muy adecuada. Tecnologías biométricas asumen roles en ámbitos operativos y cabe emplear la expresión *aplicación biométrica* para referirse a la conjunción de funcionalidades utilizations y roles que desempeñan¹²⁵. Los objetivos operativos de las tecnologías biométricas son tan diversos como las razones por las cuáles deseamos identificar a las personas o tratamos de encontrar a algunas de ellas¹²⁶. Algunos sistemas buscan a personas cuya Biometría conocemos (enroladas) y otros a personas cuya Biometría —incluso cuya identidad— desconocemos. Algunos sistemas funcionan ante identidades alegadas, y otros verifican la presencia de una identidad no alegada, o puede que en sentido estricto no les importe tanto la identidad de un individuo como su pertenencia a un grupo. Desde esa circunstancia, cabe cuestionar la razón que ha llevado a los legisladores a que los listados de los apartados 2 a 8 de este Anexo III se hayan definido tan encorsetados en un elenco de escenarios operativos —algunos de los cuáles no han sido definidos a los efectos del RIA—.

Respecto a este segundo conjunto, y siempre que encajen en la descripción que se hace en el anexo de esos escenarios operativos y utilidades, puede pensarse en: supuestos de identificación biométrica no remota; categorizaciones biométricas no basadas en atributos sensibles; o detecciones de biometrías comportamentales que no encajen, propiamente, en la categoría de reconocimiento de emociones.

Un ejemplo, respecto al escenario operativo educativo (punto 3, letra d)¹²⁷, puede ser el de un sistema biométrico que, en el escenario educativo, identifica comportamientos prohibidos durante los exámenes mediante un sistema de reconocimiento biométrico adiestrado para detectar comportamientos atípicos (por ejemplo, tendencia a ocultar las manos bajo los pupitres, simulación de estar escribiendo, llamativas desviaciones en la fijación de la mirada, etc.). En el espacio laboral, por su parte, cabe pensar en sistemas como los empleados en las evaluaciones de comportamientos que, conforme al apartado b) del punto 4 pueden utilizarse para asignar tareas a las personas, promocionarlas en su carrera laboral o, incluso, provocar la rescisión de sus contratos. Ahora bien, no siempre se tratará de sistemas

125. Escajedo San Epifanio, L., *Tecnologías Biométricas*, cit., 2017,244-248.

126. Wayman, J. / Jain, A. K. / Maltoni, D. / Maio, D., cit., 2005,4-5.

127. Anexo III. 3. d) Sistemas de IA destinados a ser utilizados para el seguimiento y la detección de comportamientos prohibidos por parte de los estudiantes durante los exámenes en el contexto de los centros educativos y de formación profesional o dentro de estos a todos los niveles.

remotos o categorizantes, planteándose algunas dudas sobre el impacto que en la interpretación puede tener la exclusión de sistemas singularizantes no remotos.

Asimismo, en relación con el bloque 5 del Anexo III, puede suceder que sistemas de identificación no remota puedan ser empleados respecto a situaciones como las descritas respecto a las llamadas de emergencia (Anexo III.5. letra d) o categorizaciones biométricas y detecciones comportamentales diferentes a las descritas en el punto 1 del anexo puedan entenderse abarcadas en enunciados como los de los sistemas destinados a evaluar los riesgos de comisión delictiva o de reincidencia (punto 6, letra d), la fiabilidad de un testigo o perito (como parte del punto 6, letra c, que se refiere a la evaluación de la fiabilidad de las pruebas) y, sin duda, en la gestión del tránsito fronterizo cuando no se disponga de documentos de viaje que puedan ser autenticados mediante verificación biométrica (punto 7, letra d).

Entre las entidades de supervisión previstas para estos sistemas, a la autoridad nacional notificante y la autoridad de supervisión del mercado, el RIA añade otras autoridades de supervisión, como aquellas que se encargan de supervisar las actividades de seguridad, migración o asilo, o bien las agencias de protección de datos. No ha de olvidarse, en cualquier caso, que estas últimas tienen, a su vez, competencias asignadas por el RCPD.

VI. LOS GRANDES SISTEMAS DE RECONOCIMIENTO OPERATIVOS ANTES DE LA ENTRADA EN VIGOR DEL REGLAMENTO (ART. 111 Y ANEXO X)

En coherencia con la expansión de biometrías de identificación a los documentos de viaje de los nacionales —cosa que los EE.UU. aún no han llegado a hacer—, los Estados miembros comenzaron a desarrollar sistemas de información con componentes biométricos en el contexto del Acuerdo de Schengen o el Tratado de Prüm, el sistema de visados¹²⁸ o el Convenio de Dublín, en el seno del cual surgió Eurodac. Al servicio o como parte de tales sistemas de información existen en la actualidad —con un mayor o menor nivel operativo, según el caso— una serie de herramientas de reconocimiento biométrico que se utilizan, entre otros, frente a la inmigración ilegal, el terrorismo o el tráfico de personas. Esas aplicaciones, se ha dicho también, aplican fundamentalmente identificaciones uno-a-uno e identificaciones uno-a-muchos con la participación de las personas, con muy pocas excepciones (limitadas a la persecución criminal o, en algún caso, a la identificación de posibles víctimas. Esta circunstancia, aplicando la previsión del RIA respecto a las verificaciones —*vid.* II.3— deja fuera de su órbita buena parte de estas modalidades de reconocimiento biométrico o, a lo sumo, a la espera de códigos de conducta o directrices voluntarias.

Pese a las resistencias que hubo en otro tiempo, en los últimos años se han adoptado una serie de decisiones que, sometidas a muchos requisitos en materia de protección de datos, avanzan hacia la interoperabilidad de estos grandes sistemas informáticos que son estratégicos en el espacio europeo de libertad, seguridad y justicia. A lo largo de la elaboración del RIA, en especial en lo que se refiere a la inmigración no vinculada a ningún tipo de *conducta delictiva*, algunas voces se

128. Reglamento CE n.º 767/ 2008.

mostraron partidarias de desmontar este tipo de sistemas de reconocimiento o, cuando menos, excluirlos de ese macroproyecto de interoperabilidad. Por muchas razones, sin embargo, el proceso legislativo no podía detenerse a revisar la compleja realidad de estos sistemas.

Se llegó así a la conclusión que recoge el artículo 111 RIA en conexión con su anexo X. Conforme a estas normas, los sistemas informáticos a gran escala ya introducidos o de introducción prevista antes de los 36 meses posteriores a la entrada en vigor de la Ley de IA¹²⁹ están en una especie de ficción jurídica que hace que se los considere «previos» al RIA y, por tanto, exentos de su aplicación hasta el año 2030, salvo en el caso de que, a futuro se proceda a una reforma sustancial de alguno de sus elementos. Este último inciso, ha de señalarse, es lo suficientemente ambiguo como para no saber, en realidad, si el RIA terminará aplicándose a esos grandes sistemas informáticos o no.

Con todo, no es esa la única duda sobre la aplicabilidad real del RIA a estos sistemas. Las búsquedas en diferido, además de las identificaciones uno-a-uno y uno-a-muchos son objeto de notables exclusiones en la aplicabilidad del RIA, sin que haya resultado relevante que manejen datos biométricos ajustados a la noción del 4.14 del RGPD y protegidos, en coherencia, como categorías especiales de datos personales (9.1 RGPD).

1. EL ART. 111 Y EL ANEXO X DEL REGLAMENTO

El artículo 111.1 establece que, sin perjuicio de que se aplique el artículo 5 con arreglo a lo dispuesto en el artículo 113¹³⁰, apartado 3, letra a), los sistemas de IA que sean componentes de los sistemas informáticos de gran magnitud establecidos en virtud de los actos legislativos enumerados en el anexo X, y que se hayan introducido en el mercado o se hayan puesto en servicio antes de que transcurran treinta y seis meses desde la fecha de entrada en vigor del RIA —fecha aún pendiente de determinarse—, deberán estar en conformidad con el presente Reglamento a más tardar el 31 de diciembre de 2030. La posibilidad de aplicación del artículo 5, sin embargo, puede encontrar importantes dificultades atendiendo a una interpretación sistemática que comprenda el artículo 2.3. relativo al ámbito de aplicación y la extensa lista de excepciones que, como hemos visto, permiten aplicarse, por ejemplo, a supuestos como los recogidos en el art.5.1. h). Los plazos son algo más amplios que los que se permiten para otros sistemas de IA que, sin perjuicio de las prohibiciones del artículo 5, estén ya operativos cuando el RIA entre en vigor.

Volviendo a los grandes sistemas informáticos, a partir del 1 de enero de 2031, por tanto, se prevé una evaluación de conformidad de estos sistemas informáticos, sea de conformidad con lo dispuesto en los actos jurídicos que actualmente los

129. Véase infra.

130. El artículo 113, Entrada en vigor y aplicación, establece que el Reglamento entrará en vigor a los 20 días de su publicación en el Diario Oficial de la UE, si bien establece una entrada en vigor escalonada para varios apartados del articulado. Así, la entrada en vigor de los Capítulos II (prohibiciones) y III (IA de alto riesgo) se prevé a los 6 meses de la entrada en vigor, si bien en los Considerandos, como se ha visto, se reconoce que en tanto no se adopten los actos derivados tal entrada en vigor resultará limitada.

regulan, sea de conformidad con los actos jurídicos que a futuro los modifiquen o sustituyen.

A los efectos de la aplicación del artículo 111 RIA, su anexo X recoge, agrupados en 7 bloques, un listado de actos legislativos, entre los que se incluyen los Reglamentos de Interoperabilidad y las Bases de Datos vinculadas por dichos reglamentos, que se abordan a continuación, en los apartados VI.2 y VI.3.

2. LOS REGLAMENTOS DE INTEROPERABILIDAD

En mayo de 2019 se aprobaba en la UE una importante iniciativa de interoperabilidad —en dos Reglamentos¹³¹, uno relativo a fronteras y visados, y otro sobre cooperación policial y judicial, asilo e inmigración—.

Los sistemas de información policial y de control de migraciones disponibles en la UE fueron creados en diferentes momentos y para dar respuesta a diferentes iniciativas, lo cual ha dado lugar a una arquitectura fragmentada —expresión esta de la Comisión— en la que la información no sólo se almacena por separado sino en formas desconectadas que facilitan la generación de puntos ciegos¹³².

La idea de trabajar para que esos sistemas puedan interoperar, no es nueva. Ya en 2004 el Consejo Europeo invitaba a la Comisión a realizar propuestas sobre la interoperatividad del sistema Eurodac (y VIS), junto con otras bases de datos, *a fin de poder poner dicha información al servicio de la prevención y lucha contra el terrorismo*¹³³. Una estrategia de estas características, sin embargo, resultaba compleja de promover dadas las divergencias políticas y las diferencias de capacidad técnica de los diferentes Estados Miembros. A diferencia de los enfoques que empleaban países como Israel, Canadá o los EE.UU., en especial después de los atentados del 11-S¹³⁴, en la UE el SEPD y el GT-WP 29 fueron contundentes al expresar su preocupación por los almacenamientos masivos tanto en casos como los de los visados¹³⁵, como en los de los pasaportes o los salvoconductos y afirmaron que no estaba justificada la creación de una base de datos centralizada que contuviera los datos personales y, en particular, los

131. Reglamento (UE) 2019/817 del Parlamento Europeo y del Consejo, de 20 de mayo de 2019, relativo al establecimiento de un marco para la interoperabilidad de los sistemas de información de la UE en el ámbito de las fronteras y los visados (DO L 135 de 22.5.2019, p. 27), y Reglamento (UE) 2019/818 del Parlamento Europeo y del Consejo, de 20 de mayo de 2019, relativo al establecimiento de un marco para la interoperabilidad entre los sistemas de información de la UE en el ámbito de la cooperación policial y judicial, el asilo y la migración (DO L 135 de 22.5.2019, p. 85).

132. Leese, M. «Fixing State Vision: Interoperability, Biometrics, and Identity Management in the EU», *Geopolitics*, 27(1), 2020, 113-133.

133. de Hert, P./ Gutwirth, S.: «Interoperability of Police Databases within the EU?», *International Review of Law Computers & Technology*, vol 20, 1-2/ 2006, 21-25; J. A. Lewis, J. A.: «Biometrics and Security», *Center for Strategic & International Studies*.

134. Escajedo San-Epifanio, L., *Tecnologías biométricas*, cit., 2017, 258-261.

135. Opinión 3/ 2005 del GT-WP 29, relativa al sistema SIS II; Opinión 7/ 2004 sobre la inclusión de elementos biométricos en los permisos de residencia y visados teniendo en cuenta el establecimiento del Sistema Europeo de Información Sobre visados (WP 96), adoptado el 11 de agosto de 2004.

datos biométricos de todas las personas autorizadas a recibir un pasaporte¹³⁶, un visado o un salvoconducto. Entendían que ello vulneraba el principio de proporcionalidad¹³⁷.

Habida cuenta de que en 2022 la UE establecía en Tallín (Estonia) una inmensa base de datos, gestionada por la Agencia eu-LISA¹³⁸, destinada a recoger información biométrica dactilar y facial de más de cuatrocientos millones de personas de terceros países, es claro que la perspectiva de los Estados Miembros y de las instituciones europeas ha sufrido una drástica modificación.

El camino hacia la interoperabilidad comenzó a clarificarse a partir de las recomendaciones del Grupo de Expertos de Alto Nivel sobre Sistemas de Información e Interoperabilidad, creado en 2016 por la Dirección General de Migración y Asuntos de Interior¹³⁹. El hito definitivo, con todo, llegó con la aprobación de los dos Reglamentos de Interoperabilidad. La UE, definitivamente, acogía un nuevo paradigma en el tratamiento del reconocimiento biométrico con fines policiales y de control de la migración¹⁴⁰. El criterio de limitación de la finalidad de los datos pasa a ser interpretado en un modo más flexible que en otros ámbitos, cediendo el importante rol de garantía que había desempeñado desde finales de los años 90¹⁴¹.

La estrategia de interoperabilidad persigue mejorar la capacidad de los sistemas de información para intercambiar datos, pero no implica que todos los datos se pongan en común. De forma selectiva, y con base en los diferentes niveles de acceso de los usuarios (como policías, funcionarios de migración y guardias de fronteras) se persigue proporcionar un acceso más rápido, fluido y sistemático a la información que necesitan para hacer su trabajo, garantizando al mismo tiempo el respeto a los derechos fundamentales.

Desde el punto de vista técnico, el componente clave de esta estrategia es la creación de un portal único, que permita realizar una única búsqueda en el conjunto de los sistemas que interoperan y recibir conjuntamente todos los resultados disponibles. Por cuanto se refiere al reconocimiento biométrico, la interoperabilidad incluye un servicio de cotejo de datos biométricos obtenidos de las huellas dactilares y de las imágenes faciales. Este servicio de cotejo o comparación biométrica es conocido por las siglas sBMS (por sus siglas en inglés, *EU shared Biometric Matching System*) y una vez activo será uno de los

136. Dictamen de 23 de marzo de 2005 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo sobre el Sistema de Información de Visados (VIS); Dictamen de 19 de octubre de 2005 sobre tres propuestas relativas al Sistema de Información de Schengen de segunda generación (SIS II), COMO 2005 230 final, COM 2005 236 final, y COM 2005 237 final, DOUE C 91.

137. C 313/ 38, inciso 3º del apartado 12.

138. Agencia Europea para la Gestión Operativa de Sistemas Informáticos de Gran Magnitud en el Espacio de Libertad, Seguridad y Justicia.

139. Directorate-General for Migration and Home Affairs, *High-level expert group on information systems and interoperability: Final report*, 2017.

140. Oliveira Martins, B., Lidén, K./ Jumbert, M. G. «Border security and the digitalisation of sovereignty: insights from EU borderwork». *European Security*, 31(3), 2022, 475-494.

141. Hartmut A. «Interoperability Between EU Policing and Migration Databases: Risks for Privacy», *European Public Law*, 2020, 26 (1), 93-108.

sistemas de reconocimiento biométrico más grandes del mundo —superado sólo por el Aadhaar de India¹⁴²—. La base integrará, se ha dicho ya, datos biométricos dactilares y faciales de más de 400 millones de nacionales de terceros países¹⁴³, por el momento, eso sí, en ningún caso datos de nacionales de los Estados Miembros. Entre los procesos disponibles estará la extracción de plantillas biométricas de diferentes bases de datos de la UE, con el objetivo de simplificar la búsqueda y la comparación cruzada de datos biométricos. Un amplio y complejo sistema de agencias (por ejemplo, Interpol, Europol), y numerosas bases de datos (como EES, ECRIS-TCN, VIS, EURODAC y ETIAS, que se describen *infra*) serán la referencia en esta infraestructura interoperable de vigilancia de la migración y el control de la delincuencia en la UE¹⁴⁴.

Existirá un repositorio común de identidades, al que se incorporará la información biográfica de ciudadanos extracomunitarios ya disponible en las bases abarcadas en la estrategia, pero, ha de señalarse que, en principio, los Reglamentos de interoperabilidad no son una base jurídica que permita añadir más información a la ya disponible. La base legítima que permite almacenar, añadir, modificar o suprimir datos en cada una de las bases seguirá siendo el acto legislativo que regula cada una de ellas.

Se impulsa asimismo un mecanismo de detección de identidades múltiples y un conjunto de mecanismos de control de la calidad de los datos, y se prevén una serie de partidas presupuestarias, tanto para eu-LISA, Europol, CEPOL y la Agencia Europea de la Guardia de Fronteras y Costas, como para que los Estados Miembros se doten de componentes técnicos y formación que permitan a sus agentes participar en la comunidad de usuarios, aunque no todos los Estados Miembros participan en las mismas condiciones en los acuerdos de Schengen¹⁴⁵.

142. *Aadhaar* es considerado a día de hoy el mayor sistema de identificación biométrica del mundo, creado por el gobierno con el objetivo de incorporar los datos de la totalidad de personas residentes en India, con independencia de su ciudadanía, se estima que en la actualidad ha enrolado a más de mil doscientos millones de personas. Vid. Escajedo San-Epifanio, L., *Tecnologías Biométricas*, cit., 2017, 275-276; Kloppenburg, S./ Van der Ploeg, I., «Securing Identities: Biometric Technologies and the Enactment of Human Bodily Differences», *Science as Culture*, 29(1), 2018, 57-76.

143. Jones, C., «Data protection, immigration enforcement and fundamental rights: what the EU's regulations on interoperability mean for people with irregular status», *Statawatch and Platform for International Cooperation on Undocumented Migrants*, 2019, 6.

144. Oliveira Martins, B., Liden, K./ Jumbert, M. G. «Border security and the digitalisation of sovereignty: insights from EU borderwork», *European Security*, 31(3), 2022, 475-494.

145. El Acuerdo de Schengen es un acuerdo por el que varios países de Europa suprimieron los controles en las fronteras interiores (entre esos países) y trasladaron esos controles a las fronteras exteriores (con terceros países). En la actualidad forman parte del espacio Schengen los siguientes países: Alemania, Austria, Bélgica, Bulgaria, Croacia, Dinamarca, Eslovenia, España, Estonia, Finlandia, Francia, Grecia, Hungría, Islandia, Italia, Letonia, Liechtenstein, Lituania, Luxemburgo, Malta, Noruega, Países Bajos, Polonia, Portugal, República Checa, República Eslovaca, Rumanía, Suecia y Suiza.

3. SISTEMAS DE RECONOCIMIENTO BIOMÉTRICO COMPRENDIDOS EN EL ANEXO X: ALGUNOS DATOS RELEVANTES

En el momento de aprobación de los Reglamentos de Interoperabilidad existían tres de los seis sistemas que se esperaba abarcar en la estrategia: el SIS, Eurodac y el VIS.

El Sistema de Información de Schengen (SIS)¹⁴⁶ contiene un amplio espectro de descripciones de personas (denegaciones de entrada o estancia, órdenes de detención de la UE, personas desaparecidas, asistencia en procedimientos judiciales, controles discretos) y objetos (incluidos documentos de identidad o de viaje perdidos, robados o invalidados).

El sistema Eurodac¹⁴⁷, por su parte, contiene los datos dactiloscópicos de los solicitantes de asilo y de los nacionales de terceros países que hayan cruzado irregularmente las fronteras exteriores o que se encuentren en situación irregular en un Estado miembro. Eurodac fue el primer sistema automatizado de reconocimiento biométrico de base institucional en la UE, y entró en funcionamiento en enero de 2003, con base en el Reglamento (UE) 2752/ 2000, para la aplicación de la Convención de Dublín¹⁴⁸. Se trata de un sistema de chequeo que comprueba, en una base de datos centralizada, si unas huellas dactilares concreta están o no registradas¹⁴⁹ y consta que pertenecen a alguien que ha solicitado asilo en otro país, alguien a quien se le denegó o alguien a quien, por algún motivo, no le está permitido solicitar asilo. Con ello trata de evitar el *asylum Shopping*, esto es, la solicitud simultánea de asilo en varios países con el objetivo de escoger el más favorable o los intentos de acceder con diferentes identidades¹⁵⁰. No obstante, su uso se ha extendido hasta abarcar no sólo a los que

146. Reglamento (UE) 2018/1860 sobre la utilización del Sistema de Información de Schengen para el retorno de nacionales de terceros países en situación irregular; Reglamento (UE) 2018/1861 relativo al establecimiento, funcionamiento y utilización del Sistema de Información de Schengen (SIS) en el ámbito de las inspecciones fronterizas; Reglamento (UE) 2018/1862 relativo al establecimiento, funcionamiento y utilización del Sistema de Información de Schengen (SIS) en el ámbito de la cooperación policial y de la cooperación judicial en materia penal.

147. Propuesta modificada de Reglamento relativo a la creación del sistema «Eurodac» para la comparación de datos biométricos para la aplicación efectiva de dos reglamentos futuros, el Reglamento sobre la gestión del asilo y la migración, y el Reglamento sobre reasentamiento, y por el que se modifican los Reglamentos (UE) 2018/1240 y (UE) 2019/818 — COM(2020) 614 final.

148. La Convención de Dublín se sustituyó por el Reglamento del (CE) n° 343/2003 del Consejo.

149. Van der Ploeg, I./ Sprenkels, I. «Migration and the Machine-Readable Body», en Van der Plöeg/ Sprenkels (eds), *Migration and the New Technological Borders of Europe*, Springer, 2011, 83-84.

150. El 1 de junio de 2013, eu-LISA asumió la gestión operativa diaria de EURODAC de la Comisión. El servidor central es un sistema totalmente automatizado. En 2015 entró en vigor el nuevo Reglamento EURODAC (603/2013), que permite a las fuerzas policiales nacionales y a EUROPOL acceder a la base de datos para la prevención, investigación y detección de actividades delictivas. El 4 de mayo de 2016, la Comisión Europea propuso (COM 2016/0132 COD) para reforzar y ampliar el Reglamento EURODAC, y en 2018, en un acuerdo provisional, el Parlamento y el Consejo acordaron una ampliación del sistema. Como parte del pacto más amplio sobre migración y asilo, la Comisión presentó una propuesta modificada el 23 de septiembre de 2020 (COM (2020) 614). Si

efectivamente solicitan asilo, sino a todos los inmigrantes *potencialmente* irregulares de cada vez más corta edad.

El tercer y último de los sistemas previos a los Reglamentos de Interoperabilidad, es el Sistema de Información de Visados (VIS)¹⁵¹, que funciona con datos relativos a los titulares de visados de corta duración. El sistema VIS es un sistema creado como apoyo para emitir un tipo de visado que es válido para toda la UE y que sustituye a los que emitían con anterioridad los Estados. El Reglamento CE n.º 767/ 2008 sobre el sistema de Información de Visados previó una aplicación progresiva del sistema VIS, de acuerdo con criterios tales como el riesgo de inmigración ilegal, las amenazas a la seguridad interna de los Estados miembros y la viabilidad para una recogida de datos biométricos de todas las localidades de una región. Este sistema de información de visados se puso en marcha en 2009 respecto de algunas regiones en el mundo, en especial pensando en aquellos con más dificultades para proporcionar documentos de viaje indubitados a sus nacionales. Fueron seleccionadas el Norte de África, Oriente Próximo y la región del Golfo.

En el momento de elaboración de los Reglamentos de Interoperabilidad, estaban, por su parte, en preparación tres sistemas adicionales, dos de los cuáles —desde su diseño— presentaban un elevado grado de interoperabilidad tanto entre sí como respecto al VIS. Son los siguientes:

1. Un Sistema de Entradas y Salidas (SES)¹⁵², que ha sido adoptado y sustituirá al actual sistema de sellado manual de pasaportes y registrará electrónicamente el nombre, el tipo de documento de viaje, los datos biométricos y la fecha y el lugar de entrada y salida de los nacionales de terceros países que visiten el espacio Schengen para una estancia de corta duración.

2. Un Sistema Europeo de Información y Autorización de Viajes (ETIAS)¹⁵³, que, una vez adoptado, será un sistema automatizado en gran medida que recopilará y

se acepta, la propuesta introduciría la obligación de almacenar datos sobre nombres, nacionalidades, lugar y fecha de nacimiento, e información sobre documentos de viaje; para los solicitantes de asilo, la obligación es conservar el número de solicitud de asilo y el Estado miembro responsable según el Reglamento de Dublín.

151. Sistema de información de visados: Propuesta de Reglamento que modificará el Reglamento (CE) n.º 767/2008, el Reglamento (CE) n.º 810/2009, el Reglamento (UE) 2017/2226, el Reglamento (UE) 2016/399, el Reglamento XX/2018 [Reglamento sobre interoperabilidad] y la Decisión 2004/512/CE, y se deroga la Decisión 2008/633/JAI del Consejo — COM (2018) 302 final.
152. Reglamento (UE) 2017/2226 por el que se establece un Sistema de Entradas y Salidas (SES) para registrar los datos de entrada y salida y de denegación de entrada relativos a nacionales de terceros países que crucen las fronteras exteriores de los Estados miembros, se determinan las condiciones de acceso al SES con fines policiales y se modifican el Convenio de aplicación del Acuerdo de Schengen y los Reglamentos (CE) n.º 767/2008 y (UE) n.º 1077/2011.
153. Sistema Europeo de Información y Autorización de Viajes. Reglamento (UE) 2018/1240 del Parlamento Europeo y del Consejo, de 12 de septiembre de 2018, por el que se establece un Sistema Europeo de Información y Autorización de Viajes (SEIAV) y por el que se modifican los Reglamentos (UE) n.º 1077/2011, (UE) n.º 515/2014, (UE) 2016/399, (UE) 2016/1624 y (UE) 2017/2226 (DO L 236 de 19.9.2018, p. 1).. Reglamento (UE) 2018/1241 del Parlamento Europeo y del Consejo, de 12 de septiembre de 2018, por el que se modifica el Reglamento (UE) 2016/794 con objeto

verificará la información relacionada con la seguridad presentada por los nacionales de terceros países exentos de visado antes de su viaje al espacio Schengen.

3. Y un Sistema Europeo de Información de Antecedentes Penales para nacionales de terceros países (sistema ECRIS-TCN)¹⁵⁴, que, una vez adoptado, sería un sistema electrónico de intercambio de información sobre condenas anteriores dictadas contra nacionales de terceros países por tribunales penales de la UE.

Junto con las citadas bases, la estrategia de interoperabilidad preveía incluir también conexiones con la base de datos de Interpol sobre documentos de viaje robados y perdidos (SLTD), que debería consultarse sistemáticamente en las fronteras exteriores de la UE, y los datos de Europol. No se preveía interoperabilidad, en cambio, con los sistemas de información nacionales ni con los sistemas de información descentralizados de la UE.

El trabajo legislativo y operativo hacia la interoperabilidad, así como los fondos empleados al efecto, son prueba de la relevancia estratégica que la UE reconoce a estas herramientas de reconocimiento. De hecho, además de las bases relativas al tránsito transfronterizo, se avanza de forma decidida en la cooperación policial para la persecución de los delitos graves. A finales de 2023, por ejemplo, se llegó a un acuerdo sobre el intercambio automatizado de datos para la cooperación policial el marco del Tratado de Prüm, que permite a las autoridades encargadas de hacer cumplir la ley consultar las bases de datos nacionales de otros estados miembros en lo que respecta a datos de ADN, huellas dactilares y matriculación de vehículos. El nuevo Reglamento Prüm, en el que se recogerán los acuerdos alcanzados, supondrá la instalación de un enrutador por parte de eu-LISA (la agencia de la UE encargada de los grandes sistemas informáticos, como el Sistema de Información de Schengen) para facilitar el establecimiento de conexiones entre los Estados miembros (y Europol) con el fin de recuperar datos. El enrutador constará de una herramienta de búsqueda y un canal de comunicación seguro, y enviará la solicitud de consulta presentada en un Estado miembro a todos los Estados Miembros y a Europol.

Entre las mayores críticas a la interoperabilidad cabe citar aquellas que rechazan entrelazar, si quiera mediante la posibilidad de cotejo, el control de la migración y la persecución de los delitos más graves¹⁵⁵, mezclando —en un punto de no

de establecer el Sistema Europeo de Información y Autorización de Viajes (SEIAV) (DO L 236 de 19.9.2018, p. 72).

154. Sistema Europeo de Información de Antecedentes Penales de nacionales de terceros países y apátridas. Reglamento (UE) 2019/816 del Parlamento Europeo y del Consejo, de 17 de abril de 2019, por el que se establece un sistema centralizado para la identificación de los Estados miembros que poseen información sobre condenas de nacionales de terceros países y apátridas (ECRIS-TCN) a fin de complementar el Sistema Europeo de Información de Antecedentes Penales, y por el que se modifica el Reglamento (UE) 2018/1726 (DO L 135 de 22.5.2019, p. 1).
155. Vavuola, N., «The recast Eurodac regulation: are asylum seekers treated as suspected “criminals”?», en C. Bauloz et al., eds. *Seeking asylum in the European Union: selected protection issues raised by the second phase of the common European asylum system*, Brill, 2015, 247-273; Queiroz, M.B., 2019. The impact of EURODAC in European migration law: the era of crimmigration? *Market and competition Law review*, 3 (1), 157-183.

retorno— bases de datos creadas originalmente con finalidades muy diversas¹⁵⁶. Se ha criticado además que UE financie, en los países de origen, el desarrollo del programa transnacional WAPIS. Se trata éste de un sistema gestionado por INTERPOL y destinado a mejorar la capacidad de los organismos encargados de hacer cumplir la ley en África Occidental para combatir, mediante el intercambio transfronterizo, el crimen organizado transnacional y el terrorismo. Su base, no obstante, es una digitalización de los datos biométricos de ciudadanos africanos en formatos posteriormente legibles con bases de datos y organizaciones internacionales más grandes, como Frontex. Atendiendo a esa circunstancia, las voces críticas hablan de una «desterritorialización» de las fronteras exteriores de la UE¹⁵⁷, trasplantadas de algún modo a determinados países africanos. Desde 2017, WAPIS se aplica en todos los estados miembros de la Comunidad Económica de Estados de África Occidental¹⁵⁸, fundada por el Tratado de Lagos, y en Mauritania.

VII. REFLEXIONES FINALES

En el tiempo de tramitación de los actos legislativos que introdujeron en la UE el Pasaporte biométrico, el profesor Stefano Rodotà advertía que el cuerpo humano «había pasado a ser un *password*». Era un tiempo en el que la protección de datos personales comenzaba a caracterizarse por ser escenario de fuertes contradicciones, reflejando una «verdadera y propia esquizofrenia social, política e institucional»¹⁵⁹.

Las tecnologías de reconocimiento deberían haber comenzado a testarse no tanto por las necesidades que podían llegar a cumplir, como por la admisibilidad real de los usos concretos que respecto a ellos se pretendían. Aquel tiempo, el posterior a los atentados del 11-S, no parecía buena ocasión para reflexionar sobre cuánto espacio quieramos permitir a las tecnologías de reconocimiento biométrico, pero no se supo prever que terminarían llegando para usos tan banales como abrir la taquilla de un gimnasio. Los mecanismos de garantía de los derechos fundamentales no estaban preparados para atender la singularidad de la información biométrica y la experiencia ha demostrado que, en no pocos casos, ha sido necesario revisar importantes decisiones judiciales. Así, por ejemplo, hace poco más de un año España tuvo que revisar la doctrina que el Tribunal Supremo había establecido respecto al control biométrico de presencia de los trabajadores. En julio de 2007¹⁶⁰ dictó una sentencia

156. Bunyan, T., «The “point of no return” interoperability morphs into the creation of a Big Brother centralized EU state database including all existing and future», *Justice and Home Affairs databases. Statewatch Analysis*, 2018.

157. Oliveira Martins, B/ e. al., «Border security», cit., *European Security*, 2022, 31(3), 475-494.

158. Son miembros Benín, Burkina Faso, Cabo Verde, Costa de Marfil, Gambia, Ghana, Guinea, Guinea-Bisáu, Liberia, Mali (suspendido desde 2021), Níger (suspendido desde 2023), Nigeria, Senegal, Sierra Leona y Togo.

159. En el mismo sentido N. P. Musar, en AA. VV., «Workshop report. Restrictions on the Implementation of the EU Data Protection Directive for Public Interest, security and defense», en *HIDE Newsletter*, vol. 2 (7), 2009, diciembre, 2 y ss; Harb, B./ Schmid, D.: «Der Einsatz biometrischer Systeme. Verfassungsrechtliche Aspekte», cit., 2005, 158-161.

160. Sentencia de 2 de julio de 2007, visto por la Sala de lo Contencioso-Administrativo del Tribunal Supremo, recurso de casación n.º 5017/ 2003, sobre la implantación del nuevo sistema de control horario.

en la que no parecía verse mayor problema en fichar con biometrías manuales. En su fundamento jurídico 7º, llegaba a decir el TS, «parece como si los sindicatos que han promovido el proceso vieran en un código binario de la imagen tridimensional de la mano una afrenta a la dignidad humana. Pero el alcance del sistema no llega a tanto». La unificación de criterios impulsada por el Comité Europeo de Protección de Datos estableció en abril de 2023, establece que, dado que los datos biométricos identificantes pertenecen a la categoría especial del art. 9 RGPD, éstos solo podrán ser tratados en los casos en los que una medida de rango de Ley habilite específicamente su uso, así como las garantías correspondientes.

Muy probablemente, en una década o antes necesitaremos revisar también la amplitud con la que el RIA permite el manejo de datos biométricos no singularizantes por parte de operadores privados —en especial en el ámbito del comercio—. Advierte el Profesor Francisco Balaguer Callejón que el mundo digital, que ocupa una parte cada vez más importante de nuestra realidad cotidiana, «está sometido a reglas en cuya producción el Estado prácticamente no interviene y que no se ajustan a los principios y valores constitucionales»¹⁶¹. Y en el caso de la vigilancia biométrica, encontramos una preocupante tendencia a que los seres humanos se vuelvan detectables, trazables y correlacionables sin su conocimiento ni consentimiento, y para muy diversos fines¹⁶². Permitir la recopilación no controlada de datos corporales está muy lejos del principio de minimización.

Como reflexión final, y por cuanto se refiere al reconocimiento biométrico automatizado, el RIA no pasará a la historia por sus grandes aportaciones. Cuánto texto, para decir tan poco y tan mal. Tan pronto parece que entra a regular algún supuesto, inmediatamente se enreda en tantos matices y salvedades que resulta muy penoso llegar a determinar el alcance real de sus preceptos.

Las modalidades de verificación, aunque emplean datos personales de categoría especial, quedan fuera del RIA sin justificación alguna, y parece que se han englobado con ellas, las técnicas de identificación biométrica no remota en las que el sujeto a identificar participa de forma activa, sea o no voluntaria. Que en ambos casos manejen datos biométricos de categoría especial (art. 4.14 y art. 9 RGPD) no ha sido razón suficiente para proporcionar a estos sistemas los análisis de calidad y seguridad que, conforme al RIA, se aplicarán a las modalidades de alto riesgo, dejándolos, como mucho, pendientes de lo que pueda proponerse en las Directrices voluntarias.

Incluso lo más tangible, las prohibiciones y modalidades de alto riesgo, se van vaciando de contenido según se avanza en la lectura del texto finalmente aprobado. La exclusión más importante implica que ninguna de esas modalidades se aplica a los usos militares, de defensa y de seguridad nacional de los Estados Miembros (art.2.3RIA)¹⁶³; ni siquiera cuando sean actores privados los que proporcionan este

161. Balaguer Callejón, F., *La Constitución del Algoritmo*, 2ª ed., Fundación Manuel Giménez Ábad, Zaragoza, 2023, 33.

162. Gutwith, S. / De Hert, P., «Regulation Profiling in a Democratic Constitutional State», en M. Hildebrandt/ S. Gutwith (eds.) *Profiling the European Citizen*, 2008, 287.

163. En caso de que, y en la medida en que, los sistemas de IA se introduzcan en el mercado, se pongan en servicio o se utilicen, con o sin modificación, con fines militares, de defensa o de seguridad nacional, deben excluirse del ámbito de aplicación del presente Reglamento, independientemente del tipo de entidad que lleve a cabo esas

servicio a los Estados Miembros. Sólo algunas modalidades aplicables en el ámbito laboral o el educativo, aunque no muchas, parecen encontrar vías de prohibición o limitación, porque incluso lo comercial ha encontrado una vía de escape en el RIA.

Por si ello fuera poco, hemos de asombrarnos por la facilidad con que se admiten sistemas de categorización o de reconocimiento de emociones para usos en entornos comerciales bajo el argumento de que este tipo de prácticas —en especial cuando no emplea datos del art. 4.14— no generan tanto riesgo para los derechos de las personas como se pudiera pensar. Esa formulación, que habla de un menor riesgo, pero no de una ausencia de éste, no sirve de excusa para desarrollar las garantías debidas.

Es proceso de elaboración del RIA, encorsetado en el formato de elaboración de un reglamento de riesgo de productos, no era el momento adecuado para regular cuestiones relativas al reconocimiento biométrico automatizado, porque, como se ha visto, es necesaria una reflexión más sosegada. Y su fruto, el texto finalmente aprobado, así lo refleja. Es paradójico —y bastante triste— que, precisamente la alusión a las biometrías en el RIA, luzca en tantos discursos como muestra de un elevado compromiso con los valores de la Unión. Hubiera sido difícil hacerlo peor.

Atendiendo a su elevado nivel de invasividad, es necesaria, imprescindible, una regulación clara de los sistemas de reconocimiento biométrico, sólida en su formulación y con garantías tangibles y eficientes. Como primer paso, será necesario revisar tanto el RIA, en la práctica totalidad de previsiones que recoge sobre el reconocimiento biométrico, como algunas asunciones que, en su día, se hicieron en el RGPD.

actividades, por ejemplo, con independencia de que se trate de una entidad pública o de una entidad privada.

La prohibición de sistemas de inteligencia artificial que evalúan y clasifican a las personas a partir de datos que no guardan relación con el contexto donde se generaron y que provocan discriminaciones

MIGUEL ÁNGEL PRESNO LINERA¹

Catedrático de Derecho Constitucional de la Universidad de Oviedo

I. INTRODUCCIÓN

El RIA incluye un Considerando 31 donde se explica que «los sistemas de IA que permiten a agentes públicos o privados llevar a cabo una puntuación ciudadana de las personas físicas pueden tener resultados discriminatorios y abocar a la exclusión a determinados grupos. Pueden menoscabar el derecho a la dignidad y a la no discriminación y los valores de igualdad y justicia. Dichos sistemas de IA evalúan o clasifican a las personas físicas o a grupos de estas sobre la base de varios puntos de datos relacionados con su comportamiento social en múltiples contextos o de características personales o de su personalidad conocidas, inferidas o predichas durante determinados períodos de tiempo. La puntuación ciudadana resultante de dichos sistemas de IA puede dar lugar a un trato perjudicial o desfavorable de determinadas personas físicas o grupos enteros en contextos sociales que no guardan relación con el contexto donde se generaron o recabaron los datos originalmente, o a un trato perjudicial desproporcionado o injustificado en relación con la gravedad de su comportamiento social. Por lo tanto, deben prohibirse los sistemas de IA que impliquen esas prácticas inaceptables de puntuación y den lugar a esos resultados perjudiciales o desfavorables».

En otras palabras, estamos hablando de que tras la entrada en vigor del Reglamento quedarán prohibidos en el ámbito de la Unión Europea y tampoco podrán exportarse a otros países sistemas de IA que generen calificaciones o jerarquizaciones sociales de las personas derivadas de sus comportamientos o de sus características y que

1. Este trabajo es uno de los resultados del Proyecto PID2022-136548NB-I00 «Los retos de la inteligencia artificial para el Estado social y democrático de Derecho», financiado por el Ministerio de Ciencia e Innovación en la Convocatoria Proyectos de Generación de Conocimiento 2022.

puedan dar lugar a situaciones discriminatorias y, por tanto, vulneradoras de los principios de dignidad e igualdad.

Se trata de una cuestión de extraordinaria importancia porque están en juego elementos esenciales del Estado social y democrático de Derecho que quedarían seriamente menoscabados si se permitieran sistemas, como los apuntados, dirigidos a condicionar la conducta de la ciudadanía y capaces de generar, cuando menos, daños sociales y económicos, cuando no físicos y psíquicos².

En las siguientes páginas desarrollaremos la hipótesis del riesgo cierto que suponen los sistemas que nos ocupan³ y el acierto que, a nuestro juicio, constituye su introducción en el Reglamento Europeo como una forma de hacer frente a una de las manifestaciones crecientes del, en palabras de Shoshana Zuboff, el «capitalismo de vigilancia»⁴.

Como es obvio, no se trata de excluir cualquier tipo de «puntuación» personal que persiga modificaciones conductuales, puesto que hay sistemas no solo posibles sino, seguramente, necesarios; así, por ejemplo, el permiso de circulación por puntos sería un buen y conocido ejemplo: en palabras de la Dirección General de Tráfico su objetivo «es modificar los comportamientos y actitudes de los conductores infractores, sensibilizarlos sobre las graves consecuencias humanas, económicas y sociales que se derivan de los accidentes de tráfico y hacerles ver la implicación que sus conductas tienen en los accidentes»⁵.

En otros contextos los sistemas no estarán prohibidos pero sí serán calificados como de «alto riesgo»; así, «los sistemas de IA que se utilizan en los ámbitos del

2. Véase al respecto el trabajo de Paquale, F. y Keats Citron, D., «The Scored Society: Due Process for Automated Predictions», *Washington Law Review*, vol. 89, 2014, pp. 1-33.
3. Sobre el riesgo algorítmico San Martín Segura, D., *La intrusión jurídica del riesgo*, CEPC, Madrid, 2023, pp. 271 y ss.
4. «Capitalismo de la vigilancia, m. 1. Nuevo orden económico que reclama para sí la experiencia humana como materia prima gratuita aprovechable para una serie de prácticas comerciales ocultas de extracción, predicción y ventas. 2. Lógica económica parasítica en la que la producción de bienes y servicios se subordina a una nueva arquitectura global de modificación conductual. 3. Mutación inescrupulosa del capitalismo caracterizada por grandes concentraciones de riqueza, conocimiento y poder que no tienen precedente en la historia humana. 4. El marco fundamental de una economía de la vigilancia. 5. Amenaza tan importante para la naturaleza humana en el siglo XXI como lo fue el capitalismo industrial para el mundo natural en los siglos XIX y XX. 6. Origen de un nuevo poder instrumentario que impone su dominio sobre la sociedad y plantea alarmantes contradicciones para la democracia de mercado. 7. Movimiento que aspira a imponer un nuevo orden colectivo basado en la certeza absoluta. 8. Expropiación de derechos humanos cruciales que perfectamente puede considerarse como un golpe desde arriba: un derrocamiento de la soberanía del pueblo», *La era del capitalismo de vigilancia*, Paidós, Barcelona, 2022, 2ª edición, p. 9.
5. «El saldo de puntos puede cambiar: ser un buen conductor y/o realizar cursos de sensibilización te hacen ganar puntos. Cometer infracciones te resta, hasta llegar a cero. Si llegaras a esa situación, se procederá a tramitar la pérdida de vigencia de tu permiso y no podrás conducir ningún vehículo, aunque antes de que esto suceda, puedes recuperar puntos», disponible en <https://www.dgt.es/nuestros-servicios/permisos-de-conducir/tus-puntos-y-tus-permisos/como-funciona-el-permiso-por-puntos/> (a 18 de marzo de 2024).

empleo, la gestión de los trabajadores y el acceso al autoempleo, en particular para la contratación y la selección de personal, para la toma de decisiones que afecten a las condiciones de las relaciones de índole laboral, la promoción y la rescisión de relaciones contractuales de índole laboral, para la asignación de tareas a partir de comportamientos individuales o rasgos o características personales y para la supervisión o evaluación de las personas en el marco de las relaciones contractuales de índole laboral, dado que pueden afectar de un modo considerable a las futuras perspectivas laborales, a los medios de subsistencia de dichas personas y a los derechos de los trabajadores. Las relaciones contractuales de índole laboral deben incluir, de manera significativa, a los empleados y las personas que prestan servicios a través de plataformas, como indica el programa de trabajo de la Comisión para 2021. Dichos sistemas pueden perpetuar patrones históricos de discriminación, por ejemplo contra las mujeres, ciertos grupos de edad, las personas con discapacidad o las personas de orígenes raciales o étnicos concretos o con una orientación sexual determinada, durante todo el proceso de contratación y en la evaluación, promoción o retención de personas en las relaciones contractuales de índole laboral. Los sistemas de IA empleados para controlar el rendimiento y el comportamiento de estas personas también pueden socavar sus derechos fundamentales a la protección de los datos personales y a la intimidad» (Considerando 57 del Reglamento Europeo).

Igualmente, «el acceso a determinados servicios y prestaciones esenciales, de carácter público y privado, necesarios para que las personas puedan participar plenamente en la sociedad o mejorar su nivel de vida, y el disfrute de dichos servicios y prestaciones, es otro ámbito en el que conviene prestar especial atención a la utilización de sistemas de IA. En particular, las personas físicas que solicitan a las autoridades públicas o reciben de estas prestaciones y servicios esenciales de asistencia pública, a saber, servicios de asistencia sanitaria, prestaciones de seguridad social, servicios sociales que garantizan una protección en casos como la maternidad, la enfermedad, los accidentes laborales, la dependencia o la vejez y la pérdida de empleo, asistencia social y ayudas a la vivienda, suelen depender de dichas prestaciones y servicios y, por lo general, se encuentran en una posición de vulnerabilidad respecto de las autoridades responsables. La utilización de sistemas de IA para decidir si las autoridades deben conceder, denegar, reducir o revocar dichas prestaciones y servicios o reclamar su devolución, lo que incluye decidir, por ejemplo, si los beneficiarios tienen legítimamente derecho a dichas prestaciones y servicios, podría tener un efecto considerable en los medios de subsistencia de las personas y vulnerar sus derechos fundamentales, como el derecho a la protección social, a la no discriminación, a la dignidad humana o a la tutela judicial efectiva y, por lo tanto, deben clasificarse como de alto riesgo» (Considerando 58).

Como es sabido, y como se explica con más detalle en otros apartados de esta obra colectiva, la calificación de un sistema como de alto riesgo implica una serie de obligaciones; entre otras:

«Los sistemas de IA de alto riesgo irán acompañados de las instrucciones de uso correspondientes en un formato digital o de otro tipo adecuado, las cuales incluirán información concisa, completa, correcta y clara que sea pertinente, accesible y comprensible para los responsables del despliegue» (artículo 13.2);

«1. Los sistemas de IA de alto riesgo se diseñarán y desarrollarán de modo que puedan ser vigilados de manera efectiva por personas físicas durante el período que estén en uso, lo que incluye dotarlos de herramientas de interfaz humano-máquina adecuadas. 2. El objetivo de la vigilancia humana será prevenir o reducir al mínimo los riesgos para la salud, la seguridad o los derechos fundamentales que pueden surgir cuando se utiliza un sistema de IA de alto riesgo conforme a su finalidad prevista o cuando se le da un uso indebido razonablemente previsible...» (artículo 14.1 y 2);

«Los sistemas de IA de alto riesgo se diseñarán y desarrollarán de modo que alcancen un nivel adecuado de precisión, solidez y ciberseguridad y funcionen de manera uniforme en esos sentidos durante todo su ciclo de vida» (artículo 15.1);

«... Los sistemas de IA de alto riesgo que continúan aprendiendo tras su introducción en el mercado o puesta en servicio se desarrollarán de tal modo que se elimine o reduzca lo máximo posible el riesgo de que información de salida que puede estar sesgada influya en la información de entrada de futuras operaciones (“bucles de retroalimentación”) y se garantice que dichos bucles se subsanen debidamente con las medidas de reducción de riesgos oportunas» (artículo 15.4.3).

II. EL SISTEMA DE CRÉDITO SOCIAL CHINO

El sistema de puntuación social del que más se ha venido hablando, y que comenzó a desarrollarse incluso antes de la eclosión actual de la IA, es el de crédito social chino (CSCH en lo sucesivo); como explican Lauren Yu-Hsin Lin y Curtis J. Milhaupt⁶, la planificación de un programa integral de crédito social que complementara el débil sistema jurídico de China comenzó en la década de 1990 con el objetivo más ambicioso de hacer frente al fraude generalizado en la transición del país de la planificación central a una incipiente economía de mercado. Esos esfuerzos culminaron en 2014 con la publicación conjunta por el Comité Central del Partido Comunista Chino y el Consejo de Estado chino del *Esquema de planificación para la construcción de un sistema de crédito social (2014-2020)*, un programa integral para evaluar el crédito social de individuos, empresas, entidades gubernamentales y otras organizaciones.

En la actualidad, el sistema de crédito social es también la pieza central de la estrategia de gobernanza digital de China, que marca un cambio hacia un mercado autorregulado, es decir, en el que se presiona o incentiva a los agentes para que ajusten su comportamiento a las normas del partido-Estado más allá de los canales ordinarios de la ley y la regulación.

6. «China’s Corporate Social Credit System: The Dawn of Surveillance State Capitalism?», *The China Quarterly*, Cambridge University Press, 2023, pp. 1-19; en particular, pp.2-4; disponible, a 18 de marzo de 2024, en <https://www.cambridge.org/core/journals/china-quarterly/article/chinas-corporate-social-credit-system-the-dawn-of-surveillance-state-capitalism/EC80AC0CC9AE60D3D3C631A707A5CE54> (a 18 de febrero de 2024); véase también Rogier CREEMERS «China’s Social Credit System: An Evolving Practice of Control», 9 de mayo de 2018, disponible, a 18 de febrero de 2024, en <https://ssrn.com/abstract=3175792> y <http://dx.doi.org/10.2139/ssrn.3175792> (a 18 de marzo de 2024).

Por su parte, y en el ámbito privado, ya en 2015 Alibaba presentó su propio sistema de calificación crediticia personal, *Sesame Credit*, para recopilar información sobre la identidad personal, el historial crediticio, la fiabilidad contractual y los comportamientos y relaciones sociales. A partir de esta información, a los participantes se les asignan puntuaciones de crédito social visibles para los demás y a quienes las tienen altas se les ofrecen ventajas, como la aprobación más rápida de un préstamo⁷.

Zuboff explica que el sistema *Sesame Credit* genera una valoración «holística» del «carácter» de una persona por medio de un aprendizaje algorítmico que asimila mucho más que el hecho de que esta pague a tiempo sus facturas y los préstamos que contrata. Los algoritmos evalúan y clasifican compras (por ejemplo, el hecho de que sean videojuegos en vez de libros para niños), títulos educativos, y cosas como la cantidad y la «calidad» de las amistades. Los individuos bien puntuados reciben distinciones y recompensas de los clientes de *Sesame Credit* en sus mercados de futuros conductuales. Pueden así alquilar un coche sin pagar fianza, o recibir unos términos más favorables en ese préstamo o en ese alquiler de piso que soliciten, o ver acelerados los trámites para la obtención de visado, o ser objeto de una exposición más destacada en las aplicaciones de citas, etcétera. Sin embargo, algunos testimonios apuntan a que los privilegios asociados a una reputación personal elevada pueden tornarse súbitamente en penalizaciones por motivos en absoluto relacionados con

7. En su página web *Sesame Credit* explica: «The concept of a credit score may feel complicated, but in essence it looks simply at your payment history, amount of debt, how long you have had debt and how many recent applications you have made for credit accounts. Information about these items are reported to the three credit bureaus, Experian, TransUnion and Equifax, who compile your credit report. The information on your credit report is used to calculate your credit score. Your three-digit credit score captures your experiences with credit and debt and can help you track changes in your financial history over time, from the very first debt you encounter—such as the credit card you opened in college—up to the present. Credit score is a powerful tool that signals to prospective lenders your ability to make payments in a timely manner. This number is unique to you but publicly available under federal law to lenders considering you as a borrower. Your score can be a point of personal pride for good financial management and a point of public documentation. A credit score is an easy way to explain to another person or prospective lender that you can honor your commitment to make timely payments on outstanding debts. In turn, higher scores might lead a lender to extend interest rates lower than they would for consumers with less-favorable credit scores. You can get your credit score as part of a request for a credit report or independently of a credit report. A comprehensive solution is to open a free Credit Sesame account. This provides you with fast access to everything you need to know about your credit history, including your credit score. It includes helpful supporting information that makes sense of your score and report...

Legally, a variety of entities and people can request a copy of your credit report, which is the information that feeds into your credit score. According to the Consumer Financial Protection

Bureau (CFPB), this list includes: Businesses to whom you owe money, Government agencies.

Landlords, Employers, Insurance providers, Banks and financial providers, Legal entities (in the event of court orders, for example), Others you have authorized in writing to receive a copy»; disponible, a 18 de marzo de 2024, <https://www.credit-sesame.com/knowledge-hub/what-is-credit-score/>

el comportamiento de la persona en su faceta como consumidor: basta, por ejemplo, con que haya hecho trampa en un examen en la universidad⁸.

Volviendo al CSCH, tiene dos características principales: la primera es la recopilación de datos a escala nacional procedentes de un amplio abanico de organismos reguladores, Gobiernos centrales y locales, el Poder Judicial y plataformas privadas. Cuando esté plenamente operativo, el sistema recopilará dos tipos básicos de información: la crediticia pública, generada por las interacciones de una empresa con órganos gubernamentales y agencias reguladoras (multas, sentencias, licencias comerciales...), y la información crediticia de mercado, generada por las interacciones de una empresa con otros agentes del mercado (reclamaciones de consumidores, datos generados por agencias de calificación de créditos...). Los datos se utilizarán en sistemas de puntuación gestionados por las administraciones locales, la mayoría de los cuales están en fase de construcción.

El segundo elemento principal del CSCH es un régimen de recompensas y castigos (en forma de «listas rojas» y «listas negras») mantenido por organismos gubernamentales. Algunas listas tienen un amplio alcance, como el incumplimiento de sentencias judiciales, mientras que otras se aplican a sectores específicos de la economía, como la alimentación o la medicina.

La inclusión en una lista roja o negra es pública; en el primer caso puede implicar diversos beneficios, que van desde la ampliación del acceso a los préstamos hasta una reducción de la frecuencia de las inspecciones o el aumento de las oportunidades en los procesos de contratación pública y acceso a la financiación, sobre todo para las pequeñas y medianas entidades. La inclusión en una lista negra origina barreras de mercado, como restricciones para obtener autorizaciones gubernamentales, mayor frecuencia de inspecciones y prohibiciones para obtener financiación. Cuando una entidad es incluida en una lista negra, su representante legal y las personas directamente responsables de la infracción también se incluirán en la lista⁹.

III. EL DESARROLLO DE LOS SISTEMAS DE CALIFICACIÓN COMO UNA VÍA DE EXPANSIÓN DEL CAPITALISMO DE VIGILANCIA

En una nota al principio de estas páginas recogimos las definiciones de «capitalismo de vigilancia» que propone Zuboff y las dos primeras acepciones, con algunas puntualizaciones, parecen englobar prácticas como las que caracterizan el sistema de crédito social chino: serían, en primer lugar, parte de un nuevo orden

8. *Ob. cit.*, pp. 520 y 521.

9. Yu-Hsin Lin y Curtis J. Milhaupt, *ob. Cit.*, pp. 3-4; con más amplitud, Schaffer, K., «China's social credit system: context, competition, technology and geopolitics.» *Trivium China*, 16 de noviembre de 2020, disponible, a 18 de marzo de 2024, en https://www.uscc.gov/sites/default/files/2020-12/Chinas_Corporate_Social_Credit_System.pdf Véanse también Lam T. «The People's Algorithms: Social Credits and the Rise of China's Big (Br)other», en Mennicken, A. Salais, R. (eds) *The New Politics of Numbers. Executive Politics and Governance*, Palgrave Macmillan, 2022; pp. 71-95; en especial, pp. 78 ss.; Xu XU, Kostka, G. y Cao, X. «Information Control and Public Support for Social Credit Systems in China», *The Journal of Politics*, Vol. 84, n.º 4, 2022, pp. 2231-2245, <https://www.journals.uchicago.edu/doi/10.1086/718358> (a 18 de marzo de 2024).

económico-político que reclama para sí la experiencia humana como materia prima gratuita aprovechable para una serie de prácticas políticas, sociales y comerciales ocultas de extracción, predicción y ventas; en segundo término, estarían presididas por una lógica parasítica en la que la producción de bienes y servicios se subordina a una nueva arquitectura global de modificación conductual.

No parece que la previsión del RIA prohibiendo los sistemas de IA que proporcionan calificaciones sociales de personas físicas para su uso con fines generales esté pensando en la implantación o el uso en Europa de sistemas como el del crédito social chino: en los países de la Unión Europea y en otros Estados democráticos la vida privada y los datos personales gozan de un elevado nivel de protección jurídica y existe mayor grado de preocupación social por las amenazas que herramientas de esta naturaleza suponen para esos derechos y para el propio libre desarrollo de la personalidad individual; como resultado no se han desarrollado prácticas propias de sociedades totalitarias como el llamado «*dang'an*», el expediente personal de múltiples y variados aspectos de cada uno de los cientos de millones de habitantes urbanos que se va actualizando desde su infancia y durante el resto de sus vidas. Este “sistema de la era Mao para el registro de los más íntimos detalles de la vida” se nutre de la información actualizada que suministran los profesores, los funcionarios del Partido Comunista y los empleadores. Los ciudadanos no tienen derecho alguno a revisar el contenido de sus propios ficheros ni, menos aún, a impugnarlo¹⁰.

No obstante estas diferencias entre el «ecosistema» europeo y el chino, debe tenerse en cuenta para matizarlas que, en primer lugar, entre nosotros está presente la llamada «paradoja de la privacidad»: mientras los individuos se declaran preocupados por su privacidad y le otorgan una valoración importante, sus decisiones son significativamente incoherentes con la valoración que confiesan, puesto que hacen poco o, esencialmente, nada para proteger sus datos personales y, por lo tanto, su privacidad¹¹.

Y, en segundo lugar, aunque no parezca próxima en Europa la consolidación de un capitalismo de vigilancia estatal autoritario como el chino no quiere decir que no haya ya prácticas de capitalismo de vigilancia empresarial que, parafraseando de nuevo a Zuboff, se valen de la experiencia humana como materia prima gratuita aprovechable para una serie de prácticas comerciales y laborales ocultas de extracción, predicción y ventas presididas por una lógica parasítica en la que la producción de bienes y servicios y las relaciones laborales poco a poco se van subordinando a una nueva arquitectura global de modificación conductual.

En palabras de Creemers, esta «tendencia a la ingeniería social y a “empujar” a los individuos hacia un comportamiento “mejor” también forma parte del enfoque de Silicon Valley, que sostiene que los problemas humanos pueden resolverse

10. Zuboff, *ob. cit.*, p. 524.

11. Artigot Golobardes, M. «Mercados digitales, inteligencia artificial y consumidores», *El Cronista El Cronista del Estado social y democrático de Derecho*, n.º 100, 2022, pp. 130 y 131; de manera más extensa, Barth y De Jong, «The privacy paradox — Investigating Discrepancies between expressed privacy concerns and actual online behavior — A systematic literature review», *Telematics and Informatics*, 34(7) (2017); Norberg, P. A. y Horne D. A. «The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors», *Journal of Consumer Affairs*, 41 (1), 2007, pp. 100-126.

de una vez por todas mediante el poder perturbador de la tecnología. Los seres humanos se reducen a un conjunto de números que indican su rendimiento en escalas preestablecidas, en sus hábitos alimentarios, por ejemplo, o en su régimen de ejercicio físico, que luego se les reta a mejorar. El mero hecho de que exista información significa que las empresas y los gobiernos tratarán de aprovecharla para sus propios fines, ya sean políticos o comerciales. En ese sentido, quizá el elemento más chocante de la historia no sea la agenda del gobierno chino, sino lo similar que es al camino que está tomando la tecnología en otros lugares»¹².

Y para mencionar un ejemplo concreto en España de la utilización de datos que una empresa ha venido empleando con el objetivo de evitar las exigencias derivadas de una relación laboral dependiente y al tiempo para «empujar» conductualmente a los trabajadores a estar disponibles el mayor tiempo posible para conseguir más encargos y, en suma, mayor retribución, cabe recordar, aunque resulte un poco extenso, lo dicho por la Sala de lo Social del Tribunal Supremo español en su sentencia de 25 de septiembre de 2020 que resolvió sobre la condición de trabajadores por cuenta ajena de los repartidores de GLOVO:

«Antecedente de hecho séptimo.- La Empresa tiene establecido un sistema de puntuación de los “glovers”, clasificándolos en tres categorías: principiante, junior y senior. Si un repartidor lleva más de tres meses sin aceptar ningún servicio, la Empresa puede decidir bajarle de categoría. (Cláusula cuarta del contrato de prestación de servicios). El sistema de ranking utilizado por GLOVO ha tenido dos versiones diferentes: la versión *fidelity*, que se utilizó hasta julio de 2017, y la versión *excellence*, utilizada desde dicha fecha en adelante. En ambos sistemas la puntuación del repartidor se nutre de tres factores: La valoración del cliente final, la eficiencia demostrada en la realización de los pedidos más recientes, y la realización de los servicios en las horas de mayor demanda, denominadas por la Empresa “horas diamante”. La puntuación máxima que se puede obtener es de 5 puntos. Existe una penalización de 0,3 puntos cada vez que un repartidor no está operativo en la franja horaria previamente reservada por él. Si la no disponibilidad obedece a una causa justificada, existe un procedimiento para comunicarlo y justificar dicha causa, evitando el efecto penalizador... Los repartidores que tienen mejor puntuación gozan de preferencia de acceso a los servicios o recados que vayan entrando...

Fundamento jurídico decimotavo.- ... En la práctica este sistema de puntuación de cada repartidor condiciona su libertad de elección de horarios porque si no está disponible para prestar servicios en las franjas horarias con más demanda, su puntuación disminuye y con ella la posibilidad de que en el futuro se le encarguen más servicios y conseguir la rentabilidad económica que busca, lo que equivale a perder empleo y retribución. Además la empresa penaliza a los repartidores, dejando de asignarles pedidos, cuando no estén operativos en las franjas reservadas, salvo causa justificada debidamente comunicada y acreditada.

12. *China's chilling plan to use social credit ratings to keep score on its citizens*, CNN, 27 de octubre de 2015, <https://edition.cnn.com/2015/10/27/opinions/china-social-credit-score-creemers/index.html> (a 18 de marzo de 2024).

La consecuencia es que los repartidores compiten entre sí por las franjas horarias más productivas, existiendo una inseguridad económica derivada de la retribución a comisión sin garantía alguna de encargos mínimos, que propicia que los repartidores intenten estar disponibles el mayor período de tiempo posible para acceder a más encargos y a una mayor retribución.

Fundamento jurídico vigésimo primero.- Glovo no es una mera intermediaria en la contratación de servicios entre comercios y repartidores. No se limita a prestar un servicio electrónico de intermediación consistente en poner en contacto a consumidores (los clientes) y auténticos trabajadores autónomos, sino que realiza una labor de coordinación y organización del servicio productivo. Se trata de una empresa que presta servicios de recadería y mensajería fijando el precio y condiciones de pago del servicio, así como las condiciones esenciales para la prestación de dicho servicio. Y es titular de los activos esenciales para la realización de la actividad... La empresa ha establecido instrucciones que le permiten controlar el proceso productivo. Glovo ha establecido medios de control que operan sobre la actividad y no solo sobre el resultado mediante la gestión algorítmica del servicio, las valoraciones de los repartidores y la geolocalización constante... Para prestar estos servicios Glovo se sirve de un programa informático que asigna los servicios en función de la valoración de cada repartidor, lo que condiciona decisivamente la teórica libertad de elección de horarios y de rechazar pedidos. Además Glovo disfruta de un poder para sancionar a sus repartidores por una pluralidad de conductas diferentes, que es una manifestación del poder directivo del empleador. A través de la plataforma digital, Glovo lleva a cabo un control en tiempo real de la prestación del servicio, sin que el repartidor pueda realizar su tarea desvinculado de dicha plataforma...».

Cabe mencionar otros ejemplos en el ámbito laboral; así, Todolí Signes explica, en una cita también extensa, que «el trabajo en un *Call center* es uno de los más afectados por este alto nivel de monitorización. Los algoritmos controlan el número de llamadas atendidas, la duración de las mismas, las pausas, incluso el contenido de la llamada a través de la detección de palabras clave, el tono de voz y la entonación... La empresa CallMiner anuncia que su software puede evaluar y poner nota —y hacer un ranking entre los trabajadores— en profesionalidad, cortesía y empatía en la atención mostrada durante las llamadas... En el mismo sentido, en los supermercados se contabiliza cómo de rápido cada cajero escanea los productos de la cesta de la compra pudiendo compararlos con el resto de trabajadores a efectos de retribuciones, asignación de turnos de trabajo, despidos de los menos rápidos y para hacer competir entre sí a los cajeros para que aceleren el ritmo de trabajo. El trabajo con ordenadores, sea en oficina o teletrabajo, es otro de los que están sujetos a un control absoluto de los tiempos de trabajo y una posterior evaluación del mismo mediante algoritmos a través de índices de productividad. La empresa Crossover ofrece una herramienta llamada *WorkSamart* para monitorizar ordenadores. Este programa cuenta las pulsaciones del teclado y el ratón, la pantalla del ordenador, los emails enviados e incluso realiza una foto cada diez minutos a través de la webcam del ordenador. De esta forma, cada segundo de inactividad con el ordenador —que no significa que el trabajador no esté pensando o trabajando con una libreta— se penaliza...

Los trabajos presenciales no se libran de este tipo de controles y rankings de productividad. Existen en el transporte, la limpieza, la hostelería, etc. El ejemplo más conocido es el control que realiza Amazon a los mozos de almacén midiendo el número y velocidad de cajas empaquetadas, los paso dados en un día dentro de los almacenes, las pausas para ir al baño o socializar, etc. Así, mediante pulseras inteligentes o chips en las botas se realiza un exhaustivo recuento del trabajo prestado y junto otras variables se elabora un índice de productividad que es usado para generar avisos automáticos (la pulsera vibra o se le envía un mensaje a la misma) o se despiden de forma automatizada a las personas que no alcanzan una mínima productividad. De acuerdo con los datos, el 10% de los trabajadores de almacén de Amazon en EEUU han sido despedidos por el índice de productividad»¹³.

Finalmente, y por acercarnos de manera breve un ámbito diferente como es el de los contratos de seguro, un ejemplo clásico es el uso de la calificación crediticia de los asegurados para fijar la prima en los seguros de automóvil, que como recuerda María Luisa Muñoz Paredes, dio lugar a un movimiento de rechazo en Estados Unidos, tras la averiguación hecha por la Asociación Consumer Reports en el año 2015 de que se atendía más a ese factor que a otros más influyentes en el riesgo, como es el historial de conducción del asegurado¹⁴. A este respecto, en el Considerando 37 del RIA se recuerda que los sistemas de IA destinados a ser utilizados para la evaluación de riesgos y la fijación de precios en relación con las personas físicas para los seguros de salud y de vida también pueden tener un impacto significativo en los medios de vida de las personas y, si no se diseñan, desarrollan y utilizan debidamente, pueden vulnerar sus derechos fundamentales y acarrear graves consecuencias para la vida y la salud de las personas, incluida la exclusión financiera y la discriminación.

Con las previsiones contenidas en el Reglamento, algunas de estas herramientas, como ya se ha apuntado antes, serán consideradas sistemas de «alto riesgo» si los datos que se usan provienen del propio contexto en el que se aplican los resultados de las evaluaciones y podrán prohibirse en el supuesto de que provengan de contextos diferentes y generen discriminaciones.

13. «La inteligencia artificial no te robará tu trabajo, sino tu salario. Retos del Derecho del Trabajo frente a la dirección algorítmica del trabajo», *El Cronista del Estado social y democrático de Derecho*, n.º 100, 2022, pp. 155 y 156; de manera más extensa, y del mismo autor, *Algoritmos productivos y extractivos. Cómo regular la digitalización para mejorar el empleo e incentivar la innovación*, Aranzadi, 2023.

14. «Big Data, IA y seguro: riesgos de inasegurabilidad y discriminación entre asegurados», *El Cronista del Estado social y democrático de Derecho*, n.º 100, 2022, p. 122; de manera más extensa, y de la misma autora, «“Big Data” y contrato de seguro: los datos generados por los asegurados y su utilización por los aseguradores», en Huergo Lora, A. H (dir.): *La regulación de los algoritmos*, Aranzadi, Cizur Menor, 2020, pp. 129-162; «El “Big Data” y la transformación del contrato de seguro», en Veiga, A. B. *Dimensiones y desafíos del seguro de responsabilidad civil*, Cizur Menor (Aranzadi), 2021, pp. 1017-1051; sobre el empleo en los contratos de seguro de lo que Caty O’neil llama «armas de destrucción matemática» véase su libro del mismo título, Capitán Swing, Madrid, 2017, pp. 199 ss.

IV. LA PROHIBICIÓN DE DETERMINADOS SISTEMAS QUE EVALÚAN O CLASIFICAN A LAS PERSONAS FÍSICAS

El apartado 1.c del artículo 5 del Reglamento ha tenido el siguiente discurrir desde la propuesta de la Comisión de 21 de abril de 2021 a la redacción definitiva, pasando antes por la Posición común («enfoque general») del Consejo Europeo sobre la Ley de IA, del 6 de diciembre de 2022, y por las enmiendas formuladas por el Parlamento Europeo el 14 de junio de 2023.

Quedan prohibidas las siguientes prácticas de IA

<i>Comisión</i>	<i>Consejo Europeo</i>	<i>Parlamento</i>	<i>Reglamento</i>
<p>La introducción en el mercado, la puesta en servicio o la utilización de sistemas de IA por parte de las autoridades públicas o en su representación con el fin de evaluar o clasificar la fiabilidad de personas físicas durante un período determinado de tiempo atendiendo a su conducta social o a características personales o de su personalidad conocidas o predichas, de forma que la clasificación social resultante provoque una o varias de las situaciones siguientes:</p> <p>i) un trato perjudicial o desfavorable hacia determinadas personas físicas o colectivos enteros en contextos sociales que no guarden</p>	<p>La introducción en el mercado, la puesta en servicio o la utilización de sistemas de IA con el fin de evaluar o clasificar a las personas físicas durante un período determinado de tiempo atendiendo a su comportamiento social o a características personales o de su personalidad conocidas o predichas, de forma que la puntuación ciudadana resultante provoque una o varias de las situaciones siguientes:</p> <p>i) un trato perjudicial o desfavorable hacia determinadas personas físicas o grupos de personas físicas en contextos sociales que no guarden relación con los contextos</p>	<p>La introducción en el mercado, la puesta en servicio o la utilización de sistemas de IA con el fin de evaluar o clasificar a las personas físicas o grupos de ellas a efectos de su calificación social durante un período determinado de tiempo atendiendo a su comportamiento social o a características personales o de su personalidad conocidas, inferidas o predichas, de forma que la puntuación ciudadana resultante provoque una o varias de las situaciones siguientes:</p> <p>i) un trato perjudicial o desfavorable hacia determinadas personas físicas o colectivos enteros en contextos sociales que no guarden relación con los</p>	<p>La introducción en el mercado, la puesta en servicio o la utilización de sistemas de IA con el fin de evaluar o clasificar a las personas físicas o grupos de personas durante un período determinado de tiempo atendiendo a su comportamiento social o a características personales o de su personalidad conocidas, inferidas o predichas, de forma que la puntuación ciudadana resultante provoque una o varias de las siguientes situaciones:</p> <p>i) un trato perjudicial o desfavorable hacia determinadas personas físicas o grupos enteros de personas en contextos sociales que no guarden relación</p>

<i>Comisión</i>	<i>Consejo Europeo</i>	<i>Parlamento</i>	<i>Reglamento</i>
relación con los contextos donde se generaron o recabaron los datos originalmente;	donde se generaron o recabaron los datos originalmente;	contextos donde se generaron o recabaron los datos originalmente;	con los contextos donde se generaron o recabaron los datos originalmente;
ii) un trato perjudicial o desfavorable hacia determinadas personas físicas o colectivos enteros que es injustificado o desproporcionado con respecto a su comportamiento social o la gravedad de este.	ii) un trato perjudicial o desfavorable hacia determinadas personas físicas o grupos de personas físicas que es injustificado o desproporcionado con respecto a su comportamiento social o la gravedad de este.	ii) un trato perjudicial o desfavorable hacia determinadas personas físicas o grupos de personas físicas que es injustificado o desproporcionado con respecto a su comportamiento social o la gravedad de este.	ii) un trato perjudicial o desfavorable hacia determinadas personas físicas o grupos de personas que sea injustificado o desproporcionado con respecto a su comportamiento social o la gravedad de éste.

Cuadro de elaboración propia.

Aunque no estamos ante uno de los preceptos que haya experimentado más cambios entre la propuesta de la Comisión y las enmiendas aprobadas por el Parlamento sí cabe destacar los realizados y, en primer lugar, uno de los más importantes es el relativo al sujeto al que se le prohíbe introducir estos sistemas: mientras que en la propuesta de la Comisión se mencionaba a «las autoridades públicas» o a quien actuara «en su representación», la posición común del Consejo, así como la enmienda del Parlamento y la redacción final resultante del acuerdo interinstitucional suprimen esa concreción y la prohibición afectará tanto a las autoridades públicas como a sujetos particulares, físicos o jurídicos, incluyendo, por tanto, a las empresas.

Esta modificación parece muy positiva porque los riesgos que se tratan de combatir pueden provenir tanto de sujetos públicos como de particulares y, como ya hemos visto, encontramos ejemplos de utilización de sistemas de puntuación por parte de empresas muy relevantes.

En segundo lugar, la propuesta de la Comisión se refería a la evaluación o clasificación de «la fiabilidad» de personas físicas mientras que la posición común del Consejo, la enmienda del Parlamento y el texto definitivo hablan de «evaluar o clasificar a las personas físicas o a grupos de personas», es decir, el análisis no se circunscribe a la «confianza» que puede generar una persona sino que alcanza a la persona como tal y, además, a partir de la enmienda del Parlamento se incluye a las personas «o grupos de personas» (piénsese, por ejemplo, en colectivos de consumidores, trabajadores, asegurados...).

En tercer lugar, los textos de la Comisión y del Consejo aunque no son idénticos —en el primer caso se atiende «a su conducta social o a características personales o de su personalidad» y en el segundo «a su comportamiento social o a características

personales o de su personalidad»— aluden a características «conocidas o predichas» mientras que la enmienda del Parlamento y la redacción final del Reglamento también incluyen las «inferidas», algo relevante porque las inferencias son conclusiones que se obtienen a partir del tratamiento de datos y esa es una de las propiedades de los sistemas de IA: la capacidad de extraer de los datos existentes nuevas informaciones.

En cuarto lugar, mientras la propuesta de la Comisión habla de «clasificación social» el Consejo y el Parlamento utilizan la expresión «puntuación ciudadana», que será la incluida finalmente en el «Reglamento», aunque no parece que la idea a la que se refieren sea distinta: la jerarquización de las personas a partir de los datos conocidos, predichos o inferidos.

Como quinta cuestión a comentar está la relativa a la generación de una o varias de las situaciones que se describen a continuación y que son las que justificarían la prohibición; la primera de ellas es que resulte un trato perjudicial o desfavorable para concretas personas o colectivos enteros en contextos sociales que no guarden relación con aquéllos donde se generaron o recabaron los datos originalmente. Se está pensando en que la puntuación resultante del tratamiento de los datos provoque una discriminación o, en palabras de los textos examinados, un «trato perjudicial o desfavorable».

A este respecto, y como hemos visto al principio, la redacción final del Considerando 31 explica que «los sistemas de IA que permiten a agentes públicos o privados llevar a cabo una puntuación ciudadana de las personas físicas pueden tener resultados discriminatorios y abocar a la exclusión a determinados grupos. Pueden menoscabar el derecho a la dignidad y a la no discriminación y los valores de igualdad y justicia».

Una matización significativa, a la que ya nos hemos referido con anterioridad, es que los datos que generan ese trato desfavorable deben haberse obtenido en contextos diferentes al de aquel en el que provocarían ese perjuicio pero nada impediría su uso en el contexto de origen; a este respecto, parece que sí se podrían emplear datos conseguidos en el seno de una relación laboral para llevar a cabo una puntuación de quienes trabajan en esa empresa o datos obtenidos en una relación contractual de prestación de servicios (por ejemplo, el suministro eléctrico) para establecer una jerarquía de precios diferentes a clientes en situaciones distintas porque una cosa es la diferencia de precios y otra la discriminación; en esta línea, la Ley 3/1991, de 10 de enero, de Competencia Desleal, en artículo 16.1 establece que «el tratamiento discriminatorio del consumidor en materia de precios y demás condiciones de venta se reputará desleal, a no ser que medie causa justificada», es decir, no sería desleal aquel trato diferente para el que exista justificación ni la mera diferencia de precios¹⁵.

Ahora bien, la no existencia de discriminación o trato perjudicial contrario a la prohibición del artículo 5.1.c no excluye que los datos empleados lo sean sin conocimiento o, incluso, sin consentimiento de la persona afectada, lo que puede situarle en una posición de especial vulnerabilidad en los mercados digitales. Por este motivo, «es necesario crear los mecanismos para evitar que dicha vulnerabilidad se

15. Véase al respecto Muñoz Paredes, M. L. «Big Data, IA y seguro: riesgos de inasegurabilidad y discriminación entre asegurados», *El Cronista del Estado social y democrático de Derecho...*, p. 123.

materialice en una expropiación de excedente contractual que el consumidor esperaba obtener de la transacción y que solo con instrumentos puramente contractuales no podrá recuperar»¹⁶.

Por otra parte, y como también se ha apuntado más arriba, que el sistema en cuestión no sea objeto de prohibición no excluye que pueda ser calificado como de «alto riesgo» en los términos ya vistos.

Finalmente, y como ya se ha apuntado, lo que no cabría es la utilización de los datos de alguien para llevar a cabo evaluaciones o clasificaciones en un contexto diferente al de su generación u obtención y que le supongan un perjuicio o trato desfavorable¹⁷; así, por ejemplo, la mayor o menor calificación crediticia de una persona no debería ser un condicionamiento para un ascenso dentro de una empresa¹⁸.

La segunda situación que justificaría la prohibición de un sistema de IA es que genere «un trato perjudicial o desfavorable hacia determinadas personas físicas o grupos de personas que sea injustificado o desproporcionado con respecto a su comportamiento social o la gravedad de este». Aquí lo que se tiene en cuenta es la manera en que una persona física interactúa con otras personas físicas o con la sociedad e influye en ellas y de ahí se deriva un trato desfavorable que o bien no está justificado o las consecuencias son desproporcionadas a su gravedad; por ejemplo, que las opiniones políticas o manifestaciones ideológicas, religiosas, sociales o culturales expresadas en una red social impliquen, con carácter general, una causa de exclusión para que una persona sea contratada o su expulsión de un centro educativo o que las calificaciones sobre amabilidad de un trabajador llevadas a cabo por los clientes supongan causa suficiente para su despido o para una sanción económica desmedida.

16. Golobardes, A. «Mercados digitales, inteligencia artificial y consumidores», *El Cronista El Cronista del Estado social y democrático de Derecho...* p. 135.

17. Como es obvio, sí pueden tener repercusión en la relación laboral comentarios o comportamientos que supongan una transgresión de la buena fe contractual o sean ofensivos para el empleador (54.2 c) y d) del Estatuto de los Trabajadores).

18. Cathy O'NEIL ofrece numerosos ejemplos de los resultados perversos derivados del uso, entre otros, de criterios de calificación crediticia en el ámbito laboral y en el del consumo en *Armas de destrucción matemática. Cómo el big data aumenta la desigualdad y amenaza la democracia...* pp. 181 ss.

El contenido de las llamadas «técnicas subliminales» y las vulnerabilidades de grupo específico de personas en el Reglamento de inteligencia artificial

LUIS MIGUEL GONZÁLEZ DE LA GARZA¹

Profesor Titular de Derecho Constitucional UNED

I. INTRODUCCIÓN

En el capítulo que vamos a estudiar analizaremos los apartados a) y b) del artículo cinco del RIA cuyas notas quizás más destacadas son los conceptos de «*técnicas subliminales*» y «*vulnerabilidades de grupo*». Antes no obstante conviene señalar dónde se ubica sistemáticamente este artículo en el marco de la norma. El uso de la IA, con sus características específicas tales como: opacidad, complejidad, dependencia de datos, comportamiento autónomo, puede afectar negativa y gravemente a una serie de derechos fundamentales y a la seguridad de los usuarios. Para abordar esas preocupaciones, el RIA sigue un sensato enfoque basado en el riesgo mediante el cual la intervención legal se adapta al nivel concreto de riesgo. Con ese fin, el RIA distingue entre sistemas de IA que presentan (i) un riesgo inaceptable (ii) un riesgo alto (iii) un riesgo limitado y (iv) un riesgo bajo o mínimo. Las aplicaciones de IA se regularían solo en la medida estrictamente necesaria para abordar niveles específicos de riesgo. El Título II (Artículo 5) de la ley de IA prohíbe explícitamente las prácticas dañinas de IA que se consideran una clara amenaza para la seguridad, los medios de vida y los derechos de las personas, debido al «riesgo inaceptable» que supone su uso. En consecuencia, estaría prohibido comercializar, prestar servicios o utilizar esas prácticas en la UE.

De lo anterior se desprende que estamos en presencia de técnicas de «riesgo inaceptable» es decir de las de mayor restricción que prevé la norma. Pero qué son —sintéticamente— esas técnicas, al margen de lo que más adelante veremos que constituyen un riesgo inaceptable. Se trata en esencia de técnicas o formas de manipulación mental destinadas a alterar de forma sustancial o relevante el comportamiento de una persona o de un grupo de personas alterando su capacidad

1. Este trabajo es resultado del proyecto de investigación «Educar en valores, construir ciudadanías», Ministerio de Ciencia e Innovación. Agencia Estatal de Investigación. Proyectos de Generación de Conocimiento 2021. Referencia: PID2021-127680OB-I00.

de formación de preferencias siendo estas conducidas mediante estrategias conductuales conocidas o que puedan desarrollarse en un futuro y en las que los sistemas de IA sean adecuados para su aplicación. El artículo que consideramos no las describe pero proyecta los principios esenciales para su identificación y concreción ya que estas pueden desplegarse de diversos modos y, sobre todo, según diversas tecnologías cualitativamente muy diversas.

No cabe a nuestro juicio duda alguna de la oportunidad y conveniencia de la necesidad de esta previsión normativa ya que como veremos dos grupos de tecnologías operan sinérgicamente en el ámbito de riesgos para las personas, el primero es la IA con su inmensa capacidad de procesamiento de datos *cuantitativos* precisos sobre las personas, grupos o colectivos caracterizados por rasgos comunes, por ejemplo los psicológicos y, por otro lado, el desarrollo de las neurotecnologías que no están destinadas al tratamiento médico de los pacientes sino aquellas que están orientadas a usos aparentemente lúdicos o de terapias no comprendidas en las regulaciones médicas pero que bajo regulaciones *laxas* —lo que sucede singularmente en los Estados Unidos como advierte Farahany²— pueden obtener tanto datos mentales como modificar conductas mediante la generación de campos electro magnéticos en respuesta al procesamiento de datos mentales procesados mediante IA. Capítulo aparte pero en conexión con cuanto consideramos se encuentran las tecnologías médicas basadas en el acceso directo al cerebro mediante bioimplantes en los que una fase esencial de procesamiento de la información se realizará mediante IA. Estas tecnologías tienen una capacidad disruptiva muy significativa ya que poseen la característica de la permanencia —al constituir sistemas de implantación fija— y no puntual como las tecnologías basadas en radiofrecuencias.

II. EVOLUCIÓN DE LA TRAMITACIÓN Y CONTENIDO

Vamos a considerar seguidamente la evolución de los apartados a) y b) del artículo 5 objeto de nuestro comentario, para ello prestaremos atención a la Propuesta de RIA y se modifican determinados actos legislativos de la Unión de 19 de octubre de 2022, documento de la Presidencia a las Delegaciones. En este texto prestamos atención al considerando 16 que señala:

«Las técnicas de manipulación basadas en la IA pueden utilizarse para persuadir a las personas para que adopten comportamientos no deseados, o para engañarlas empujándolas a tomar decisiones de una manera que subverta y perjudique su autonomía, su capacidad de decisión y su libre elección. La comercialización, la puesta en servicio o la utilización de determinados sistemas de IA destinados a distorsionar materialmente el comportamiento humano, con la probabilidad de que se produzcan daños físicos o psicológicos, son especialmente peligrosos y, por tanto, deben prohibirse. Dichos sistemas de IA despliegan componentes subliminales, como estímulos de audio, imagen o vídeo, que las personas no pueden percibir, ya que dichos estímulos están más allá de la percepción humana, u otras técnicas subliminales que subvierten o perjudican la autonomía, la toma de decisiones o la libre elección de las personas, de forma que éstas no son conscientes de ello, o incluso,

2. Farahany, Nita A, *The Battle for your Brain. Defending the right to think freely in the age of neurotechnology*, St. Martin's Press, 2023, New York, pp. 29-35.

si son conscientes, no son capaces de controlar o resistir, por ejemplo, en los casos de interfaces máquina-cerebro o realidad virtual.

Además, los sistemas de IA pueden explotar las vulnerabilidades de los niños y de un grupo específico de personas debido a su edad o a sus incapacidades físicas o mentales. Lo hacen con la intención de distorsionar materialmente la discapacidad en el sentido de la Directiva (UE) 2019/882, o una situación social o económica específica que puede hacer que esas personas sean más vulnerables a la explotación, como las personas que viven en la pobreza extrema³».

Como podemos observar una idea directriz se basa en que el vector de aplicación de estas tecnologías son los sistemas audiovisuales, es decir, la tecnología de pantallas tales como las de los ordenadores, los teléfonos móviles inteligentes en todas sus posibles configuraciones y en estas deben considerarse incluidas las diademas tipo Muse-2 entre otras o las gafas de *realidad aumentada* y las gafas de *realidad virtual* que recordemos son diversas de las primeras como en la actualidad las Apple Visión Pro o las Meta Quest 3, entre otras tecnologías. Observamos también una referencia breve a los interfaces máquina-cerebro más conocidos como BCI y al que el legislador comunitario dedicará mucha más atención en las futuras modificaciones de la norma singularmente en los considerandos, no así en el texto legal.

Se fijan en los niños y en grupos específicos de personas que debido a su edad o a sus capacidades diversas deben ser especialmente tutelados al mostrarse más vulnerables ante el potencial uso de este tipo de tecnologías, estos grupos se definen en el apartado 1⁴ del artículo 3 de la Directiva (UE) 2019/882.

En lo que respecta al contenido normativo la redacción de la Comisión de 19 de octubre de 2022 se expresa en los siguientes términos:

Artículo 5 [...] 1. Quedan prohibidas las siguientes prácticas de inteligencia artificial.

a) La comercialización, la puesta en servicio o la utilización de un sistema de IA que despliegue técnicas subliminales más allá de la conciencia de una persona con el objetivo o el efecto de distorsionar materialmente el comportamiento de una persona de manera que cause o sea *razonablemente* probable que cause a esa persona o a otra un daño físico o psicológico;

b) La comercialización, puesta en servicio o utilización de un sistema de IA que explote cualquiera de las vulnerabilidades de un grupo específico de personas debido a su edad, discapacidad física o mental o a una situación social o económica específica, con el objetivo o el efecto de distorsionar materialmente el comportamiento de una persona perteneciente a ese grupo de manera que cause o sea *razonablemente* probable que cause a esa persona o a otra un daño físico o psicológico;

Con fecha 6 de diciembre de 2022 la Secretaría General del Consejo fija un nuevo texto para las Delegaciones, en este se aprecian algunas modificaciones de interés.

3. Las itálicas son nuestros en los textos corresponden a negrillas.

4. «personas con discapacidad»: aquellas personas que tienen deficiencias físicas, mentales, intelectuales o sensoriales a largo plazo que, al interactuar con diversas barreras, puedan impedir su participación plena y efectiva en la sociedad, en igualdad de condiciones con las demás.

Comenzando por los considerandos éste es considerablemente más detallado que el de 19 de octubre de 2022 y señala con idéntica numeración:

«Las técnicas de manipulación que posibilita la IA pueden utilizarse para persuadir a las personas de que adopten comportamientos no deseados o para engañarlas empujándolas a tomar decisiones de una manera que socava y perjudica su autonomía, su toma de decisiones y su capacidad de elegir libremente. La introducción en el mercado, la puesta en servicio o el uso de determinados sistemas de IA que alteran de manera sustancial el comportamiento humano, lo que hace probable que se produzcan perjuicios físicos o psicológicos, son especialmente peligrosos y, por tanto, deben prohibirse. Estos sistemas de IA utilizan componentes subliminales, como sonidos, imágenes o estímulos de vídeo, que las personas no pueden percibir, ya que dichos estímulos trascienden la percepción humana, u otras técnicas subliminales que socavan o perjudican la autonomía, la toma de decisiones o la capacidad de elegir libremente de las personas de maneras de las que las personas no son realmente conscientes, e incluso si son conscientes de ellas, no pueden controlarlas o resistirse a ellas, por ejemplo, en los ámbitos de las interfaces cerebro-máquina o la realidad virtual. Además, los sistemas de IA también pueden explotar de otro modo las vulnerabilidades de un grupo específico de personas, derivadas de su edad, su discapacidad en el sentido de la Directiva (UE) 2019/882 o de una situación social o económica específica que puede hacerlas más vulnerables a la explotación, como las personas que viven en condiciones de pobreza extrema o las minorías étnicas o religiosas. Estos sistemas de IA pueden introducirse en el mercado, ponerse en servicio o utilizarse con el objetivo o el efecto de alterar de manera sustancial el comportamiento de una persona y de un modo que provoque o sea razonablemente probable que provoque perjuicios físicos o psicológicos a esa persona o a otra persona o grupo de personas, en particular perjuicios que pueden acumularse a lo largo del tiempo. La intención de distorsionar el comportamiento no puede darse por supuesta si la alteración es el resultado de factores externos al sistema de IA que escapen al control del proveedor o del usuario, es decir, factores que el proveedor o el usuario del sistema de IA no pueden prever ni mitigar razonablemente. En cualquier caso, no es necesario que el proveedor o el usuario tengan la intención de causar los perjuicios físicos o psicológicos, siempre que dichos perjuicios se deriven de las prácticas de manipulación o explotación que posibilita la IA. Las prohibiciones de tales prácticas de IA complementan las disposiciones de la Directiva 2005/29/CE, en particular la prohibición, en cualquier circunstancia, de las prácticas comerciales desleales que causan perjuicios económicos o financieros a los consumidores, hayan sido establecidas mediante de sistemas de IA o de otra manera. La prohibición de las prácticas de manipulación y explotación contenida en el presente Reglamento no debe afectar a las prácticas legales en el contexto de un tratamiento médico, por ejemplo, el tratamiento psicológico de una enfermedad mental o la rehabilitación física, cuando dichas prácticas se lleven a cabo de conformidad con las normas y la legislación médicas aplicables. Asimismo, no debe considerarse que las prácticas comerciales

comunes y legítimas, conformes con la legislación aplicable son, en sí mismas, prácticas de manipulación de la IA perjudiciales».

Como podemos advertir al margen de giros específicos en la redacción sobre lo ya establecido en la redacción de 19 de octubre de 2022 se añade que la prohibición de tales prácticas de IA complementaran las disposiciones de la Directiva 2005/29/CE del Parlamento Europeo y del Consejo de 11 de mayo de 2005 relativa a las prácticas comerciales desleales de las empresas en sus relaciones con los consumidores en el mercado interior, que modifica la Directiva 84/450/CEE del Consejo, las Directivas 97/7/CE, 98/27/CE y 2002/65/CE del Parlamento Europeo y del Consejo y el Reglamento (CE) n.º 2006/2004 del Parlamento Europeo y del Consejo, por ejemplo en lo que respecta a las prohibiciones consignadas en el apartado 3 de su artículo 5 cuando señala: «*Las prácticas comerciales que puedan distorsionar de manera sustancial, en un sentido que el comerciante pueda prever razonablemente, el comportamiento económico únicamente de un grupo claramente identificable de consumidores especialmente vulnerables a dichas prácticas o al producto al que se refieran, por padecer estos últimos una dolencia física o un trastorno mental o por su edad o su credulidad, deberán evaluarse desde la perspectiva del miembro medio de ese grupo. Ello se entenderá sin perjuicio de la práctica publicitaria habitual y legítima de efectuar afirmaciones exageradas o afirmaciones respecto de las cuales no se pretenda una interpretación literal.*».

Se excluye de las citadas prácticas prohibidas aquellas prácticas legales en el contexto de un tratamiento médico, citando como ejemplo el tratamiento psicológico de una enfermedad mental o la rehabilitación física, cuando dichas prácticas se lleven a cabo de conformidad con las normas, es decir la *lex artis* de los profesionales de la salud mental, pensemos en tratamientos en los que pudiese emplearse IA con pacientes que precisan manejar modelos de información apropiada a las patologías que sufren. En estos supuestos los colegios profesionales pueden tener una responsabilidad en el conocimiento, supervisión y adecuación de estas prácticas en su uso por parte de los profesionales colegiados, responsabilidad que puede compartirse con los centros universitarios de formación. Señala igualmente el considerando dieciséis que no debe considerarse que las prácticas comerciales comunes y legítimas, conforme con la legislación aplicable, son en sí mismas, prácticas de manipulación de la IA perjudiciales. Tenemos que recordar que la publicidad subliminal constituye una figura prohibida en la normativa publicitaria, en España esta prohibición se encuentra en la letra c) del artículo 3 de la Ley 34/1988, de 11 de noviembre General de Publicidad y se define en su artículo 4 para el que: «A los efectos de esta ley, será publicidad subliminal la que mediante técnicas de producción de estímulos de intensidades fronterizas con los umbrales de los sentidos o análogas, pueda actuar sobre el público destinatario sin ser conscientemente percibida». También lo está en el apartado cuarto⁵ del artículo 122 «Prohibiciones absolutas de determinadas comunicaciones audiovisuales» de Ley 13/2022, de 7 de julio, General de Comunicación Audiovisual. Hay que distinguir cuidadosamente los conceptos de la *publicidad subliminal* con otros conceptos limítrofes para no confundirlos con los de

5. 4. Se prohíbe la comunicación comercial audiovisual subliminal que mediante técnicas de producción de estímulos de intensidades fronterizas con los umbrales de los sentidos o análogas, pueda actuar sobre el público destinatario sin ser conscientemente percibida.

publicidad encubierta así como con la *publicidad sugestiva* que son a los que se refiere la parte final del considerando dieciséis y que suponen estrategias de neuromarketing legítimas, Diotto⁶. Sobre estas distinciones puede verse Tato Plaza.⁷

En lo que respecta al articulado de la redacción de 6 de diciembre de 2022, la redacción experimenta variaciones no substanciales, modificando comercialización por introducción en el mercado de forma que se avanza la frontera del acceso de la tecnología subliminal, se modifica igualmente la expresión daños por perjuicios ampliando de esta forma las posibles dimensiones de las esferas de afectación de las personas como de los grupos.

Por último veremos el texto aprobado por el Parlamento Europeo y las enmiendas aprobadas el 14 de junio de 2023 sobre sobre la propuesta de RIA.

Veremos en primer lugar las modificaciones operadas sobre el considerando dieciséis por la enmienda 38. Uno de los elementos más relevantes de la redacción del Parlamento es la introducción no ya de las prótesis cerebrales de conexión BCI como en las redacciones anteriores, aquí ya se indica expresamente que: Debe entenderse que esta *limitación comprende las neurotecnologías asistidas mediante sistemas de IA que se utilizan para supervisar, utilizar o influir en los datos neuronales recopilados a través de interfaces cerebro-ordenador, en la medida en que alteren sustancialmente el comportamiento de una persona física de una manera que cause o pueda probablemente causar un perjuicio significativo a esa misma persona o a otra*. Es decir el Parlamento Europeo tiene en consideración la interrelación neurotecnologías e inteligencia artificial de forma clara lo que nos parece importante porque una de las grandes áreas en las que la IA puede alterar sustancialmente el comportamiento de las personas vendrá determinado por las neurotecnologías y el hecho de que operen mediante sistemas de IA es y será lo habitual por la inmensa complejidad de datos que es necesario procesar para recolectar y procesar —traducir eléctricamente— las señales de salida como de entrada al cerebro. Señalado lo anterior se echa en falta mayor ambición para haber conectado este tipo de tratamientos con los denominados neuroderechos que veremos más adelante.

Es importante distinguir diversas áreas que aborda la redacción que estamos considerando, las técnicas subliminales podrían ser empleadas mediante sistemas de visualización de diversos tipos de dispositivos como ya consideramos más arriba, sin embargo las técnicas de manipulación a través del procesamiento de datos es un aspecto muy distinto ya que se trataría de operar en las entradas electroquímicas que serían traducciones de las señales manipuladas por la IA con lo que la consciencia no tendría la oportunidad si quiera de detectar tales manipulaciones y por ello y a nuestro juicio se trata de un tipo de manipulación cualitativamente distinto de las técnicas subliminales, más adelante veremos este tema con más detalle.

En lo que respecta al articulado de la redacción de 14 de junio de 2023, cabe tener en cuenta la enmienda 215 correspondiente al apartado 1, letra a) y a la enmienda 216 apartado 1, letra b que da redacción parcial al artículo cinco. Del texto de las mismas

6. Diotto, M, «*Neuromarketing. Las herramientas técnicas de una estrategia de marketing eficaz para creativos y especialistas en marketing*», Hoepli Ediciones, Madrid, 2022, pp. 131-157.

7. Tato Plaza, A, en J A, García-Cruces, *Tratado de Derecho de la Competencia y de la Publicidad*, Tomo II, Tirant lo Blanch, Valencia, 2014, pp. 1964-1967.

se puede observar que el apartado a) hace referencia a las técnicas subliminales que hagan uso de IA. Igualmente se hace referencia a *técnicas deliberadamente manipuladoras o engañosas* con el objetivo o el efecto de alterar de manera sustancial el comportamiento de una persona o un grupo de personas mermando de manera apreciable su capacidad para adoptar una decisión informada y causando así que la persona tome una decisión que de otro modo no habría tomado, técnicas estas últimas que empleen IA pero que no necesariamente sean subliminales y entre las que podría encontrarse las de base neurotecnológica a través de sistemas BCI e IA y otras técnicas menos invasivas como el *primado* o todas aquellas que explotan *sesgos cognitivos* conocidos Garrigues y de la Garza⁸ que tienen por objeto la explotación de formas de pensamiento automática características del procesamiento de la información *inconsciente*. El último párrafo señala la prohibición de los sistemas de IA que se sirvan de las técnicas subliminales que no se aplicará a los sistemas de IA destinados a ser utilizados para fines terapéuticos autorizados sobre la base de un consentimiento informado específico de las personas expuestas a ellos o, en su caso, de su tutor legal, es decir en casos de formas terapéuticas que empleen IA en el marco de la salud mental como vimos más arriba.

La letra b) conserva con carácter general la redacción de 6 de diciembre de 2022 a su vez tributaria de la de 19 de octubre de 2022 si bien añade: incluidas las características conocidas o predichas de los rasgos de personalidad o la situación social o económica de esa persona o grupo, la edad y la capacidad física o mental, en este sentido podemos pensar en la limitación de procesamiento de sistemas basados en IA que mediante BigData procesen información de personalidad pero también económica o ambas. Esta información es aquella que ya ha sido utilizada en los ámbitos de la propaganda electoral cognitiva virtual a través del microtargeting.

Finalmente y con fecha 14 de marzo de 2024 se publica el RIA adoptado por el Parlamento.⁹ En esta última modificación se cambia el considerando 16 que pasa a ser el 29 y no se producen modificaciones sustanciales con respecto a las redacciones anteriores, se mantiene la preocupación por las neurotecnologías emergentes en relación con la IA y en su capacidad de producir modificaciones sustanciales en la conducta de las personas. Creemos que hubiese sido prudente no sólo considerar los efectos de las BCI (*interfaces cerebro-máquina*) sino también contemplar la influencia a través de campos electromagnéticos inducidos que podrían ser el equivalente actual, en el siglo XXI de las técnicas subliminales del siglo XX. Las técnicas de acceso al cerebro subliminales del siglo XX se producen a través del sistema visual y auditivo, las del siglo XXI añadirán las BCI en forma de neuroimplantes y diademas mediante lectura y modificación de conducta a través campos electromagnéticos procesados por IA, como se estudia detalladamente en Garrigues y de la Garza¹⁰.

8. Garrigues Walker, A, González de la Garza, L. M, *El derecho a no ser engañados. Y cómo nos engañan y nos autoengañamos*, Thomson Reuters Aranzadi, Navarra, 2020, p. 87.

9. Reglamento de Inteligencia Artificial. Resolución legislativa del Parlamento Europeo, de 13 de marzo de 2024, sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión (COM(2021)0206 — C9-0146/2021 — 2021/0106(COD)).

10. Garrigues Walker, A, L M, González de la Garza, *Qué son los neuroderechos y cuál es su importancia para la evolución de la naturaleza humana*, Aranzadi, Navarra, 2024.

La propaganda computacional es una realidad como precisan entre otros Wooley y Howard si bien lo novedoso del tipo de propaganda que consideramos no es que sea una propaganda pasiva, sino que es una propaganda que podríamos denominar activa e inteligente debido a que se aprovecha de los *sesgos caracterológicos* de los electores para diseñar una campaña de muy alta granularidad y precisión a la medida precisa del elector y de sus preferencias emocionales y políticas. Si, por ejemplo, es un elector que se ha abstenido en otras elecciones pueden ofrecérsele argumentos basado en sus preferencias emocionales para que vote. Podemos pensar en votantes que expresen caracteres que puedan ser explotables por agentes de propaganda automatizada, votantes que no tienen una clara preferencia y a los que este tipo de propaganda puede «seguir» de forma que mediante el «*microtargeting*» o *la microsegmentación éste busque al elector para ofrecerle propaganda activa de su agrado*, capaz de aprender de la interacción con el elector en base a su personalidad y readaptarse y *refinarse* en función de las respuestas del votante en un diálogo virtual de acompañamiento propagandístico dirigido por IA *predictiva* que con anterioridad al advenimiento de estas tecnologías era inexistente.

Se denomina microtargeting porque tiene por objetivo agrupar a los electores en muy pequeños segmentos o *clústeres*¹¹ sincronizados con los 20 modelos de tipos de personalidades o *perfiles psicométricos* ya elaborados a los que se dirige este tipo de propaganda electoral. Para que la información personalizada alcance y siga a su objetivo electoral. Es usual observar en cualquier navegación por Internet que tras visitar un comercio virtual posteriormente aparece en nuestros ordenadores o teléfonos móviles información del producto o servicio que hemos visitado anteriormente, en horas, días o semanas anteriores, la publicidad sigue al navegador en determinadas páginas Web merced al uso de cookies¹² previamente aceptadas e instaladas en los equipos de los usuarios en los que esta publicidad contextual «que nos busca y acompaña» aparece. Ese seguimiento sería el equivalente del microtargeting electoral en su dimensión comercial. Pero a diferencia de ese microtargeting comercial, el electoral interactúa y aprende del elector al que tratará de *persuadir* con argumentos *racionales* intentando imitar los intereses personales, sociales y emocionales de éste y ofrecer al mismo variantes de campaña propagandística adaptadas a su perfil psicológico. Los experimentos de manipulación y contagio masivo de emociones en las redes sociales como el que se produjo en Facebook el año 2012 y que afectó a 700.000 sujetos como estudiaron Kramer, Guillory y Hancock (8788-9790:2014)¹³ demuestran convincentemente la gran efectividad de lo que se puede lograr en el ámbito de la transformación de motivaciones y preferencias mediante contagio emocional inducido.

11. De tipo sociodemográfico como distritos electorales o circunscripciones específicas disputadas, en las que pocos votos pueden producir la asignación de un escaño y en las que una actividad de propaganda cognitiva puede justificar un esfuerzo suplementario de campaña electoral para conducir a votantes que dudan en si ejercerán o no su voto a ser motivados a decantarse hacia una tendencia electoral determinada.
12. López Jiménez, D, Las cookies como instrumento para la monitorización del usuario en la red: La publicidad personalizada, *Ciencias Económicas* 29, n.º 2, 2011.
13. Kramer, Adam D.I, J E. Guillory y J T. Hancock, Experimental evidence of massive-scale emotional contagion through social networks, *PNAS*, Vol. 111, n.º 24, 17 de junio 2014.

En la campaña electoral de Donald Trump del año 2016, Cambridge Analytica, en la actualidad Emerdata tal y como señala Cadwalladr¹⁴ estaba empleando entre cuarenta y cincuenta mil variantes de diferentes argumentos electorales informativos de los que se medía su respuesta en tiempo real de los destinatarios, readaptándose a sus respuestas de forma evolutiva. La granularidad de las acciones de estos mensajes está estructurada por áreas geográficas de hasta una radio de 5 millas en las que se agrupan los perfiles psicográficos¹⁵ que se evalúan por el algoritmo de Cambridge Analytica cuyo origen se encuentra en la Universidad de Cambridge¹⁶ y que analiza mediante OCEAN¹⁷ los tipos de personalidad de los electores que posteriormente se quiere influenciar. Además, las variantes de los mensajes propagandísticos empleados actualmente no pueden ser conocidos por otros electores ya que se basan, por ejemplo, en Facebook en las publicaciones invisibles¹⁸ o *dark post* que inicialmente fueron y son un instrumento para la publicidad personalizada pero que también se puede emplear en las campañas electorales cognitivas personalizadas y que son de difícil fiscalización por una futura autoridad electoral.

Wolley y Howard¹⁹ señalan y nos adherimos a sus conclusiones que la propaganda computacional es una de las herramientas más poderosas contra la democracia ya que hace posible una genuina y nueva forma de «*ingeniería social*» *capaz de romper por completo los modelos de opinión pública y de su manipulación como han estudiado Bond, Fariss y colaboradores*²⁰. En efecto los sistemas propaganda cognitiva electoral parece que funcionan en paralelo a poderosas y profundas distorsiones de la opinión pública que están siendo originadas por muy diversos grupos de interés de alcance nacional e internacional capaces de modificar, por ejemplo —mediante granjas de ordenadores— la agenda de la opinión pública en temas de interés político mediante la manipulación de tendencias basadas en generación de hashtags hasta lograr posicionamientos

-
14. Cadwalladr, C, Google, «democracy and the truth about internet search» Internet, The Observer, 4 de diciembre de 2016. <https://www.theguardian.com/technology/2016/dec/04/google-democracy-truth-internet-search-facebook> (20 de marzo de 2024).
 15. La segmentación psicográfica es una herramienta que permite profundizar en los grupos de referencia para encontrar sus motivaciones de voto.
 16. El lector puede experimentar un análisis psicográfico básico de sus redes sociales con este algoritmo en: <https://www.psychometrics.cam.ac.uk/productsservices/apply-magic-sauce>
 17. Según Goldberg, los cinco grandes rasgos de personalidad, también llamados factores principales, reciben los siguientes nombres: factor O (apertura a las nuevas experiencias), factor C (responsabilidad), factor E (extroversión), factor A (amabilidad) y factor N (neuroticismo o inestabilidad emocional), formando así el acrónimo «OCEAN».
 18. <https://www.facebook.com/business/a/online-sales/unpublished-page-posts> (20 de agosto de 2023).
 19. Woolley, Samuel C, y Philip N. Howard, *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*, Oxford Studies, 2018.
 20. Bond, Robert M, Christopher J. Fariss, Jason J. Jones, Adam D.I. Kramer, Cameron Marlow, Jaime E. Settle & James H. Fowler, A 61-million-person experiment in social influence and political mobilization, *Nature*, Vol 489, 13 septiembre 2012.

como Trending Topics como señala Nimmo²¹. Si bien, esas tendencias son creadas de forma artificial e intencionada por medio de IA tanto por las señaladas granjas de ordenadores como por *bots*²² automatizados como advierte Ferrara²³ u otros vectores tecnológicos de generación y difusión al servicio de sus creadores. El fenómeno ha sido estudiado por Bradshaw y Howard²⁴ en el contexto internacional, hallándose un cuerpo de evidencias muy preocupante ya que la principal tarea de estas plataformas, que fue en su origen dar forma a la opinión pública a través del uso de «narrativas dinámicas» para combatir la propaganda diseminada en las redes por las organizaciones terroristas, ha cambiado en la actualidad alcanzado otras actividades completamente diversas como las de naturaleza política al demostrarse una efectividad o eficacia de estas técnicas en finalidades distintas de para las que fueron originariamente diseñadas.

Lo señalado hace referencia a la elaboración de la información falsa, posteriormente esta información es difundida o vectorizada en las redes sociales por grupos o personas, recientemente Guess, Nagler y Tucker²⁵ han estudiado qué grupos sociales —por edad— son los agentes de difusión más característicos en redes como Facebook, llegando a la conclusión de que un pequeño porcentaje de estadounidenses, menos del 8,5 por ciento compartió enlaces a los sitios de «noticias falsas» durante la campaña electoral de 2016, pero este comportamiento fue desproporcionadamente común entre las personas mayores de 65 años con independencia de la afiliación ideológica o política, los jóvenes tuvieron un papel muy inferior. Estas conductas anteriormente no reguladas pueden ser eficiente prohibidas con la redacción que hemos considerado.

III. ¿EN QUÉ CONSISTEN LAS TÉCNICAS SUBLIMINALES?

Los psicólogos saben desde hace tiempo que los estímulos débiles, degradados o de corta duración a menudo no llegan a percibirse conscientemente, pero pueden sin embargo conducir a una conducta de respuesta, como señalan Edelman y Tonomi²⁶ Hace más de cuarenta años Vance Packard en su exitoso libro *The Hidden Persuaders* hizo popular esta percepción subliminal con su célebre mensaje «Beba Coca-Cola²⁷» que se mostraba muy brevemente durante la proyección de una película con la

21. Nimmo, B, *Measuring Traffic Manipulation on Twitter*, Computational Propaganda Research Project, Oxford University, 2019.

22. Para una taxonomía de los diversos tipos de Bots aptos para ingeniería social, puede verse: Ferrara, E, Onur Varol, C Davis, F Menczer y A Flamini, *The Rise of Social Bots*, *Communications of the ACM*, julio 2016, Vol. 59, n.º 7.

23. Ferrara, E. y otros, *The Rise of Social Bots*, *Communications of the ACM*, julio 2016, Vol. 59, n.º 7.

24. Bradshaw, Samantha y Philip N. Howard, *Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation*, Working paper n.º 2017.12, Computational Propaganda Research Project, Oxford University, UK.

25. Guess A, J Nagler y J Tucker, *Less than you think: Prevalence and predictor of fake news dissemination on Facebook*, *Sci. Adv.* 2019; 5: eaau4586 9 January 2019. Puede consultarse en: <https://advances.sciencemag.org/content/advances/5/1/eaau4586.full.pdf%20> (8 de mayo de 2022).

26. Edelman Gerald, M y G Tonomi, *El universo de la conciencia*, Crítica, Barcelona, 2005, p. 88.

27. Hay autores que dudan de que existiese realmente el experimento de CocaCola.

intención de despertar la sed de los espectadores sin que ellos pudieran reconocer conscientemente el mensaje. Durante muchos años la precaria evidencia científica en apoyo de la percepción subliminal fue objeto de abundante escepticismo, pero estudios posteriores han establecido el fenómeno mediante experimentos controlados. En el laboratorio, la percepción subliminal —que en la actualidad se suele llamar *percepción inconsciente*— se suele demostrar mediante la presentación de estímulos que son demasiado débiles, cortos o ruidosos como para ser percibidos conscientemente, pero que son suficientes para aprestar o sesgar la habilidad del sujeto en la ejecución de una tarea de decisión léxica o pruebas equivalentes. Edelman y Tononi²⁸ nos recuerdan, por ejemplo, que si la palabra *paseo* se muestra durante un tiempo muy breve, la persona negará haber visto nada, si se le pregunta después una palabra que cuadre con *banco*, es más probable que la persona responda *asiento* que *dinero*. Parece evidente, entonces, que los estímulos subliminales producen suficiente activación neuronal para desencadenar una respuesta conductual apropiada. Sin embargo, hay algo en la activación neuronal —nos recuerdan los autores— producida por tales estímulos que es inadecuado o insuficiente para que surja una experiencia consciente ¿Qué es lo que falta?

Una serie de experimentos iniciada hace unos 30 años por Benjamín Libet arroja algo de luz sobre esta cuestión. En uno de ellos Libet enviaba impulsos eléctricos a 72 pulsos por segundo a través de electrodos implantados crónicamente en el tálamo del paciente para control terapéutico del dolor. Se sabe que la estimulación de ciertas partes del tálamo activa las vías neuronales que se ocupan de los estímulos táctiles y producen fácilmente una sensación identificable. El descubrimiento más sorprendente fue que unos estímulos tan débiles requerían una cantidad de tiempo notable de la actividad apropiada del cerebro, unos 500 mseg (medio segundo) antes de producir una experiencia sensorial consciente.

Libet demostró que la intención consciente de actuar sólo aparece tras un retardo de unos 350 mseg desde el principio de la actividad cerebral específica que precede a un acto voluntario. Concluyó que la iniciación cerebral de un acto voluntario espontáneo y libre puede comenzar de forma inconsciente, es decir, antes de que el sujeto se dé cuenta conscientemente de que se su intención de actuar se ha iniciado ya en el cerebro Edelman y Tononi²⁹.

Carl Jung escribió que «hay ciertos eventos de los que no nos percatamos conscientemente, que se mantienen, por así decirlo, por debajo del umbral de la conciencia. Se producen pero quedan absorbidos subliminalmente». Para Mlodinow³⁰ nuestro cerebro subliminal nos resulta invisible, pero influye en nuestra experiencia consciente del mundo de la más fundamental de las maneras: en cómo nos vemos a nosotros mismos y en cómo vemos a los demás, en los significados que atribuimos a los acontecimientos cotidianos de nuestras vidas, en nuestra capacidad para hacer juicios rápidos y decisiones que a veces significan la diferencia entre la vida y la muerte y en las acciones que iniciamos a consecuencia de todas esas experiencias instintivas.

28. Op Cit, p. 89.

29. Op Cit, pp. 90-91.

30. Mlodinow, L, *Subliminal. Cómo tu inconsciente gobierna tu comportamiento*, Crítica, Barcelona, 2018, p. 11.

Zimmerman³¹ estima que el sistema sensorial humano envía al cerebro cada segundo unos once millones de bits de información, sin embargo nuestra mente consciente no puede procesar una cantidad tan ingente de información, la que se ha estimado que puede procesar se encuentra entre dieciséis y cincuenta bits por segundo, por lo que la mente consciente no puede manejar esa inmensa cantidad de información que procesa el sistema *inconsciente*. La evolución, precisa Mlodinow³² nos ha dotado de una mente inconsciente porque el inconsciente es lo que nos permite sobrevivir en un mundo que nos exige procesar un inmenso caudal de información. La percepción sensorial, la invocación de recuerdos, las actividades, las decisiones y juicios de cada día, todo eso parece que lo hagamos sin esfuerzo, pero sólo porque el esfuerzo que exigen se realiza mayoritariamente en partes del cerebro que funcionan fuera de nuestra conciencia. Esta automatización como la describe Edelman y Tononi³³ tan generalizada en nuestra vida adulta sugiere que el control consciente sólo se ejerce en momentos críticos, cuando es necesario tomar una decisión específica o elaborar un plan concreto. Entremedio se ejecutan innumerables rutinas inconscientes que permiten que la conciencia flote libre de las ataduras de todos esos detalles y pueda dedicarse a entresacar el significado y elaborar los planes de actuación dentro del esquema global de las cosas. Según parece, tanto en la acción como en la percepción, la conciencia sólo tiene a su disposición los niveles más altos de control y análisis: todo lo demás se ejecuta automáticamente. Este rasgo ha hecho pensar a muchos que somos conscientes de los *resultados* de las *operaciones* del cerebro, pero no de las propias operaciones.

Pero además de lo anterior y como señala Metzinger³⁴ según varios estudios científicos, nuestras mentes están extraviadas entre el 30 y el 50% de nuestras fases de vigilia conscientes. Si tomamos en consideración todos los hallazgos empíricos concernientes a la *divagación mental*, alcanzamos un sorprendente resultado, de una relevancia filosófica difícil de exagerar: la autonomía mental es la excepción, la pérdida de control es la regla. Un buen número de estudios empíricos muestra que las áreas del cerebro involucradas en la divagación mental —es decir los estados mentales conscientes pero sin referente alguno— se solapan en gran escala con la conocida red de *modo-por-defecto*. La red del *modo por defecto* se activa, por lo general, durante los periodos de descanso y, como resultado, la atención se dirige hacia adentro. Esto es lo que ocurre por ejemplo cuando soñamos despiertos, se producen recuerdos inesperados o cuando pensamos sobre nosotros mismos o el futuro. En el momento en el que surge una tarea concreta que realizar se desactiva esta área del cerebro y nos concentramos inmediatamente en el problema a resolver. La hipótesis de Metzinger es que el modo por defecto sirve fundamentalmente para mantener estable y en buena forma el auto modelo autobiográfico: como un programa automático de mantenimiento, generando historias renovadas para hacernos creer que somos una y la misma persona a lo largo del tiempo, es decir, crear continuidad psicológica.

31. Zimmerman M, The Nervous System in the Context of information theory, en R.F. Schmidt y G. Thews, eds., *Human Psychology*, Springer, Berlin, 1989, pp. 166-176.

32. Op Cit, Mlodinow, L, *Subliminal. Cómo tu inconsciente gobierna tu comportamiento*, p. 45.

33. Op Cit, Edelman Gerald, M y G Tononi, *El universo de la conciencia*, p. 75.

34. Metzinger, T, *El túnel del yo. Ciencia de la mente y mito del sujeto*. Enclave, Madrid, 2018, pp. 170-171.

Bernard Baars elaboró en 1988 la *teoría del área de trabajo global* como señala Kandel³⁵ según esta teoría, la conciencia implica la difusión de información previamente inconsciente (preconsciente) a través de la corteza cerebral. Baars sugiere que el área de trabajo global abarca un sistema de circuitos neuronales que se extienden desde el tronco encefálico hasta el tálamo y desde este hasta el córtex. El neurocientífico cognitivo francés Stanislas Dehaene extrapoló el modelo psicológico de Baars al modelo biológico, Dehaene descubrió que lo que percibimos como un estado consciente es el resultado de un conjunto de circuitos neuronales que seleccionan unos datos, los amplifican y distribuyen por la corteza cerebral. La teoría de Baars y los hallazgos de Dehaene muestran que tenemos dos formas distintas de pensar en las cosas: una es *inconsciente* e implica percepción; la otra es consciente e implica la difusión de la información percibida. ¿Qué ocurre en el cerebro cuando vemos una palabra de manera subliminal, por debajo del nivel de la conciencia? En primer lugar la corteza visual se vuelve muy activa. Se trata de una actividad neuronal inconsciente: la palabra que hemos visto alcanza el centro de procesamiento visual primario de la corteza cerebral, pero al cabo de 200 o 300 milisegundos, desaparece lentamente sin alcanzar los centros superiores del córtex. Cuando una percepción se vuelve consciente ocurre otro escenario. La percepción consciente también se inicia con signos de actividad en la corteza visual, pero esa actividad, en vez de disminuir se intensifica. Al cabo de unos 300 milisegundos es muy intensa; es como un *tsunami*, no como una ola moribunda. Se propaga hacia arriba hasta la corteza prefrontal. Desde allí vuelve a donde empezó, creando un resonante circuito de actividad. Tal es la difusión de información que se produce cuando somos consciente de ella. Esta llega al *área de trabajo global* donde queda a disposición de otras regiones del cerebro.

El procesamiento inconsciente de información añade Kandel³⁶ por el contrario, se produce simultáneamente en muchas áreas distintas, pero esa información no es enviada a otras regiones. Mientras leemos estas palabras, por ejemplo, somos conscientes de nuestro entorno: el sonido ambiente, la temperatura, la humedad, los niveles de luminosidad etc. Esa información sensorial acerca del entorno se procesa de manera inconsciente en el cerebro, pero, puesto que no se difunde ampliamente no nos damos cuenta de ella mientras estamos leyendo. Los experimentos que se han realizado demuestran que la información puede entrar en el cerebro *sin que se produzca percepción consciente*. Sin embargo, esa información *puede afectar al comportamiento*. Ello se debe a que la *cerebración* inconsciente no se limita a la información sensorial. Mientras que el simple reconocimiento de una palabra se produce de manera inconsciente, a su significado se accede en niveles muchos más altos del procesamiento cerebral sin que nos demos cuenta de nada. Parece que la forma de afectar al comportamiento tendría lugar a través del *inconsciente adaptativo* idea que introdujo el psicólogo cognitivo Timothy Wilson. La función biológica del inconsciente adaptativo en la toma de decisiones, fue descubierta por Libet a quien nos hemos referido anteriormente y a la que no podemos dedicar más atención en esta breve consideración.

35. Kandel, E R, *La nueva biología de la mente. Qué nos dicen los trastornos cerebrales sobre nosotros mismos*, Paidós. Barcelona, 2022, p. 239.

36. Op Cit, pp. 241-242.

Durante el curso de la evolución humana como advierte Bargh³⁷ nuestros sistemas básicos psicológicos y de conducta fueron originariamente inconscientes y existieron antes de la tardía aparición del lenguaje y el uso consciente e intencionado de esos sistemas. El instinto fundamental de la seguridad física es un poderoso legado de nuestro pasado evolutivo y ejerce una influencia constante, respondiendo a la vida moderna a menudo de maneras sorprendentes, como por ejemplo influyendo en el voto político. La región derecha de la amígdala —el cuartel general neuronal del miedo— es más grande en las personas que se identifican como políticamente conservadoras. En las tareas de laboratorio que implicaban correr riesgos, este centro del miedo del cerebro se activa mucho más en los que se declaran republicanos que en los que se declaran demócratas. De manera que existe una conexión entre la fuerza de la motivación inconsciente por la seguridad física y las actitudes políticas de una persona. Y la investigación ha mostrado que se puede hacer a los progresistas más conservadores amenazándolos y provocándolos miedo.

Es claro que estos temores pueden, primero identificarse hoy a escala individual tan sólo con analizar con detalle los datos de navegación de millones de ciudadanos y procesarlos para identificar como veremos más adelante sus tendencias psicológicas y posteriormente manipularlos tanto de formas subliminales como directas como veremos seguidamente.

La facilidad con la que algo acude a nuestra mente se denomina «*heurística de disponibilidad*», la frecuencia con la que algo como una imagen o conjunto de ellas se empleen puede ser perfectamente inducida con precisión a través de diversos vectores informativos. La heurística de disponibilidad como argumenta Bargh³⁸ fue descubierta por Daniel Kahneman y Amos Tversky. Estos juicios de frecuencia importan en nuestra vida cotidiana porque tomamos decisiones basadas en la frecuencia con la que diversas cosas sucedan o es probable que suceda.

La conducta es un efecto inconsciente e involuntario del estado o de los estados emocionales que se procesan en la corteza insular anterior o ínsula como precisa Kandel (248: 2022)³⁹ que es una pequeña isla situada entre los lóbulos parietal y temporal. En la ínsula se reflejan los sentimientos; es como la toma de conciencia de las reacciones fisiológicas ante los estímulos emocionales. La ínsula no sólo evalúa e integra la importancia emocional o motivacional de esos estímulos, sino que también coordina la información sensorial externa y los estados motivacionales internos. Esa conciencia de los estados fisiológicos es una medida del conocimiento de uno mismo. Hay evidencias como señala Bargh⁴⁰ que apuntan a que los compradores compulsivos suelen estar deprimidos, y que las compras les ayudan a sentirse más contentos (o por lo menos no tan tristes). Que la tristeza está en la base de una gran parte de las compras compulsivas lo prueba el hecho de que los antidepresivos son

37. Bargh, John, *Sin darse cuenta. El poder del inconsciente para descubrir por qué hacemos lo que hacemos*, Penguin, Barcelona, 2023, p. 52.

38. Op Cit, p.163.

39. Op Cit, *La nueva biología de la mente. Qué nos dicen los trastornos cerebrales sobre nosotros mismos*, p. 344.

40. Op Cit, *Sin darse cuenta. El poder del inconsciente para descubrir por qué hacemos lo que hacemos*, pp. 156-159.

efectivos para reducir esa conducta compulsiva. Sin olvidar que la tristeza también predispone a pagar más por los mismos productos.

Hoy es posible conocer el estado emocional de las personas, por ejemplo, con las *diademas* neurotecnológicas de precisión como las unidades *Kernel*⁴¹ que leen los estados mentales siendo esa información muy valiosa para las organizaciones comerciales; pues bien, estas serían áreas en las que la IA que estamos considerando no debería tener acceso si la parametrización de datos estuviera fuera de la esfera de datos médicos lo que sucede con dispositivos como NeoRythm o Muse 2 entre muchos otros dispositivos que no quedan regulados por las reglamentaciones estrictas de protección de datos médicos y en los que los datos neurobiológicos de los usuarios quedan a disposición y son tratados por multinacionales externas a la UE y bajo regulaciones de *dispositivos de uso lúdico* lo que representa un riesgo de la máxima importancia para la privacidad mental de los ciudadanos y que por primera vez se incluye en la declaración de León sobre neurotecnología europea.⁴²

Pero existen muchas otras formas de manipulación inconsciente, es importante en este sentido señalar la investigación de Zajonc sobre el efecto de la *mera exposición* el cual fue muy relevante por diversas razones Bargh.⁴³ En primer lugar mostraba cómo podemos desarrollar gustos y preferencias de forma inconsciente, sin pretenderlo, tan sólo según la frecuencia de una experiencia. Zajonc sostenía que a menudo mostramos reacciones afectivas inmediatas ante estímulos como cuadros, atardeceres, alimentos u otras personas sin pensar antes sobre ello cuidadosamente, sería lo que unos años más tarde denominaría Russell Fazio como «*actitudes automáticas*» que se identificaría posteriormente como el paradigma del *priming afectivo*. Un estudio posterior de Chris Fritch y sus colegas en el University College de Londres concluía que nuestro cerebro almacena nuestras intenciones actuales de conducta en las áreas del córtex prefrontal y premotor, pero las áreas que se utilizan para guiar ese comportamiento se encuentran en una zona anatómica distinta del cerebro: el córtex parietal. Este descubrimiento ayuda a explicar cómo pueden influir en nuestra conducta el *primado* y otras influencias inconscientes. El *primado* y las influencias externas sobre nuestra conducta pueden activar la conducta guiada —*pensamiento motivado*— en una parte del cerebro con *independencia de la intención* de realizar esa conducta, que está localizada en otra zona muy distinta del cerebro. Parece que William James tenía razón cuando en su famoso capítulo de: «*La voluntad*» (The Will) sostenía que nuestra conducta surge en realidad de fuentes inconscientes y no intencionadas, incluidas las conductas apropiadas y sugeridas por lo que estamos viendo y experimentando en un momento dado en nuestro mundo. Nuestros actos conscientes de voluntad, decía James, son actos de control sobre esos impulsos inconscientes, que permiten que algunos se manifiesten y otros no.

La mente humana es una especie de espejo: genera conductas potenciales que reflejan las situaciones y circunstancias del ambiente en el que nos encontramos:

41. <https://www.kernel.com/>

42. Declaración de León sobre la neurotecnología Europea: Un enfoque centrado en la persona y basado en los derechos humanos. Reunión informal de los Ministros de Telecomunicaciones, León 23-24 de octubre de 2023, p. 2.

43. Op Cit, *Sin darse cuenta. El poder del inconsciente para descubrir por qué hacemos lo que hacemos*, p. 179.

un vaso de agua dice «bébeme», un parterre de flores dice «ríegame», una cama dice «túmbate» y los museos dicen «admírame». Estamos todos así programados, para reaccionar a los estímulos externos. Sin que nos demos cuenta lo que vemos es lo que hacemos. Lo señalado tiene mucha importancia para alterar de manera sustancial el comportamiento de una persona o un grupo de personas de manera apreciable tomando decisiones que *no manipuladas* no habrían tomado. Cuando las modificaciones conductuales introducidas de forma inconsciente a través del *primado* y mediante diferentes medios que se refuerzan tratan de modificar ideas sociales es posible conjugarlas con las pautas de guiado conductual de la ventana de Overton a escala social⁴⁴. El cerebro humano es muy sensible a la manipulación inconsciente por razones que, como podemos ver son meramente neurobiológicas. Para concluir este apartado recordemos la investigación neurocientífica sobre los circuitos motivacionales del cerebro llevada a cabo por Mathias Pessiglione y Chris Fritch, Bargh⁴⁵ quienes han confirmado que la percepción de una recompensa activa los centros de recompensa del cerebro tanto si la persona percibe conscientemente la recompensa externa como si no. Los participantes mostraban mejor rendimiento en la tarea que tenían que realizar cuando antes de la tarea aparecía la imagen subliminal de una moneda de una libra (la recompensa por hacer la actividad bien), frente a cuando aparecía la imagen de un penique. Además, el centro de recompensa del cerebro, en el posencéfalo, estaba más activo en la condición de la libra que en la del penique. Como concluye Dehaene⁴⁶ nuestro cerebro cuenta con un conjunto de *dispositivos inconscientes inteligentes* que monitorean constantemente el mundo que nos rodea y le asigna valores que guían nuestra atención y dan forma a nuestro pensamiento. Gracias a esas etiquetas subliminales, los estímulos amorfos que nos bombardean se vuelven un paisaje de oportunidades cuidadosamente ordenadas según la relevancia para nuestras metas actuales. Por debajo de nuestro nivel de conciencia, nuestro cerebro inconsciente evalúa en todo momento las oportunidades

44. La ventana de Overton es un modelo para comprender cómo las ideas en la sociedad cambian o se cambia intencionalmente por *lobbys de poder* con el tiempo e influyen en la política. El concepto central es que los políticos están limitados en cuanto a las ideas políticas que pueden apoyar; por lo general, solo persiguen políticas que son ampliamente aceptadas en toda la sociedad como opciones políticas legítimas. Estas políticas se encuentran dentro de la Ventana Overton. Existen otras ideas políticas, pero los políticos corren el riesgo de perder el apoyo popular si defienden estas ideas. Estas políticas se encuentran fuera de la ventana Overton. Pero la ventana Overton puede cambiar y expandirse, ya sea aumentando o reduciendo el número de ideas que los políticos pueden apoyar sin arriesgar excesivamente su apoyo electoral. A veces los políticos pueden mover ellos mismos la ventana de Overton respaldando con valentía una política que se encuentra fuera de la ventana, pero esto es infrecuente. Lo más frecuente es que la ventana se mueva debido a un fenómeno mucho más complejo y dinámico, que no es fácil de controlar desde arriba: la evolución de los valores y normas sociales, si bien, los grandes grupos de poder serán los que apoyen, por ejemplo, mediante el *primado* las nuevas ideas que se quieren apoyar con múltiples finalidades. Puede verse con más detalle en: <https://www.mackinac.org/overtonwindow> (visualizado 10 octubre de 2023).

45. Op Cit, *Sin darse cuenta. El poder del inconsciente para descubrir por qué hacemos lo que hacemos*, p. 289.

46. Dehaene, Stanislas, *La conciencia en el cerebro. Descifrando el enigma de cómo el cerebro elabora nuestros pensamientos*, Siglo XXI editores, Argentina, 2015, pp. 106-107.

latentes, lo que atestigua que nuestra atención opera, en gran medida, de manera subliminal.

IV. UNA INTELIGENCIA ARTIFICIAL QUE LO PROCESA TODO

Asumamos con Metzinger⁴⁷ que pudiera identificarse el correlato neuronal de la experiencia consciente que acompaña la *mentira deliberada o a cualquier otro tipo de pensamiento inconsciente* (de hecho, los primeros candidatos están ya en discusión). A partir de ahí podríamos construir eficientes detectores de alta tecnología que no dependan ya de efectos psicológicos superficiales, como la conductividad eléctrica capilar y los cambios en el flujo sanguíneo periférico. Podría ser este un instrumento extremadamente útil en la lucha contra el crimen y el terrorismo, a la vez que cambiaría en sus fundamentos nuestro mundo social. Algo que había sido hasta ahora el paradigma de la privacidad —los contenidos de la mente— se convertirían repentinamente en un asunto público. Las formas más simples de resistencia política, como confundir a las autoridades durante los interrogatorios, desaparecerán. Por otro lado, la sociedad se beneficiará del incremento de transparencia de muchas maneras. Presos inocentes podrían salvarse de sus penas. Imaginemos que durante los debates de las campañas presidenciales se encendiera una luz roja frente a uno de los candidatos cada vez que el correlato neuronal de la mentira se activara en su cerebro. Pero la detección prácticamente infalible de la mentira haría más que eso cambiaría nuestros automodelos. Si, como ciudadanos, supiéramos que en principio los secretos han dejado de existir, que no podemos ya ocultar información al Estado, uno de los pilares de la vida cotidiana (por lo menos en occidente) el disfrute de la autonomía intelectual y de la libertad de pensamiento, añadimos nosotros, desaparecería. La mera percepción de la existencia de tales tecnologías neuroforenses sería suficiente para traer el cambio. ¿Queríamos vivir en esa sociedad?

Decimos libertad de pensamiento porque esa libertad supone infinidad de opciones mentales que aunque nunca se materialicen forman parte de nuestros ensayos mentales o premodelos⁴⁸ de actuación conductual en infinidad de contextos, pero no ser materializados no significa que no hayan sido previamente pensados y descartados, pero qué sucedería si esos pensamientos —no concluidos— pudieran ser registrados y conocidos por las tecnologías de registro mental.

Es factible pensar que un sistema de transmisión de toda esa enorme cantidad de información que es preciso monitorizar tanto en su salida como en su entrada hacia el cerebro y que podría aprovechar tecnologías ya en uso pero adaptándolas a las necesidades de las neurotecnologías, así no sería descartable que en pocas décadas nuestros teléfonos móviles puedan ser implantados cerebralmente y con tecnologías de gran ancho de banda como 6 G estemos permanente conectados y la neuroprótesis se independicen de equipos como ordenadores portátiles o centros

47. Op Cit, *El túnel del yo. Ciencia de la mente y mito del sujeto*, p. 314.

48. Las personas fantasean y proyectan en la imaginación multitud de realidades neuro-virtuales que nunca realizarán pero que aunque no se realicen quedan porque se procesan en la memoria: ideaciones suicidas, pensamientos moral o socialmente reprobables de naturaleza sexual con otras personas si se exteriorizaran o inclusive pensamientos de naturaleza criminal a los que se podría acceder en un futuro y ser las personas evaluadas o juzgadas —también— por esos pensamientos internos.

médicos, las macro corporaciones oligopolistas como Google, Facebook, Meta, Microsoft, Amazon, filtrarían mediante IA toda nuestra información de máxima sensibilidad.

Opciones imposibles e impensables hace décadas empiezan a vislumbrarse en la actualidad como posibles abriéndose, a su vez, muchas preguntas esenciales. ¿Qué sucede si al enfermo de Alzheimer se le implantan *recuerdos diferentes* de los que originariamente desarrolló en su vida al procesar su cerebro la información que configuró su identidad *pre-enfermedad*? Sería la misma persona, pero su identidad habría variado. Asegurar la integridad de su memoria, de sus recuerdos y su privacidad mental son retos sumamente importantes. Los *neuro derechos* tienen por finalidad avanzar una respuesta jurídica preventiva a un inmenso conjunto de dilemas éticos con los que nos vamos a enfrentar más pronto que tarde, la prevención de técnicas invasivas inicialmente que ya se pueden apreciar hoy como contradictorias con los Derechos Humanos «clásicos» en muy diversas dimensiones y que supondrán retos espectaculares en la evolución de nuestra propia especie desde el momento en que conozcamos cómo modificar el contenido de nuestra mente —los correlatos neurobiológicos— sobre las bases de un conocimiento cada vez más detallado de la arquitectura neurobiológica de un órgano central en nuestra vida como es el cerebro, se aproxima.

Las neurotecnologías como precisan Müller y Rotter⁴⁹ nos abren un mundo nuevo que hay que configurar con ideas nuevas y herramientas novedosas, suponen una oportunidad y un reto que hay que afrontar sin miedo, pero bajo dos principios esenciales que deben operar simultáneamente: *el principio de precaución*⁵⁰ y *el principio de responsabilidad*. El recurso al principio de precaución sólo se produce en la hipótesis de riesgo potencial, aunque este riesgo no pueda demostrarse por completo, no pueda cuantificarse su amplitud o no puedan determinarse sus efectos debido a la insuficiencia o al carácter no concluyente de los datos científicos disponibles. La neurotecnología se está desarrollando con gran rapidez, posiblemente de manera exponencial. Pero los seres humanos, en nuestro proceso de adaptación a los acontecimientos y desafíos es, en cambio, *lineal*. Cuando los seres humanos *lineales* nos enfrentamos a un cambio *exponencial*, no podemos adaptarnos a ese cambio con facilidad y es lo que denominados un *cambio de paradigma* repleto de derivadas en multitud de dimensiones y ámbitos de la vida: social, económica, política, jurídica, emocional etc.

El principio de responsabilidad es claro, hay que responder de los daños que puedan causarse de las aplicaciones neurotecnológicas que se comercialicen bien

49. Müller, O y S Rotter, Neurotechnology: Current Developments and Ethical Issues, *Frontiers in Systems Neuroscience*, Diciembre 2017, Vol 11, article 93, pp. 1-4.

50. Aunque en el TFUE sólo se mencione explícitamente el principio de precaución en el terreno del medio ambiente, art. 191, su ámbito de aplicación es mucho más amplio. Este principio abarca los casos específicos en los que los datos científicos son insuficientes, no concluyentes o inciertos, pero en los que una evaluación científica objetiva preliminar hace sospechar que existen motivos razonables para temer que los efectos potencialmente peligrosos para el medio ambiente y la salud humana, animal o vegetal pudieran ser incompatibles con el alto nivel de protección elegido. «Sobre el recurso al principio de precaución», Comunicación de la Comisión, COM (2000) 1 final, Bruselas, 2, 2, 2000.

en los ámbitos médicos como en los de naturaleza lúdica que *deben reconducirse* a regulaciones de evaluación médica si queremos el máximo control sobre sus aplicaciones, en caso contrario, las aplicaciones lúdicas pueden suponer *una vía furtiva* de pérdida de información y datos personales extraordinariamente sensibles que entendemos inaceptable, al menos con el modelo técnico de protección de datos del que disponemos en la actualidad y que entendemos que no es el idóneo para el ciudadano pero si lo es para las organizaciones y empresas multinacionales que los emplean máxime cuando se puede deslocalizar tales datos agregados en un mundo global y ser tratados en localizaciones mundiales poco o nada respetuosas de los derechos humanos. La empresa 23andMe ha sufrido un hackeo el 9 de octubre de 2023, sustrayéndose millones de datos de su base de datos. 23andMe es una empresa genética encargada de analizar muestras de ADN de millones de pacientes, esta empresa se dedica precisamente a recibir las muestras de saliva de sus clientes para realizarles un genotipado con el objetivo de determinar qué genes se están expresando y cuáles están silenciados. Dependiendo de los genes que estén expresados en el sujeto se puede determinar que puede estar predispuesto a diferentes enfermedades, los datos genéticos están a la venta en la Dark Web con un precio que va de 1 a 10 dólares. Pensemos lo que supondría tener acceso o robar la experiencia vital de una persona codificada por un laboratorio extraída por una BCI. Es claro que las políticas de datos y seguridad no podrían indemnizar a una persona por una pérdida de estas características, si además posteriormente se hicieran públicas esas experiencias vitales de todo género, personales, ideológicas, sexuales pudiendo incluir imágenes. La *privacidad mental* exige nuevos desarrollos *técnicos* ya que los actuales son sumamente ineficientes. En ese sentido concreto se pronuncian los principios 6, 7, 8 y 10 del Código de Conducta global de la IA de 30 de octubre de 2023, denominado el proceso de IA de Hiroshima aprobado por el G7 si bien adolece que es un código voluntario de conducta internacional para las empresas. Creemos positiva y compartimos desde hace tiempo la idea expuesta entre otros por Acemoglu⁵¹ en relación con la creación de un *mercado de datos* en el que cada ciudadano tenga acceso a la información que cada empresa dispone sobre él y reciba una parte de los ingresos que generaran, es decir, la *patrimonializarían*⁵² de los datos personales ya que si el combustible del BigData y de la IA son los datos, quien controle los datos controlará en buena medida la IA, por ejemplo, obligando a *excluir* un dato propiedad de alguien a través del Derecho Penal si se vulnera su derecho de propiedad.

El empleo de las neurotecnologías aborda directamente los problemas que hemos considerado y muchos otros de análoga naturaleza en lo que Sadin⁵³ denomina la *condición antropológica* que entrelaza a un ritmo creciente organismos humanos y artificiales con problemas éticos fundamentales que el derecho no puede desconocer

51. D Acemoglu, *El impacto de la inteligencia artificial será una mezcla de la imprenta, la máquina de vapor y la bomba atómica*, en: <https://www.elmundo.es/papel/lideres/2023/10/16/652d5669fc6c8317598b45bf.html?fbclid=IwAR2HFvNmqqkjJhE-CjJec7EkbYOKBcTQWJsWh8MiYoHjQPKiPGqNjBJTOIIE> (Visualizado, octubre 2023).

52. <https://www.elnotario.es/opinion/opinion/743-patrimonializar-los-datos-de-caracter-personal-argumentos-para-un-debate-0-022018592825176746>

53. Sadin, E, *La humanidad aumentada*, Caja Negra, Buenos Aires, 2017, p. 152.

como precisan Lenca y Andorno⁵⁴ desde la perspectiva de tales neuroderechos. Aquí sólo podemos hacer un breve semblante de cómo los primeros intentos de regulación se están desarrollando. Los neuroderechos formarían parte de los derechos de cuarta generación y tienen que ver con bienes jurídicos a los que afecta la IA, la genética, la bioingeniería, así como todo el conjunto de neurociencias que inciden sobre el *libre desarrollo de la personalidad* (10.1 CE); la *integridad física y moral —mental de constitución ferenda—* (15 CE); *el no ser obligado a declarar sobre ideología, religión o creencias* (16.2 CE); o a *la intimidad personal* (18.1 CE) en el caso nuestro ordenamiento jurídico constitucional.

V. TECNOLOGÍAS ADICTIVAS. LA ACTUACIÓN DE LA IA SOBRE COLECTIVOS, LOS MENORES, LOS JÓVENES Y OTROS COLECTIVOS

El hecho de que China⁵⁵ haya desarrollado legislación para limitar el uso de juegos por parte de los niños a un máximo de *tres horas por semana* responde igualmente a la consciencia por parte de los poderes públicos de que estas herramientas generan *problemas adictivos subliminales* muy severos en el desarrollo de los menores con patologías graves para su normal desarrollo emocional e intelectual. De hecho, según establece la OMS esta adicción se ha contemplado en la reciente clasificación internacional de enfermedades mentales CIE.11 y, específicamente, se recoge como 6C51.0. Trastorno por uso de videojuegos, predominantemente en línea⁵⁶.

Como señala entre nosotros Echeburúa⁵⁷ conectarse a la Red nada más despertarse, al llegar a casa o justo antes de acostarse y con ello reducir el tiempo dedicado a las tareas cotidianas (comer, dormir, estudiar o charlar con la familia) son

54. Lenca M y R Andorno, *Towards new human rights in the age of neuroscience and neurotechnology*, *Life Sciences, Society and Policy*, 2017, pp. 5-13.

55. China prohíbe que los menores dediquen más de tres horas semanales a los juegos por internet. El anuncio se produce en medio de una creciente preocupación de las autoridades por la adicción a esta actividad, que han llegado a calificar de «opio espiritual» <https://elpais.com/tecnologia/2021-08-30/china-limita-a-tres-horas-semanales-la-practica-de-juegos-online-por-parte-de-los-menores.html> (visualizado 23 sept 2023).

56. *Descripción*: El trastorno por uso de videojuegos se caracteriza por un patrón de comportamiento de juego persistente o recurrente («juegos digitales» o «videojuegos»), que puede ser en línea (es decir, por internet) o fuera de línea, y que se manifiesta por: 1. deterioro en el control sobre el juego (por ejemplo, inicio, frecuencia, intensidad, duración, terminación, contexto); 2. incremento en la prioridad dada al juego al grado que se antepone a otros intereses y actividades de la vida diaria; y 3. continuación o incremento del juego a pesar de que tenga consecuencias negativas. El patrón de comportamiento del juego puede ser continuo o episódico y recurrente. El patrón de comportamiento del juego da como resultado una angustia marcada o un deterioro significativo en las áreas de funcionamiento personal, familiar, social, educativo, ocupacional u otras áreas importantes. El comportamiento del juego y otras características normalmente son evidentes durante un período de al menos 12 meses para que se asigne un diagnóstico, aunque la duración requerida puede acortarse si se cumplen todos los requisitos de diagnóstico y los síntomas son graves. <https://icd.who.int/browse11/l-m/es/#/http%3a%2f%2fid.who.int%2fid%2fentity%2f1448597234> (visualizado 09/2023).

57. Echeburúa, E, *Adictos a las nuevas tecnologías*, *Investigación y Ciencia*, (*Mente & Cerebro*), mayo-agosto, 2013, pp. 36-37.

algunos de los comportamientos usuales del adicto a las redes sociales o, en su caso, a las nuevas tecnologías o a los video juegos. Es decir, más que el número de horas, el factor determinante en la adicción es *el grado de interferencia negativa* que ejerce ese comportamiento en la vida cotidiana del afectado. Así, el teléfono inteligente —seguramente ya más inteligente que el usuario medio— crea dependencia en individuos menores como ha demostrado Haidt⁵⁸ que consideran el dispositivo indispensable para la vida y que no saben cuándo deben prescindir de él. Su atención a los mensajes que reciben es constante, por lo que con frecuencia desatienden otras actividades importantes, incluso *la comunicación cara a cara*, para contestar a los contactos virtuales o lo que se ha dado en llamar *comunicación asíncrona* por el *estrés* que produce a los jóvenes el miedo de las comunicaciones en tiempo real⁵⁹. Las consecuencias del uso abusivo de un teléfono inteligente suponen, asimismo, una diversidad de efectos negativos: existe una focalización atencional en torno al dispositivo y sus aplicaciones, se reduce la actividad física, y no se es capaz de diversificar el tiempo e interesarse por otras actividades o temas. El sujeto muestra ansiedad por las redes sociales y se produce un flujo de *transrealidad* que recuerda la experiencia adictiva de las drogas. Se crea un efecto bola de nieve, ya que los problemas se extienden a todas las parcelas personales (salud, familia, escuela y relaciones sociales). En definitiva *la dependencia* y la supeditación del estilo de vida al mantenimiento del hábito conforman el núcleo central de la adicción. Así, la dependencia a las redes sociales no se caracteriza tanto por el tipo de conducta implicada, sino por la forma de relación que el sujeto establece con ella. Todo esto puede desembocar en una especie de *analfabetismo relacional* y facilitar la construcción de *relaciones sociales ficticias, deficientes y defectuosas*.

El uso indiscriminado y sin control alguno por parte de los menores de las redes —como en el caso del juego ya considerado— puede generar graves problemas de salud mental en la juventud. Como ha desvelado en fechas recientes *The Wall Street Journal*⁶⁰ Facebook la empresa dueña de Instagram tiene un claro conocimiento de que Instagram es una aplicación *tóxica* para los adolescentes, el uso de la aplicación por millones de jóvenes en el mundo genera un importante problema de salud mental ya que una gran parte de ellos, pero singularmente en los menores de 22 años les genera adicción a la misma que la compañía minimiza de cara al público. Quienes sufren un daño más acusado son, singularmente, las jóvenes que han convertido en adicción su presencia en esta red social diseñada para captar y explotar la dependencia emocional de la misma a través de los círculos de amistades que también las emplean en esas edades. El diario citó estudios internos de Facebook durante los últimos tres años que examinaron cómo Instagram afecta a su base de usuarios jóvenes. Una presentación interna de Facebook señaló que entre los adolescentes que informaron

58. Haidt, J, *La generación ansiosa. Por qué las redes sociales están causando una epidemia de enfermedades mentales entre nuestros jóvenes*, Deusto, Barcelona, 2024.

59. No llames, manda un audio: los milenials ya no hablan por teléfono por estrés. En la era del postexto y del consumo frenético de información, la comunicación asíncrona (o sea, la conversación fragmentada) gana terreno. Llamar se percibe como algo casi invasivo. <https://elpais.com/ideas/2021-11-21/no-llames-manda-un-audio-los-milenials-ya-no-hablan-por-telefono-por-estres.html> (visualizado el 22-9-2023).

60. <https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739>

tener pensamientos suicidas, el 13% de los usuarios eran británicos y el 6% de los usuarios estadounidenses rastrearon el problema del suicidio hasta Instagram.

TikTok ganó 682 millones de nuevos usuarios el año pasado, cada uno de los cuales pasa un promedio de 50 minutos al día en la aplicación Koetsier⁶¹ pero qué la hace tan adictiva. Simplemente se está por el usuario *en un estado mental placentero* generando dopamina Haynes⁶² al visualizar y observar esas imágenes dejándose llevar. Es casi hipnótico, seguirás mirando y mirando. Cuando te desplazas de una imagen a otra a veces ves una foto o algo que es llamativo, atractivo y te llama la atención, *en ese momento* se obtiene ese pequeño incremento de dopamina en el cerebro en el centro de placer del cerebro el *núcleo accumbens*. Así se desea seguir desplazándose, navegando y visualizando imágenes por esas redes sin fin de imágenes. Sigues desplazándote porque, a veces se visualiza algo que es agradable pero *otras no*. Y esa *diferenciación*, muy similar a la que se produce en una máquina tragaperras de un Casino, es la clave como señalan James y colaboradores.⁶³ Plataformas como TikTok, incluidas Instagram, Snapchat y Facebook, han adoptado los mismos principios psicológicos que han hecho que los juegos de azar sean adictivos. En la demanda federal —Caso 4:23-cv-05448⁶⁴— presentada ante el distrito norte de California por 33 Fiscales Generales, los Estados alegan que los productos de Meta han perjudicado a los menores y han contribuido a una *grave crisis de salud mental en Estados Unidos* en base a los argumentos sumariamente aquí considerados. En la demanda federal multiestatal del martes 24 de octubre de 2023 participan California, Colorado, Connecticut, Delaware, Georgia, Hawái, Idaho, Illinois, Indiana, Kansas, Kentucky, Louisiana, Maine, Maryland, Michigan, Minnesota, Missouri, Nebraska, Nueva Jersey, Nueva York, Carolina del Norte, Dakota del Norte, Ohio, Oregón, Pensilvania, Rhode Island, Carolina del Sur, Dakota del Sur, Virginia, Washington, Virginia Occidental y Wisconsin. Hay que señalar que en los estudios internos de la multinacional Meta ésta era consciente de que sus redes sociales tenían y tienen capacidad adictiva sobre los usuarios como argumenta Horwitz⁶⁵.

En términos psicológicos se denomina «*refuerzo aleatorio*» y significa que a veces se gana y a veces se pierde. Los estudios sobre *el refuerzo aleatorio* provienen de las investigaciones con ratas en procesos de alimentación y recompensa. Y así es como están —conscientemente— diseñadas estas plataformas, son exactamente como una máquina de juego de azar. Somos conscientes de que existe la adicción al juego y como patología extrema la ludopatía. Pero no consideramos a menudo cómo nuestros «*teléfonos inteligentes*» las plataformas y estas aplicaciones de gran difusión tienen estas mismas cualidades adictivas integradas en lo que nos ofrecen *por diseño*, es

61. <https://www.forbes.com/sites/johnkoetsier/2020/01/18/digital-crack-cocaine-the-science-behind-tiktoks-success/?sh=40498c0678be> (visualizado el 10 de septiembre de 2023).

62. T Haynes, *Dopamine, Smartphone & You: A Battle for Your Time*, Harv. Univ. SITN Blog, May 1, 2018, <https://archive.ph/9MMhY>

63. J RJE, O'Malley C and Tunney RJ, Why are Some Games More Addictive than Others: The Effects of Timing and Payoff on Perseverance in a Slot Machine Game. *Front. Psychol.* 7:46. (2016) doi: 10.3389/fpsyg.2016.00046.

64. <https://ag.ny.gov/sites/default/files/court-filings/meta-multistate-complaint.pdf>

65. Horwitz J, *Código roto. Manipulación política, fake news, desinformación y salud pública*, Ariel, Barcelona, 2024.

decir, se ofrece de forma consciente a los usuarios tecnologías claramente adictivas. Esa es la razón que ha motivado la demanda.

El descubrimiento de esta técnica Morgan⁶⁶ vino dado por un conjunto de experimentos donde se estudiaba la respuesta de las ratas al condicionamiento. Mediante un condicionamiento específico se proporcionaba una recompensa a las ratas cada vez que ejecutaban una simple tarea. Gratificar es una buena forma de incentivar a alguien a que haga algo. Pero los experimentadores se dieron cuenta de que había *algo mejor* que el «refuerzo positivo». Era el «refuerzo aleatorio». Esto supone que, bajo refuerzo aleatorio, cuando la rata realizaba la tarea *unas veces obtenía recompensa y otras no*. Aunque pueda parecer contraintuitivo que gratificando a alguien siempre o en todas las ocasiones se conseguirá una finalidad que se desea que el sujeto realice, lo cierto es que es mucho más efectivo *recompensar sólo a veces*. En el caso de los seres humanos son las emociones las que son la materia prima con la juegan estos sistemas de condicionamiento electrónico. La emoción es lo realmente relevante y es la «*incertidumbre*» la que desencadena la sensación emocionante y adictiva de que al conseguir el objetivo se produce la *liberación de dopamina* Eimeren.⁶⁷ Si nuestro equipo de fútbol o de baloncesto gana todos los partidos pronto deja de interesarnos. La persona que siempre está encima de nosotros, atendiéndonos y dándonos todas las atenciones todo el tiempo, nos empieza a molestar. De hecho, siempre nos atrae aquello que parece más difícil de conseguir. Si siempre hay gratificación pronto alcanzamos un *nivel de saturación* y nos condicionamos de manera que la recompensa ya no nos produce el mismo nivel de bienestar y placer que la primera vez. Valoramos pues las cosas de acuerdo con lo que nos cuesta conseguirlas, por eso valoramos más lo que tenemos o podemos conseguir «*en algunas ocasiones, cuando ganamos*» que no lo que tenemos siempre.

Es posible que pensemos que somos mucho más sofisticados que las ratas, pero para este aspecto concreto de la conducta humana no es el caso, operamos como ellas y otros muchos animales. La *emoción* de los partidos en directo, el hecho de que nos apasionen los juegos de todo tipo de apuestas o azar, buscando premios que podemos *conseguir o no*, la cuestión de que todos hemos querido lo que no podemos alcanzar y no valoramos lo que siempre tenemos o ya hemos conseguido. Todos y cada uno de esos hechos tan humanos son la demostración empírica de cómo el *refuerzo aleatorio* es lo más adictivo y es, precisamente, lo que ofrecen a las personas estas plataformas, naturalmente el impacto en la juventud es cuantitativa y cualitativamente mucho más adictivo que en los adultos porque sus cerebros no están aún desarrollados. Pero entre los grupos que pueden ser objeto de manipulación también se encuentran colectivos como personas adultas afectadas por la ciber ludopatía electrónica, colectivos que deben ser objeto de protección.

VI. CONCLUSIONES

Kant señalaba que, *se mide la inteligencia del individuo por la cantidad de incertidumbre que es capaz de soportar*, casi cien años después F. Scott Fitzgerald acuñó una de las

66. Morgan, M. J, Effects of random reinforcement sequences, *Journal of the experimental analysis of behavior*, n.º 2, (septiembre), 22, 1974.

67. Eimeren, T V, Renunciar temporalmente a los dispositivos tecnológicos, el llamado «ayuno de dopamina», puede prevenir la adicción a estos aparatos, en *Mente & Cerebro*, noviembre-diciembre, n.º 111, 2021, pp. 66-67.

definiciones de la inteligencia más famosas de la historia: *la prueba de un intelecto de primer nivel es su capacidad de manejar dos ideas contrapuestas al mismo tiempo y aun así mantener la capacidad de funcionar*. En los entornos VUCA (volatilidad, incertidumbre, complejidad y ambigüedad) como es específicamente el de la IA existen ventajas realmente útiles para toda la sociedad y riesgos muy importantes que oscilan al mismo tiempo en un horizonte de indeterminación que se consolida paso a paso en dependencia de las decisiones sociales que sobre ellas adoptemos. La regulación de la IA que supone la Ley de IA de la Unión Europea es una cristalización temporal feliz de una regulación que favorece las ventajas sociales de la IA y minimiza sus riesgos, en ese sentido hay razones para considerar que es una buena regulación *inicial* singularmente en la prevención de riesgos inaceptables de manipulación mental que son los que hemos abordado en este capítulo, sin perjuicio de que deberemos permanecer atentos al desarrollo de la norma en su encaje con la realidad tecnológica objeto de regulación a menudo resistente a dejarse regular, pero esa es la vocación del Derecho y quizá su único mérito realmente consistente.

El resto de los sistemas de inteligencia artificial prohibidos o inaceptables en el Reglamento de inteligencia artificial

PERE SIMÓN CASTELLANO¹

Profesor Titular de Derecho Constitucional de la Universidad Internacional de la Rioja - UNIR

I. INTRODUCCIÓN

El artículo 5 del RIA prohíbe expresamente la introducción en el mercado, la puesta en servicio o la utilización de unos sistemas de IA y determinados usos y «prácticas». De hecho, el capítulo II del RIA lleva por título, precisamente, «prácticas de inteligencia artificial prohibidas». Buena parte del alcance de la prohibición de determinadas prácticas de IA ha sido objeto de estudio en los capítulos que han precedido *ut supra* al que ahora tiene el lector entre manos y el placer de redactar un servidor.

Otros estudios en esta obra sobre IA prohibida se centran en el reconocimiento mediante biometrías; en los sistemas de IA que se sirvan de técnicas subliminales que trasciendan la conciencia de una persona o de técnicas deliberadamente manipuladoras o engañosas con el objetivo o el efecto de alterar de manera sustancial el comportamiento de una persona o un grupo de personas, mermando de manera apreciable su capacidad para tomar una decisión informada y haciendo que una persona tome una decisión que de otro modo no habría tomado, de un modo que provoque, o sea probable que provoque, perjuicios considerables a esa persona, a otra persona o a un grupo de personas. También se ha examinado la prohibición de los sistemas que permiten explotar vulnerabilidades de una persona o un grupo específico de personas derivadas de su edad o discapacidad, o de una situación social o económica específica, con el objetivo o el efecto de alterar de manera sustancial el comportamiento de dicha persona o de una persona que pertenezca a dicho grupo de un modo que provoque, o sea razonablemente probable que provoque, perjuicios considerables a esa persona o a otra; en los sistemas de IA con el fin de evaluar o clasificar a personas físicas o a grupos de personas durante un período determinado

1. El presente trabajo se realiza en el marco del Proyecto «Derechos y garantías públicas frente a las decisiones automatizadas y el sesgo y discriminación algorítmicas» 2023-2025 (PID2022-136439OB-I00) financiado por MCIN/AEI/10.13039/501100011033/FEDER, UE.

de tiempo atendiendo a su comportamiento social o a características personales o de su personalidad conocidas, inferidas o predichas, de forma que la puntuación ciudadana resultante provoque una situación de trato perjudicial o desfavorable. También, cabe recordar que son analizadas las evaluaciones de riesgos de persona físicas y la comisión de infracciones penales en el estudio de Fernando Miró Llinares y Mario Santisteban Galarza, en el bloque temático de sistemas de IA de alto riesgo.

No obstante, el RIA prevé otras prácticas prohibidas o IA inaceptables que son objeto del presente estudio: prohibición de IA basada en reconocimiento facial con imágenes extraídas de Internet o televisión; sistemas de IA de reconocimiento de emociones; sistemas de categorización biométrica que clasifiquen individualmente a persona físicas en base a sus datos biométricos.

El tema es importante puesto que en la práctica y a nivel internacional ya se han empezado a implementar estos sistemas basados en IA, lo que ha generado polémica y a vueltas cierto rechazo. Más allá del uso tradicional para identificar individuos, los sistemas biométricos avanzados pueden ahora reconocer emociones, clasificar personas, detectar comportamientos y pensamientos, e incluso evaluar personalidades e incluye incluso los llamados polígrafos inteligentes. La implementación de estos sistemas en el control fronterizo no es una práctica reciente; ya hace más de una década se introdujeron en Arizona. Más recientemente, en Estados Unidos, el Agente Virtual Automatizado para la Evaluación de la Verdad en Tiempo Real (AVATAR) ha sido empleado para analizar tanto el comportamiento verbal como el no verbal de los viajeros, y también se ha probado en el aeropuerto de Bucarest, entre otros. La Comisión Europea, por su parte, financió el proyecto *Intelligent Portable Control System (iBorderCtrl)*², que utiliza herramientas para la detección de engaños y evaluaciones basadas en el riesgo, lo que ha provocado una notable reacción por parte de la sociedad civil, incluyendo una iniciativa ciudadana europea y la campaña *reclaimyourface.eu*.

Otro ejemplo lo encontramos en Brasil. El 7 mayo 2021 el Tribunal de Justicia de São Paulo prohibió a la concesionaria del Metro de Sao Paulo que utilizara el «Sistema Digital Interactivo de Puertas» (DID) con reconocimiento facial, el sistema infería emociones, género y edad de las personas para personalizar la publicidad³.

Grandes empresas y plataformas disponen también de sistemas de reconocimiento biométrico y facial no sólo para identificar personas sino para detectar emociones, estados de ánimo, etc. En junio de 2022, Microsoft anunció que retiraba sus sistemas *Azure Face*⁴, previamente había dejado de vender este tipo de tecnología a la policía de Estados Unidos. Meta-Facebook dispone desde 2017 de patentes de reconocimiento de emociones⁵. En noviembre de 2021 eliminó el polémico uso del reconocimiento facial⁶.

2. Sobre el uso en fronteras, véase Sánchez Monedero, J. y Dencik, L. «The Politics of Deceptive Borders: “Biomarkers of Deceit” and the Case of iBorderCtrl’». *Information, Communication & Society*, vol. 1, 2020. <https://doi.org/10.1080/136911>
3. <https://www.accessnow.org/sao-paulo-court-bans-facial-recognition-cameras-in-metro/>
4. <https://azure.microsoft.com/es-es/products/cognitive-services/face/>
5. <https://www.cbinsights.com/research/facebook-emotion-patents-analysis/>
6. <https://about.fb.com/news/2021/11/update-on-use-of-face-recognition/>

Con todo, procede advertir al lector que el RIA contiene normas específicas para la protección de las personas en relación con el tratamiento de datos personales que restringen el uso de sistemas de IA para la identificación biométrica remota con fines de aplicación de la ley, el uso de sistemas de IA para la realización de evaluaciones de riesgos de personas físicas con fines de aplicación de la ley y el uso de sistemas de IA de categorización biométrica con fines de aplicación de la ley. Es por ello por lo que el RIA, en lo que atañe a dichas normas específicas, encuentra su base y fundamento en el artículo 16 del TFUE. A la luz de dichas normas específicas y del recurso al artículo 16 del TFUE, debemos tener en cuenta en especial el criterio al respecto del Comité Europeo de Protección de Datos.

II. DEFINICIONES Y TERMINOLOGÍA: TECNOLOGÍAS DE CATEGORIZACIÓN BIOMÉTRICA, EMOCIONES, DATOS BIOMÉTRICOS Y RECONOCIMIENTO FACIAL

El reconocimiento facial automático o automatizado hay que incluirlo dentro de toda una serie de «técnicas biométricas» («identificación biométrica», «categorización biométrica», «detección de comportamientos», «reconocimiento de emociones», tratamiento de «datos biométricos», «elaboración de perfiles biométricos», etc.).

Se vienen definiendo como sistemas de identificación biométrica a los procesos automatizados utilizados para reconocer a un individuo a partir de medir, almacenar y comparar sus datos biométricos relativos a sus características físicas, fisiológicas o de comportamiento⁷. Estas características biométricas son universales —las tienen todos los humanos—, singulares y únicas e invariables a lo largo de la vida.

Los sistemas de procesamiento de datos biométricos se basan en recoger y procesar datos personales relativos a las características físicas, fisiológicas o conductuales de las personas físicas, entre las que cabe incluir, como se ha puesto de manifiesto recientemente, las características neuronales de estas, mediante dispositivos o sensores, creando plantillas biométricas (también denominadas firmas o patrones) que posibilitan la identificación, seguimiento o perfilado de dichas personas (esto es, «tratar», art. 4.2 del RGPD). El RGPD define en el artículo 4.14 datos biométricos como «datos personales obtenidos a partir de un tratamiento técnico específico, relativos a las características físicas, fisiológicas o conductuales de una persona física que permitan o confirmen la identificación única de dicha persona, como imágenes faciales o datos dactiloscópicos», y en la definición se establece que son datos biométricos todos aquellos que permitan la identificación o autenticación de una persona.

Sin embargo, los datos biométricos pueden permitir la autenticación, la identificación o la categorización de las personas físicas y el reconocimiento de las emociones de las personas físicas. Decimos «pueden» porque como se verá, la versión final del RIA no mantiene esa referencia a que los datos biométricos permiten la identificación o autenticación de una persona. Los datos biométricos cuentan con

7. Véase la recopilación de conceptos afines disponible en Gallego Rodríguez, P. «Los registros biométricos y su aplicación al proceso penal desde una perspectiva constitucional», en Calaza López, S. y Llorente Sánchez-Arjona, M. (dirs.), *Inteligencia artificial legal y administración de justicia*. Aranzadi, Cízur Menor, 2022, pp. 211-255, véase pp. 234 y ss.

la misma definición en el art. 3.13 de la Directiva (UE) 2016/680⁸, el artículo 4.14 del RGPD y, salvo la referencia a la identificación unívoca, también en el RIA, en el art. 3.34. Se trata de «aquellos datos personales obtenidos a partir de un tratamiento técnico específico, relativos a las características físicas, fisiológicas o conductuales de una persona física, como imágenes faciales o datos dactiloscópicos». En la última versión aprobada se ha eliminado del texto original del artículo 3.33 del RIA la mención a «que permitan o confirmen la identificación única de dicha persona». La razón que subyace a la exclusión de esa mención en el RIA, no es por la redundancia en la medida que sólo son datos personales aquellos que permiten identificar una persona, y con ello se infiere que no son datos biométricos aquellos que no permiten la identificación unívoca de una persona. La razón es que cabe excluir los sistemas de mera verificación biométrica, que comprende la autenticación, cuyo único propósito es confirmar que una persona física concreta es la persona que dice ser, así como la identidad de una persona física con la finalidad exclusiva de que tenga acceso a un servicio, desbloquee un dispositivo o tenga acceso de seguridad a un local. Esos sistemas no están prohibidos en la medida que el riesgo se sitúa por razones obvias dentro de un umbral aceptable o tolerable.

El considerando 15 del RIA ayuda a delimitar el contenido al establecer que «el concepto de “identificación biométrica” a que hace referencia el presente Reglamento debe definirse como el reconocimiento automatizado de características humanas de tipo físico, fisiológico o conductual, como la cara, el movimiento ocular, la forma del cuerpo, la voz, la entonación, el modo de andar, la postura, la frecuencia cardíaca, la presión arterial, el olor o las características de las pulsaciones de tecla, a fin de determinar la identidad de una persona comparando sus datos biométricos con los datos biométricos de personas almacenados en una base de datos de referencia, independientemente de que la persona haya dado o no su consentimiento. Quedan excluidos los sistemas de IA destinados a la verificación biométrica, que comprende la autenticación, cuyo único propósito es confirmar que una persona física concreta es la persona que dice ser, así como la identidad de una persona física con la finalidad exclusiva de que tenga acceso a un servicio, desbloquee un dispositivo o tenga acceso de seguridad a un local».

Resulta fundamental también atender a las dos definiciones de los artículos 3.35 y 3.36 del RIA, que establecen que la identificación biométrica debe ser considerada como «el reconocimiento automatizado de características humanas de tipo físico, fisiológico, conductual o psicológico para determinar la identidad de una persona física comparando sus datos biométricos con los datos biométricos de personas almacenados en una base de datos» y la verificación biométrica debe ser entendida como «la verificación automatizada y uno-a-uno, incluida la autenticación, de la identidad de las personas físicas mediante la comparación de sus datos biométricos con los datos biométricos facilitados previamente».

8. Directiva (UE) 2016/680 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por parte de las autoridades competentes para fines de prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de sanciones penales, y a la libre circulación de estos datos. En adelante, Directiva (UE) 2016/680.

Un dato biométrico contenido en un sistema se almacena en forma de una plantilla o patrón biométrico. Una plantilla biométrica es una forma de escritura de una característica biométrica humana, como un rostro o una huella dactilar, de manera que sea interpretable por una máquina de forma eficiente y eficaz para un propósito o propósitos determinados. La plantilla biométrica no está orientada a ser interpretada por una persona, como una fotografía, sino que está orientada a ser tratada en un proceso automatizado, es decir, ser eficiente y eficazmente interpretable por una máquina. Esta forma de almacenamiento permitiría singularizar a un individuo y ejecutar acciones de forma automática, perfilar o inferir información sobre un sujeto como actitudes o patrones de comportamiento, etc.

En el caso de operaciones de identificación o autenticación, para que una plantilla biométrica sea eficaz es necesario que las plantillas generadas a partir de dos individuos distintos sean claramente distinguibles. En ese caso, la plantilla actúa como un identificador único de la persona. El hecho de que, a partir de una plantilla biométrica, por ejemplo, de reconocimiento facial, no se pueda reconstruir el rostro original carece de relevancia, pues es un identificador único que lo singulariza unívocamente, al menos, en el marco de un tratamiento automatizado. De igual forma, a partir de únicamente del número de DNI no se puede reconstruir un nombre o un rostro. A ambos identificadores únicos, plantilla biométrica o número del DNI, se les puede asociar datos personales y atributos adicionales en un fichero. A diferencia de un número de DNI, la plantilla biométrica no es asignada a una persona, sino que se genera directamente de la observación de características físicas únicas e inalterables del propio individuo, sin necesidad de recurrir a documentos, otros dispositivos o bases de datos de terceros.

Hablamos así de identificadores a partir de características físicas, fisiológicas o de comportamiento. Se llega a distinguir entre identificadores «fuertes» especialmente utilizados con las tecnologías de primera generación especialmente destinadas a la identificación (huellas dactilares, ADN, estructura del iris, rostros, voz) e identificadores «débiles» que cada vez cobran más protagonismo (formas de andar, patrones de vasos sanguíneos, pulsaciones de teclas etc.). Con la nueva generación de tecnologías se va más allá de la finalidad de identificación y se habla de «biometría del comportamiento» para el perfilado, reconocimiento de emociones o categorización de personas. Es por ello que, frente a los datos biométricos ligados únicamente a la identificación, se propone el concepto más amplio inclusivo de «datos basados en la biometría»⁹.

9. Las enmiendas del Parlamento de junio de 2023 incorporar el concepto, en el artículo 3.33 bis, de «datos basados en la biometría», a los que definía como los datos personales resultantes del tratamiento técnico específico de señales o características físicas, fisiológicas o de comportamiento de una persona física, como las expresiones faciales, los movimientos, la frecuencia del pulso, la voz, las pulsaciones o la forma de andar, que pueden permitir o no la identificación única de una persona física. Tal definición se ha subsumido dentro del considerando 15 que se refiere a características humanas de tipo físico, fisiológico o conductual, como la cara, el movimiento ocular, la forma del cuerpo, la voz, la entonación, el modo de andar, la postura, la frecuencia cardíaca, la presión arterial, el olor o las características de las pulsaciones de tecla, a fin de determinar la identidad de una persona comparando sus datos biométricos con los datos biométricos de personas almacenados en una base de datos de referencia, independientemente de que la persona haya dado o no su consentimiento. Se excluyen de

La cuestión es importante porque sólo los datos biométricos para la identificación son datos especialmente protegidos bajo el régimen especial del artículo 9 RGPD o art. 10 Directiva (UE) 2016/680. El Comité Europeo de Protección de Datos sigue siendo claro en el sentido de que, si no se vinculan a la identificación, no son datos sensibles¹⁰. Lo que encaja con la definición incluida en el RIA que excluye los sistemas de mera verificación biométrica, que comprende la autenticación, cuyo único propósito es confirmar que una persona física concreta es la persona que dice ser, así como la identidad de una persona física con la finalidad exclusiva de que tenga acceso a un servicio, desbloquee un dispositivo o tenga acceso de seguridad a un local. Llama la atención que no estén entre las categorías especiales de datos los que revelan nuestras emociones, los pensamientos y las intenciones¹¹. También procede manejar el concepto de «inferencias biométricas» relativas a las conclusiones o resultados de los tratamientos permanentes o a largo plazo de estos datos basados en la biometría.

III. PRÁCTICAS INACEPTABLES: OBJETO Y CONTENIDO DE LA PROHIBICIÓN

El RIA prohíbe la introducción en el mercado, la puesta en servicio para este fin específico o el uso de sistemas de IA que creen o amplíen bases de datos de reconocimiento facial mediante la extracción no selectiva de imágenes faciales de internet o de circuitos cerrados de televisión; la introducción en el mercado, la puesta en servicio para este fin específico o el uso de sistemas de IA para inferir las emociones de una persona física en los lugares de trabajo y en los centros educativos, excepto cuando el sistema de IA esté destinado a ser instalado o introducido en el mercado por motivos médicos o de seguridad; la introducción en el mercado, la puesta en servicio para este fin específico o el uso de sistemas de categorización biométrica que clasifiquen individualmente a las personas físicas sobre la base de sus datos biométricos para deducir o inferir su raza, opiniones políticas, afiliación sindical, convicciones religiosas o filosóficas, vida sexual u orientación sexual; esta prohibición no incluye el etiquetado o filtrado de conjuntos de datos biométricos adquiridos lícitamente, como imágenes, basado en datos biométricos ni la categorización de datos biométricos en el ámbito de la garantía del cumplimiento del Derecho; el uso de sistemas de identificación biométrica remota «en tiempo real» en espacios de acceso público con fines de garantía del cumplimiento del Derecho.

estos los sistemas de mera verificación biométrica, que comprende la autenticación, cuyo único propósito es confirmar que una persona física concreta es la persona que dice ser, así como la identidad de una persona física con la finalidad exclusiva de que tenga acceso a un servicio, desbloquee un dispositivo o tenga acceso de seguridad a un local.

10. Véanse las *Guidelines 05/2022 on the use of facial recognition technology in the area of law enforcement, Version 1.0*, 12 mayo, https://edpb.europa.eu/our-work-tools/documents/public-consultations/2022/guidelines-052022-use-facial-recognition_en
11. Sobre la expansión o insuficiencia de la categoría de datos especialmente protegidos, como también ocurre con los datos de salud, véase Cotino Hueso, L. «El alcance e interacción del régimen jurídico de los datos personales y big data relacionado con salud y la investigación biomédica», *Revista de derecho y genoma humano: genética, biotecnología y medicina avanzada*, n.º 52, pp. 57-96, n.º 52 enero-junio 2020.

1. RECONOCIMIENTO FÁCIL VÍA «SCRAPING» O EXTRACCIÓN NO SELECTIVA DE IMÁGENES FACIALES EN INTERNET Y CIRCUITO CERRADO DE TELEVISIÓN

En el considerando 43 del RIA se indica que «la introducción el mercado, la puesta en servicio para este fin concreto o la utilización de sistemas de IA que creen o amplíen bases de datos de reconocimiento facial mediante la extracción no selectiva de imágenes faciales a partir de internet o de imágenes de circuito cerrado de televisión deben estar prohibidas, pues estas prácticas agravan el sentimiento de vigilancia masiva y pueden dar lugar a graves violaciones de los derechos fundamentales, incluido el derecho a la intimidad».

El reconocimiento facial mediante técnicas de *scraping*, también conocido como extracción no selectiva de imágenes faciales en Internet y circuitos cerrados de televisión, es una tecnología avanzada que permite identificar o verificar la identidad de una persona analizando sus características faciales. Este proceso se lleva a cabo mediante algoritmos de IA que extraen y procesan datos faciales a partir de imágenes disponibles en línea o capturadas por cámaras de vigilancia.

El *scraping* facial implica la recolección masiva de imágenes faciales de personas a partir de diversas fuentes digitales, como redes sociales, sitios web públicos y cámaras de vigilancia. A diferencia de la recopilación de datos autorizada, el *scraping* no se realiza con el consentimiento explícito de los individuos cuyas imágenes son recolectadas. Los algoritmos de IA analizan estas imágenes para crear perfiles faciales detallados, que luego pueden ser utilizados para diversas aplicaciones, desde la seguridad hasta la publicidad personalizada.

Este método de extracción no selectiva plantea serias preocupaciones éticas y legales, dado que invade la privacidad de las personas y puede dar lugar a usos indebidos de la información recopilada. La capacidad de identificar a individuos sin su consentimiento y en contextos no autorizados supone un riesgo significativo para los derechos fundamentales de las personas y el libre desarrollo de su personalidad, además de para su libertad personal.

La prohibición de esta técnica en el RIA encuentra su fundamentación nuclear, más allá de la privacidad, en la necesidad de prevenir abusos, manipulación, discriminación o ciertos usos vinculados con la ciberdelincuencia. La recopilación y uso no autorizado de datos faciales representa, ante todo, una violación directa de la privacidad. El RIA trata de cerrar el círculo normativo y complementar con esta prohibición un uso difícilmente compatible con el RGPD, tutelando proteger a los ciudadanos europeos de la intrusión en su vida privada y garantizando que sus datos biométricos no sean utilizados sin su conocimiento y consentimiento explícito. Sin embargo, como decía más arriba, la fundamentación esencial para esta prohibición va más allá de la privacidad y la protección de datos, y es que las imágenes y los datos biométricos de reconocimiento facial pueden ser utilizados para fines discriminatorios, como la vigilancia selectiva de ciertos grupos étnicos o la discriminación en el acceso a servicios y oportunidades. Al prohibir estas prácticas, el RIA intenta prevenir el abuso de la tecnología y asegurar un trato justo y equitativo para todos los individuos. Además, la recolección masiva de datos faciales sin control puede llevar a la creación de bases de datos susceptibles de ser hackeadas o explotadas. Al establecer restricciones claras, el RIA busca garantizar

que las tecnologías de IA se desarrollen y utilicen de manera segura y confiable, fortaleciendo la confianza del público en estas innovaciones. Hay que tener en cuenta, también, que el uso indiscriminado del *scraping* facial contraviene principios éticos fundamentales, como el respeto por la dignidad humana y la autonomía personal, y en la práctica tendría efectos nocivos con un potencial efecto lesivo sobre el libre desarrollo de la personalidad.

Consideramos que la prohibición del reconocimiento facial vía *scraping* en el RIA es una medida crucial para proteger los derechos y libertades de los individuos, prevenir el abuso de la tecnología y fomentar un desarrollo ético y seguro de las innovaciones en IA. Al establecer estos límites, la UE se posiciona a la vanguardia en la regulación de tecnologías emergentes, asegurando un equilibrio entre progreso tecnológico y respeto por los derechos fundamentales.

2. LA PROHIBICIÓN DEL USO DE SISTEMAS DE INTELIGENCIA ARTIFICIAL PARA INFERIR EMOCIONES EN EL REGLAMENTO

El RIA establece una prohibición clara y estricta sobre el uso de sistemas de IA para inferir las emociones de las personas en lugares de trabajo y centros educativos, salvo en situaciones donde dichos sistemas se empleen con fines médicos o de seguridad. Esta disposición, incluida en el considerando 44 del RIA, subraya la necesidad de proteger los derechos fundamentales y la privacidad de los individuos frente a prácticas que podrían llevar a una vigilancia masiva y a violaciones graves de su intimidad.

Más concretamente, el legislador europeo nos dice que «existe una gran preocupación respecto a la base científica de los sistemas de IA que procuran detectar o deducir las emociones, especialmente porque la expresión de las emociones varía de forma considerable entre culturas y situaciones, e incluso en una misma persona. Algunas de las deficiencias principales de estos sistemas son la fiabilidad limitada, la falta de especificidad y la limitada posibilidad de generalizar. Por consiguiente, los sistemas de IA que detectan o deducen las emociones o las intenciones de las personas físicas a partir de sus datos biométricos pueden tener resultados discriminatorios y pueden invadir los derechos y las libertades de las personas afectadas. Teniendo en cuenta el desequilibrio de poder en el contexto laboral o educativo, unido al carácter intrusivo de estos sistemas, dichos sistemas podrían dar lugar a un trato perjudicial o desfavorable de determinadas personas físicas o colectivos enteros. Por tanto, debe prohibirse la introducción en el mercado, la puesta en servicio y el uso de sistemas de IA destinados a ser utilizados para detectar el estado emocional de las personas en situaciones relacionadas con el lugar de trabajo y el ámbito educativo. Dicha prohibición no debe aplicarse a los sistemas de IA introducidos en el mercado estrictamente con fines médicos o de seguridad, como los sistemas destinados a un uso terapéutico».

La prohibición se centra en los sistemas de IA que analizan y determinan las emociones humanas a través de diversas señales biométricas y de comportamiento. Estos sistemas son capaces de interpretar expresiones faciales, tonos de voz, gestos y otras características físicas y comportamentales para inferir estados emocionales, como felicidad, tristeza, estrés, y otros. La tecnología que permite esta inferencia se basa en el análisis de grandes volúmenes de datos, a menudo recopilados sin el consentimiento explícito de los individuos afectados.

Los ámbitos de aplicación de la prohibición incluyen lugares o puestos de trabajo y centros educativos. La prohibición en el ámbito laboral responde a la preocupación de que los empleadores puedan utilizar estas tecnologías para monitorizar y evaluar continuamente el estado emocional de los empleados. Tal práctica podría llevar a un entorno de trabajo opresivo y a la discriminación basada en las emociones percibidas, afectando la salud mental y la privacidad de los trabajadores. Por lo que se refiere a las instituciones educativas, el uso de IA para inferir emociones plantea riesgos significativos para la privacidad y el bienestar de los estudiantes. El monitoreo emocional constante podría interferir con el desarrollo natural y la autonomía de los estudiantes, además de crear un ambiente de vigilancia que contraviene los principios de educación libre y abierta.

Lo anterior limita considerablemente el ámbito de aplicación de la prohibición, y de hecho, en el propio RIA, el legislador anima a su uso en otros ámbitos, en especial, cuando se refiere a los fines médicos o de seguridad. Por ejemplo, en un entorno médico, estos sistemas pueden ser cruciales para diagnosticar y tratar condiciones de salud mental o emocional. Del mismo modo, en situaciones de seguridad, la capacidad de inferir emociones puede ser vital para prevenir amenazas inmediatas, como identificar comportamientos potencialmente peligrosos en tiempo real. No es un *numerus clausus* de excepciones, pero sí es un listado cerrado, en cambio, de prohibiciones, que sólo proyecta efectos sobre centros educativos y entorno laboral.

La prohibición se fundamenta en la protección de los derechos fundamentales de los ciudadanos europeos, en particular el derecho a la privacidad y a la integridad psicológica. La inferencia de emociones puede llevar a una forma de vigilancia intrusiva que no solo invade la privacidad de las personas, sino que también puede manipular su comportamiento y decisiones. Además, la precisión de estas tecnologías no está garantizada y puede variar significativamente, lo que aumenta el riesgo de errores y malinterpretaciones.

La introducción de sistemas de IA capaces de inferir emociones también suscita preocupaciones éticas significativas. La posibilidad de que estas tecnologías se utilicen para influir en el comportamiento humano de manera subrepticia plantea cuestiones sobre el libre albedrío y la manipulación. En contextos laborales y educativos, donde las personas ya están en posiciones de menor poder en comparación con empleadores o autoridades educativas, esta tecnología podría exacerbar las desigualdades de poder y llevar a abusos. La prohibición de estos sistemas de IA en el RIA refleja un compromiso firme con la protección de la dignidad humana y la prevención de un entorno de vigilancia que podría socavar los derechos y libertades fundamentales.

3. EL USO DE SISTEMAS DE INTELIGENCIA ARTIFICIAL DE CATEGORIZACIÓN BIOMÉTRICA QUE CLASIFIQUEN INDIVIDUALMENTE A LAS PERSONAS FÍSICAS SOBRE LA BASE DE SUS DATOS BIOMÉTRICOS PARA DEDUCIR O INFERIR SU RAZA, OPINIONES POLÍTICAS, AFILIACIÓN SINDICAL, CONVICCIONES RELIGIOSAS O FILOSÓFICAS, VIDA SEXUAL U ORIENTACIÓN SEXUAL

El RIA también establece la prohibición de utilizar sistemas de IA para la categorización biométrica que clasifiquen individualmente a personas físicas en base a sus datos biométricos. Más específicamente, el RIA prohíbe los sistemas que

inferían o deduzcan la raza, opiniones políticas, afiliación sindical, convicciones religiosas o filosóficas, vida sexual u orientación sexual de una persona física a partir de sus datos biométricos, como la cara o las huellas dactilares.

El objeto o alcance material de la prohibición indicada se concreta en el uso por parte de sistemas de IA de datos biométricos para realizar categorizaciones personales. Los datos biométricos incluyen características físicas, biológicas y de comportamiento que son únicas para cada individuo, como las huellas dactilares, los rasgos faciales, el iris del ojo, entre otros. Estos sistemas pueden analizar estos datos para intentar inferir información sensible sobre las personas, como su raza, creencias políticas, afiliación sindical, creencias religiosas, vida sexual u orientación sexual. Sin embargo, esas inferencias o predicciones de los sistemas de IA pueden ser inexactos o predisponer a los usuarios hacia una toma de decisiones sesgadas.

Por ello, la prohibición de estos sistemas en el RIA se justifica por varias razones clave, centradas principalmente en la protección de los derechos fundamentales y la privacidad de las personas. La deducción o inferencia o predicción basada en información sensible a partir de datos biométricos puede llevar a violaciones significativas de los derechos fundamentales. La privacidad, la igualdad y la no discriminación son pilares esenciales de la legislación de la Unión Europea. Utilizar IA para inferir características tan íntimas puede resultar en un uso indebido de la información, en acciones discriminatorias u otras formas de injusticia social. Además, el uso de sistemas de IA para la categorización biométrica puede contribuir a un estado de vigilancia masiva, donde los individuos son constantemente monitorizados y analizados. Esto no solo infringe la privacidad, sino que también crea un ambiente de desconfianza y miedo, socavando la libertad individual y el derecho a la autonomía personal. Asimismo, los sistemas de IA no son infalibles y pueden cometer errores en la interpretación de los datos biométricos. Si el resultado de la predicción es hacer una «propuesta» comercial o una propuesta de la que no deriva una acción que produzca efectos sobre los derechos de los interesados, entonces no proyectaría efectos negativos o perniciosos. Pero la inferencia incorrecta¹² de características sensibles puede llevar a decisiones equivocadas y a la discriminación. Además, los algoritmos pueden perpetuar y amplificar sesgos existentes en los datos, resultando en tratamientos injustos para ciertas personas o grupos. Por último, la clasificación de individuos en función de sus características biométricas y la inferencia de información personal y sensible plantea también cuestiones éticas. La dignidad humana se ve comprometida cuando las personas son reducidas a un conjunto de datos biométricos y categorizadas sin su consentimiento.

La prohibición que prevé el RIA tampoco es absoluta, puesto que establece o fija notables excepciones. No se aplica al etiquetado, filtrado o categorización lícita

12. Pueden derivarse incluso del diseño de la solución tecnológica. Repárese en que la correlación alude a la correspondencia entre dos o más acciones o fenómenos; sin embargo, la correlación no implique causalidad. El output que arrojan muchos sistemas de IA, entonces, demuestran una correlación, pero no necesariamente un efecto o una consecuencia, propiamente dichos. Y este puede ser el origen de muchos equívocos o problemas por lo que se refiere a los resultados que arrojan los sistemas de IA. Véase Lehr, D. y Ohm, P. «Playing with the Data: What Legal Scholars Should Learn About Machine Learning», en *University of California Davis Law Review*, vol. 51, 2017, p. 671.

de conjuntos de datos biométricos adquiridos conforme a la legislación nacional o de la Unión Europea. Por ejemplo, la clasificación de imágenes en función del color del pelo o de los ojos puede ser permitida en el contexto de la aplicación de la ley, donde dichos datos se utilizan para finalidades legítimas y específicas que no comprometen la privacidad ni los derechos fundamentales de los individuos. En el RIA más concretamente se indica que «esta prohibición no abarca el etiquetado o filtrado de conjuntos de datos biométricos adquiridos legalmente, como imágenes, basado en datos biométricos ni la categorización de datos biométricos en el ámbito de la aplicación de la ley».

IV. LA COEXISTENCIA DE LA REGULACIÓN PREVISTA EN EL REGLAMENTO CON LA NORMATIVA DE PROTECCIÓN DE DATOS: UNA REGULACIÓN SUPERPUESTA Y, ¿COMPATIBLE?

Se ha señalado por algunos autores que el elemento esencial definitorio de los sistemas biométricos —y de reconocimiento facial— pasa por su finalidad identificación de personas¹³, y ello tiene y mucho que ver con la definición ya reproducida que contiene el RGPD de dato biométrico. Sin embargo, tal extremo choca con el RIA en la medida que este prevé tan sólo que los datos biométricos pueden permitir la autenticación, la identificación o la categorización de las personas físicas y el reconocimiento de las emociones de las personas físicas. En la última versión aprobada se ha eliminado del texto original del artículo 3.33 del RIA la mención a «que permitan o confirmen la identificación única de dicha persona». Se excluyen los sistemas de mera verificación biométrica, que comprende la autenticación, cuyo único propósito es confirmar que una persona física concreta es la persona que dice ser, así como la identidad de una persona física con la finalidad exclusiva de que tenga acceso a un servicio, desbloquee un dispositivo o tenga acceso de seguridad a un local.

Hay que distinguir además y allí donde sea posible entre la identificación biométrica (*uno a varios*) y la autenticación o verificación biométrica (*uno a uno*). Así lo han hecho las autoridades de protección de datos. Según lo han definido el Grupo del Artículo 29 ya en 2012 o la AEPD¹⁴, la identificación biométrica es el proceso de comparar los datos biométricos de un individuo, adquiridos en el momento de la identificación, con una serie de plantillas biométricas almacenadas en una base de datos de personas generalmente identificadas, esto es, un proceso de búsqueda de correspondencias uno-a-varios. Por el contrario, tradicionalmente no se ha considerado como tratamiento de datos sensibles bajo el especial régimen del artículo 9 del RGPD la «verificación» o «autenticación» biométrica «uno a uno», esto es, el proceso de confirmar que un individuo es quien dice ser en el que se comparan los datos únicamente con la identidad que se quiere contrastar. Ese sería el caso, por ejemplo, de la identificación con el teléfono móvil personal a través de la huella o el

13. Cotino Hueso, L. «Reconocimiento facial automatizado y sistemas de identificación biométrica bajo la regulación superpuesta de inteligencia artificial y protección de datos», en Balaguer Callejón, F. y Cotino Hueso, L. (coords.), *Derecho público de la inteligencia artificial*, Madrid, Marcial Pons, 2023, pp. 347.402.

14. Así, el Grupo del Artículo 29 desde el Dictamen 3/2012 sobre la evolución de las tecnologías biométricas o la AEPD en sus diversas guías.

rostro que tiene registrado. Así las cosas, el redactado actual de la propuesta de RIA tendría un encaje perfecto con la interpretación que se ha realizado de la normativa de protección de datos. No son datos sensibles y no aplica el régimen especial del artículo 9 de RGPD, puesto que es una verificación o autenticación «uno a uno», que se enmarca dentro de la exclusión prevista en el RIA cuando nos dice que no se entenderán afectados por la prohibición los sistemas de IA de mera verificación biométrica, que comprenden también la autenticación, cuyo único propósito es confirmar que una persona física concreta es la persona que dice ser, así como la identidad de una persona física con la finalidad exclusiva de que tenga acceso a un servicio, desbloquee un dispositivo o tenga acceso de seguridad a un edificio, sede o local.

En estas verificaciones uno a uno, y en sentido parecido, la AEPD ha indicado que «si bien también realiza un tratamiento de datos personales, no llega a procesar la información contra una base de datos previa que permita o confirme la identificación de personas de uno a varios»¹⁵. Por ello, que el artículo 9. 1º RGPD incluya entre los especialmente protegidos los «datos biométricos dirigidos a identificar de manera unívoca a una persona física» no es incompatible con la exclusión de la prohibición prevista en el RIA. Y eso es así en la medida que se ha venido interpretando como que el uso de datos biométricos sólo quedaba bajo el régimen de datos especialmente protegidos en las identificaciones uno a varios y no en las identificaciones uno a uno.

El Comité Europeo de Protección de Datos ha realizado un cambio interpretativo con sus directrices de 2022¹⁶. La AEPD ha exteriorizado este cambio de criterio en la Guía de Tratamientos de control de presencia mediante sistemas biométricos de 23 de noviembre de 2023, en la que indica principalmente que cabe partir de que tanto para identificación como para autenticación, estamos ante un tratamiento de alto riesgo que incluye categorías especiales de datos. Ello obliga a una habilitación normativa específica. Este cambio de criterio está vinculado al reconocimiento y tratamiento de datos biométricos para fines de control de jornada horaria en el ámbito laboral y no en un sector específico como pueda ser la identificación en el sector público y en particular en la actuación de la justicia. No obstante, sí que se tiene en cuenta¹⁷ y para

15. Véase procedimiento sancionador AEPD PS/00120/2021, p. 28.

16. Véase también la posición del Comité Europeo de Protección de Datos, que inicialmente parecía cambiar de criterio, en las *Guidelines 05/2022 on the use of facial recognition technology in the area of law enforcement*, Version 1.0, 12 mayo, https://edpb.europa.eu/our-work-tools/documents/public-consultations/2022/guidelines-052022-use-facial-recognition_en Después de distinguir autenticación-verificación de la identificación, afirma que ambos casos «constituyen un tratamiento de datos personales, y más concretamente un tratamiento de categorías especiales de datos personales». Véase también el trabajo de Santisteban Galarza, M. «Reconocimiento facial y protección de datos: una respuesta provisional a un problema pendiente». *Revista de Derecho de la UNED (RDUNED)*, n.º 28, 2022, pp. 499-526. <https://doi.org/10.5944/rduned.28.2021.32887>

17. pp. 19-20. Una limitación al uso del consentimiento se hace expresa en las mismas Directrices con relación al recurso al consentimiento en el marco de las AA.PP.:

16. El considerando 43 indica claramente que no es probable que las autoridades públicas puedan basarse en el consentimiento para realizar el tratamiento de datos ya que cuando el responsable del tratamiento es una autoridad pública, siempre hay un claro desequilibrio de poder en la relación entre el responsable del tratamiento y

un caso como el presente se incluyen afirmaciones totalmente aplicables: «se tendrán que atender los requisitos de necesidad, presente en todos ellos, además de los de reserva de ley en las letras c) y d) y también en el caso de la letra f) la superación del análisis de prevalencia entre los intereses legítimos del responsable y los intereses o los derechos y libertades fundamentales del interesado que requieran la protección de datos personales, en particular cuando el interesado sea un niño». Y que «con carácter previo a cualquier decisión de implantación de un sistema de control de presencia a través de sistemas biométricos, se realice una gestión del riesgo (art. 24.1 RGPD) y desde el diseño y por defecto (art. 25 RGPD) se apliquen las medidas técnicas y organizativas apropiadas a fin de garantizar y poder demostrar que el tratamiento es conforme con el RGPD. En particular, en caso de alto riesgo, deberá superar favorablemente una Evaluación de Impacto para la Protección de Datos (EIPD) que incluya y también supere el triple juicio de idoneidad, necesidad y proporcionalidad estricta establecido en el art. 35.7.b y también previsto por la doctrina del Tribunal Constitucional».

Se concluye también que «en el caso de que el sistema biométrico se implemente con técnicas de inteligencia artificial, para poder incluirlos en un tratamiento se deberán tener en cuenta las prohibiciones, limitaciones y exigencias establecidas en la normativa de inteligencia artificial», aportando las «medidas mínimas por defecto», que incluyen «informar a los sujetos de los datos sobre el tratamiento biométrico; implementar en el sistema biométrico la posibilidad de revocar el vínculo de identidad entre la plantilla biométrica y la persona física; implementar medios técnicos para asegurarse la imposibilidad de utilizar las plantillas para cualquier otro propósito; utilizar cifrado para proteger la confidencialidad, disponibilidad e integridad de la plantilla biométrica; utilizar formatos de datos o tecnologías específicas que imposibiliten la interconexión de bases de datos biométricos y la divulgación de datos no comprobada; suprimir los datos biométricos cuando no se vinculen a la finalidad que motivó su tratamiento; implementar la protección de datos desde el diseño; realizar previamente al inicio del tratamiento una Evaluación de Impacto para la Protección de Datos».

A juicio de quien suscribe estas líneas la guía de la AEPD supone un freno u óbice insalvable para que las empresas puedan apoyarse en determinados sistemas para cumplir con la obligación legal de mantener un control de jornada horaria, que cabe recordar es un derecho de los trabajadores y una obligación empresarial. En este contexto, las garantías desproporcionadas que se exigen en la guía no están en consonancia ni con el estado de la técnica, ni con la interpretación del CEPD y otras autoridades de control extranjeras, como la CNIL, y tampoco son susceptibles en ningún caso de permitir que se supere el juicio de necesidad, dada la interpretación severa y restrictiva realizada ante la AEPD. Y todo ello teniendo en cuenta que el riesgo de identificar con huella dactilar a un trabajador en la gran mayoría de empresas, con un sistema de identificación uno a uno, y atendiendo a matrices y métodos científicos

el interesado. Queda también claro en la mayoría de los casos que el interesado no dispondrá de alternativas realistas para aceptar el tratamiento (las condiciones de tratamiento) de dicho responsable. El CEPD considera que hay otras bases jurídicas que son, en principio, más adecuadas para el tratamiento de datos por las autoridades públicas.

para la valoración del riesgo, arrojan un escenario de riesgo inherente leve en cuanto a severidad o impacto y casi imposible en cuanto a probabilidad. Es en este escenario en el que encaja que se haya expulsado o suprimido en la última versión del RIA la referencia a la «identificación unívoca» de la persona física, y que opte por un texto literal que no deje margen a interpretaciones apocalípticas, como la de la AEPD, que actúen como freno a la innovación tecnológica en ámbitos en los que el riesgo inherente es cercano a cero y, en cualquier caso, se encuentra dentro del umbral de aquello tolerable o aceptable.

LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL DE
ALTO RIESGO: DELIMITACIÓN Y ANÁLISIS DE
ALGUNOS ÁMBITOS

Alcance y delimitación de los sistemas de alto riesgo en el Reglamento de inteligencia artificial

LORENZO COTINO HUESO

Catedrático de Derecho Constitucional de la Universitat de València. Valgrai

I. LA CONDICIÓN DE SISTEMA DE ALTO RIESGO ES ESENCIAL PARA EL REGLAMENTO

El RIA se aplica a los sistemas de IA,¹ pero lo cierto es que la mayor parte de la regulación y la imposición de obligaciones previstas giran en torno a que el sistema de IA sea considerado de alto riesgo (en adelante, SAR). La consideración de SAR ha pasado a ser el elemento clave del RIA, mencionada hasta 470 veces a lo largo del texto y forma parte del propio «ámbito de aplicación» del artículo 2, apartados 2º y 12º.

La definición de «alto riesgo» ya generó posiciones muy divergentes en el proceso previo a la propuesta de RIA.² Las nociones que se barajaron en el Libro Blanco sobre

1. cotino@uv.es. OdiseIA. El presente estudio es resultado de investigación de los siguientes proyectos: MICINN Proyecto «Derechos y garantías públicas frente a las decisiones automatizadas y el sesgo y discriminación algorítmicas» 2023-2025 (PID2022-136439OB-I00) financiado por MCIN/AEI/10.13039/501100011033/; «La regulación de la transformación digital ...» Generalitat Valenciana «Algorithmic law» (Prometeo/2021/009, 2021-24); «Algorithmic Decisions and the Law: Opening the Black Box» (TED2021-131472A-I00) y «Transición digital de las Administraciones públicas e inteligencia artificial» (TED2021-132191B-I00) del Plan de Recuperación, Transformación y Resiliencia. Estancia Generalitat Valenciana CIAEST/2022/1., Grupo de Investigación en Derecho Público y TIC Universidad Católica de Colombia; Estancia Generalitat Valenciana CIAEST/2022/1, Convenio de Derechos Digitales-SEDIA Ámbito 5 (2023/C046/00228673) y Ámbito 6. (2023/C046/00229475).
2. Cabe tener en cuenta Comisión Europea, Renda. A. (project leader), *Study to Support an Impact Assessment of Regulatory Requirements for Artificial Intelligence in Europe. Final Report (D5)*, abril 2021. <https://op.europa.eu/es/publication-detail/-/publication/55538b70-a638-11eb-9585-01aa75ed71a1>

En pp. 112 y ss. puede seguirse que en el proceso de consulta previo a la propuesta de reglamento, la definición de «alto riesgo» fue el punto más crucial. Un 18% de los encuestados (74 de 408), consideraba que esta definición de «alto riesgo» era poco clara o requería mejoras significativas. La clasificación binaria de riesgos como altos o bajos se había percibido como excesivamente simplificada, lo que llevó a varios in-

IA de la UE de 2020 y en la propuesta del Parlamento Europeo del mismo año se han concretado significativamente. Desde la Comisión se ha afirmado —no sé si con mucho fundamento— que un tercio de los sistemas de IA públicos serán SAR,³ mientras que sólo un 10% de los privados lo serían.⁴

La idea básica del RIA es que se requieren normas comunes para los SAR a fin de garantizar la salud, la seguridad y los derechos fundamentales (Cons. 7). De este modo, se garantizará que la información de salida de estos sistemas que se utilicen en la Unión no entrañe riesgos inaceptables para estos intereses públicos (Cons. 46).⁵

Obviamente, se es consciente de que este sistema de cumplimiento es una carga importante, por lo que «La clasificación de un sistema de IA como SAR debe limitarse a aquellos sistemas de IA que tengan un efecto perjudicial importante en la salud, la seguridad y los derechos fundamentales de las personas de la Unión, y dicha limitación reduce al mínimo cualquier posible restricción del comercio internacional» (Cons. 46). El Considerando 48 llega a mencionar veintiún derechos fundamentales que pueden estar afectados por los SAR, además de los derechos específicos de los menores.⁶

teresados a proponer la introducción de niveles adicionales de riesgo. Algunos argumentaban que la definición actual era demasiado amplia, mientras que otros creían que era demasiado restrictiva. Entre las propuestas alternativas, al menos seis documentos de posición habían abogado por el enfoque gradual del GDEC, que introducía cinco niveles de riesgo para un análisis más matizado. Otros sugerían la adopción de matrices de riesgo, que combinaban la intensidad del daño potencial con el nivel de implicación o control humano en las decisiones de inteligencia artificial (IA). La probabilidad de daño también se había mencionado repetidamente como un criterio esencial a considerar. Además, numerosos documentos de posición criticaban el enfoque en dos fases propuesto para determinar la IA de «alto riesgo». Al menos 19 de estos documentos lo consideraban inadecuado, y al menos cinco se oponían a un enfoque sectorial. Otras sugerencias y críticas variaban ampliamente. Una propuesta notable para mejorar la evaluación de riesgos era considerar a todos los sujetos afectados por la aplicación de la IA, destacando la importancia de tener en cuenta tanto los riesgos colectivos como los individuales, ya que las aplicaciones de IA podían implicar riesgos para la sociedad en su conjunto, incluyendo la democracia, el medio ambiente y los derechos humanos. La necesidad de clarificar la definición de «alto riesgo» había sido una preocupación compartida por todas las partes interesadas.

3. JRC, Tangi, L. y otros: *AI Watch European landscape on the use of Artificial Intelligence by the Public Sector*, JRC Science For Policy Report, Unión Europea. 2022, p. 58.
4. Comisión Europea, *Study to Support...* cit. p. 143.
5. Como se señala en el Considerando 7 «Conviene establecer normas comunes para los sistemas de IA de alto riesgo al objeto de garantizar un nivel elevado y coherente de protección de los intereses públicos en lo que respecta a la salud, la seguridad y los derechos fundamentales». Es por ello que «La introducción en el mercado de la Unión, la puesta en servicio o la utilización de sistemas de IA de alto riesgo debe supeditarse al cumplimiento por su parte de determinados requisitos obligatorios, los cuales deben garantizar que los sistemas de IA de alto riesgo disponibles en la Unión o cuya información de salida se utilice en la Unión no entrañen riesgos inaceptables para intereses públicos importantes de la UE, reconocidos y protegidos por el Derecho de la Unión.» (Considerando 46).
6. Así, se mencionan: dignidad, vida privada y familiar, protección de datos, libertad de expresión y de información, la libertad de reunión y de asociación, la no discriminación, el derecho a la educación, la protección de los consumidores, los derechos de los trabajadores, los derechos de las personas discapacitadas, la igualdad entre hombres

Como se va a exponer, el artículo 6 contiene las «Reglas de clasificación de los sistemas de IA de alto riesgo» y para ello sigue un sistema dual. De un lado, la consideración de SAR queda relacionada con los productos cubiertos por determinada legislación de armonización de la Unión enumerada o como componentes de seguridad de estos o producto independiente. Del otro lado, y si se cumplen unos requisitos generales, son SAR los sistemas de IA vinculados con finalidades y usos enumerados en el anexo III.

Buena parte de las obligaciones del RIA son relativas a los SAR y se exponen en la presente obra. Baste en todo caso recordar que se da un período de adaptación específico. En general, las obligaciones de los SAR del anexo III deben cumplirse a los 24 meses de la publicación y los del Anexo I a los 36 meses (art. 113). No obstante, el sector público contará con un privilegiado período máximo de seis años (art. 111.2º). En cualquier caso, «Se anima a los proveedores de sistemas de IA de alto riesgo a que empiecen a cumplir, de forma voluntaria, las obligaciones pertinentes del presente Reglamento ya durante el período transitorio.» (Consid. 178).

II. UNA ADVERTENCIA: LA REGULACIÓN DE SISTEMAS COMO DE ALTO RIESGO POR EL REGLAMENTO NO IMPLICA SU HABILITACIÓN LEGAL

Si ser considerado un SAR conlleva muchas consecuencias en el RIA, es muy importante recordar que la regulación de un SAR por el RIA no implica dotarle de cobertura legal. Como advierte expresamente el Considerando 63 del RIA, «El hecho de que un sistema de IA sea clasificado como un sistema de IA de alto riesgo en virtud del presente Reglamento no debe interpretarse como indicador de que su uso sea legal con arreglo a otros actos del Derecho de la Unión o del Derecho nacional compatible con el Derecho de la Unión [...] No debe entenderse que el presente Reglamento constituye un fundamento jurídico [...] salvo que el presente Reglamento disponga específicamente otra cosa.» (Considerando 63). Esta regla general es clara y considero que debe aplicarse a todos los SAR, siendo especialmente relevante respecto de los del Anexo III. Ello es así pese a que, en una clara falta de técnica legislativa, sólo en tres de los ocho apartados del Anexo III se añade la expresión «en la medida en que su uso esté permitido por el Derecho de la Unión o nacional aplicable». Así sucede en el caso de los sistemas de IA biométricos (1º), respecto de la garantía del cumplimiento del Derecho (6º) y el uso para migración, asilo y gestión del control fronterizo (7º). Sin embargo, no se añade esta necesidad de ley de regulación específica respecto del ámbito de Justicia (8º a).

En cualquier caso, hay que partir de que el RIA no sirve como norma legal que legitime un tratamiento de datos, una restricción de derechos fundamentales o que colme una exigencia de legalidad penal, sancionadora o procesal. Seguirá siendo necesaria una ley que habilite la existencia de un concreto SAR de los regulados con carácter general en el RIA.

y mujeres, los derechos de propiedad intelectual, el derecho a la tutela judicial efectiva y a un juez imparcial, los derechos de la defensa y la presunción de inocencia, y el derecho a una buena administración. Además, los derechos específicos de menores y la salud y la seguridad de las personas y la protección del medio ambiente.

Del lado contrario, y como excepción, el RIA expresamente implica una regulación que dota de base legal de legitimación en el artículo 10. 5º relativo a la posibilidad de utilizar «excepcionalmente» categorías especiales de datos «para garantizar la detección y corrección de los sesgos», bajo unos requisitos bastante precisos.

III. LOS SISTEMAS INTELIGENCIA ARTIFICIAL EN PRODUCTOS PELIGROSOS DEL ANEXO I

El primer conjunto de SAR se refiere a sistemas IA en productos que presentan ciertos niveles de peligrosidad, por lo que están sometidos al régimen de evaluación de conformidad de la UE, evaluación que debe realizarse por parte de terceros. La cuestión tiene cierta complejidad. Así, se trata de productos asociados por determinada legislación de armonización de la Unión enumerada en el Anexo I (anteriormente Anexo II en el proceso de elaboración del reglamento) o componentes de seguridad de estos productos. De acuerdo con el artículo 6.1, «un sistema de IA se considerará de alto riesgo cuando reúna las dos condiciones», es decir, que «esté destinado a ser utilizado como componente de seguridad [...] o que *el propio sistema de IA* sea uno de dichos productos» (a) y que «deba someterse a una evaluación de la conformidad realizada por un organismo independiente para su introducción en el mercado o puesta en servicio» (b).

1. EL SISTEMA INTELIGENCIA ARTIFICIAL COMO COMPONENTE DE SEGURIDAD O PRODUCTO DE PRODUCTOS SOMETIDOS A UNA «EVALUACIÓN DE CONFORMIDAD» POR UN TERCERO

El objetivo del RIA es centrarse en las aplicaciones de la IA que «pueden tener un efecto adverso para la salud y la seguridad de las personas» (Cons. 46). Para ello, se ha recurrido al concepto de «componente de seguridad», ya consolidado en el marco jurídico de las máquinas.⁷ No obstante, la idea de «componente de seguridad» se ajusta en el RIA para que sea más general, pasando a hablar de «componentes digitales»⁸ que pueden proyectarse respecto de la IA en todos los sectores, ya que cumplen una función de seguridad para el producto o sistema, o cuyo fallo o defecto de funcionamiento pone en peligro la salud y la seguridad de las personas o los bienes (artículo 3.14 del RIA).⁹

7. Reglamento (UE) 2023/1230 del Parlamento Europeo y del Consejo, de 14 de junio de 2023, relativo a las máquinas. Se puede seguir la definición en el artículo 3. 3º «“componente de seguridad”: un componente físico o digital, incluido el software, de un producto incluido en el ámbito de aplicación del presente Reglamento que esté diseñado o destinado a desempeñar una función de seguridad y que se introduzca en el mercado por separado, cuyo fallo o funcionamiento defectuoso ponga en peligro la seguridad de las personas, pero que no sea necesario para que dicho producto funcione o cuyos componentes normales puedan ser sustituidos para que dicho producto funcione».
8. Así, el Considerando 47 habla de «los riesgos de seguridad que pueda generar un producto en su conjunto debido a sus componentes digitales, entre los que pueden figurar los sistemas de IA».
9. En concreto, según el artículo 3. 14º RIA «componente de seguridad»: un componente de un producto o un sistema que cumple una función de seguridad para dicho producto o sistema, o cuyo fallo o defecto de funcionamiento pone en peligro la salud

Así, el sistema IA puede haberse introducido en el mercado integrado o no en un producto como componente digital del mismo, ya sea propiamente de seguridad o que pueda generar peligro, mencionando casos de robots o el sector sanitario (Cons. 46). Se incluyen supuestos de sistemas de IA como software o programas informáticos que se comercializan de forma independiente al producto. También el RIA se aplicaría al software independiente, que se considera un producto en sí mismo. El ejemplo más evidente sería el software independiente de dispositivos médicos, regulado por el Reglamento 745/2017 sobre productos sanitarios. Un sistema IA independiente será SAR según «la gravedad del posible perjuicio como la probabilidad de que se produzca», y la Comisión los tendrá en cuenta en su actualización de la lista de SAR según el avance tecnológico (Cons. 52). Debe tenerse en cuenta asimismo que en diversos sectores se ha preferido que el RIA actúe indirectamente, obligando a adaptar su normativa específica. Así, cuando se incorpore la IA como componente de seguridad de productos o sistemas en aviación civil, vehículos agrícolas o forestales, equipos marinos, sistema ferroviario, vehículos de motor y sus remolques. En estos casos, habrá de tenerse en cuenta las obligaciones del RIA para los SAR al adoptar actos delegados o de ejecución pertinentes, adaptados a las particularidades de dichos sectores. La idea es no «interferir con los mecanismos y las autoridades de gobernanza, evaluación de la conformidad y control del cumplimiento vigentes» (Consid. 49). Así, para ilustrar de qué se trata, la homologación de vehículos se rige por el Reglamento de la UE 2018/858, modificado por el RIA (art. 107), de modo que el RIA ha de introducirse en los actos delegados de este reglamento.¹⁰

Otro elemento básico para entender este grupo de sistemas de IA del Anexo I es que esté sujeto a una evaluación de conformidad por parte de un tercero. En este punto, cabe recordar que, respecto a determinados productos, su introducción en el mercado o la puesta en servicio solo pueden tener lugar cuando el producto cumple con toda la legislación de armonización de la Unión aplicable. Todo ello en el contexto del llamado «nuevo marco legislativo» (Cons. 46). El Anexo I es el que relaciona todos estos productos.

Para que los sistemas de IA de productos sometidos a esta legislación armonizada o componentes de seguridad de dichos productos indicados en el Anexo I se consideren SAR, la legislación armonizada de tales productos debe contemplar que la evaluación de conformidad sea realizada por un «organismo independiente de evaluación de la conformidad de acuerdo con dicha legislación de armonización de la Unión» (Cons. 50). Cabe recordar que, en la variada legislación y dentro de cada una, según el tipo de producto que se regula, en algunos casos esta evaluación de conformidad es una autoevaluación, es decir, se basa en un control interno por parte del propio proveedor.¹¹ Sin embargo, en los productos que se consideran más peligrosos, ha de intervenir un tercero independiente.

y la seguridad de las personas o los bienes;. Si se compara la definición con la del Reglamento de máquinas, la función de seguridad pasa a ser aquí una alternativa y no el elemento definitorio esencial.

10. Sobre el tema, VDA, *Position. Artificial Intelligence Act*, German Association on the Automotive Industry, Berlin, julio, 2023, <https://www.vda.de/en/news/publications/publication/artificial-intelligence-act>
11. Cabe remitir al apartado específico en esta obra de Adrián Palma. Para obtener más información sobre las evaluaciones de conformidad <https://single-market-economy>.

Así pues, para que un sistema de IA sea considerado SAR, no basta con que el sistema de IA sea un producto o componente de seguridad de un producto de la legislación de armonización del Anexo I; además, esa legislación debe contemplar que la evaluación de conformidad de éste será realizada por un tercero. Esto se justifica porque la evaluación de conformidad por parte de terceros se considera un indicio de que el producto en cuestión puede tener un impacto negativo en la seguridad o salud de las personas y, por ello, a efectos del RIA, debe considerarse SAR. Por tanto, puede haber sistemas de IA que sean productos o componentes de seguridad de productos previstos en la legislación de armonización indicados por el Anexo I, pero estos no se consideren SAR porque dicha legislación no contempla la evaluación de conformidad por parte de terceros. Como recuerda el Considerando 50: «Esos productos son, en concreto, máquinas, juguetes, ascensores, equipo y sistemas de protección para uso en atmósferas potencialmente explosivas, equipos radioeléctricos, equipos a presión, equipos de embarcaciones de recreo, instalaciones de transporte por cable, aparatos que queman combustibles gaseosos, productos sanitarios y productos sanitarios para diagnóstico in vitro».

Asimismo, un sistema IA puede ser SAR para el RIA, pero el producto del que es componente de seguridad o producto en sí mismo puede no ser de SAR en el concreto ámbito normativo que se le aplique a ese producto (Cons. 51). Este es el caso de los productos sanitarios «que prevén que un organismo independiente realice una evaluación de la conformidad de los productos de riesgo medio y alto» (Cons. 51).¹² Por lo que, si no están evaluados como SAR por ese organismo, no serían SAR a los efectos del RIA.

Si el sistema no es en sí mismo ni producto ni componente de seguridad, sino «sistemas de IA independientes», habrá que ver su finalidad y riesgo. Será la Comisión la que determinará si se considera SAR a través de actos delegados (Cons. 52).¹³ Cabe señalar que la cuestión adquiere extraordinaria complejidad, como analiza Palma en esta obra, ya que habrá que seguir de modo particular

-
- ec.europa.eu/single-market/ce-marking/manufacturers_en
12. 51 Que un sistema de IA se clasifique como de alto riesgo en virtud del presente Reglamento no significa necesariamente que el producto del que sea componente de seguridad, o el propio sistema de IA como producto, se considere de «alto riesgo» conforme a los criterios establecidos en la correspondiente legislación de armonización de la Unión que se aplique al producto. Tal es el caso, en particular, de los Reglamentos (IA) 2017/745 y (IA) 2017/746, que prevén que un organismo independiente realice una evaluación de la conformidad de los productos de riesgo medio y alto.
 13. «En cuanto a los sistemas de IA independientes, es decir, aquellos sistemas de IA de alto riesgo que no son componentes de seguridad o que no son productos en sí mismos, deben clasificarse como de alto riesgo si, a la luz de su finalidad prevista, presentan un alto riesgo de ser perjudiciales para la salud y la seguridad o los derechos fundamentales de las personas, teniendo en cuenta tanto la gravedad del posible perjuicio como la probabilidad de que se produzca, y se utilizan en varios ámbitos predefinidos especificados en el presente Reglamento. Para identificar dichos sistemas, se emplean la misma metodología y los mismos criterios previstos para la posible modificación futura de la lista de sistemas de IA de alto riesgo, que la Comisión debe estar facultada para adoptar, mediante actos delegados, a fin de tener en cuenta el rápido ritmo del desarrollo tecnológico, así como los posibles cambios en el uso de los sistemas de IA.»

cada Reglamento que regula estos productos, con especial atención al Reglamento 2023/1230 relativo a las máquinas o los referidos Reglamentos 745/2017 y 746/2017 sobre los productos sanitarios. Es más, estas normas habrán de adaptarse a la posterior aprobación del RIA, que actualmente, como es obvio, no tienen en cuenta.

2. EL PROVEEDOR DEBE CONOCER RAZONABLEMENTE SI SU PRODUCTO PUEDE SER DEL ANEXO I

Quien desarrolle un sistema de IA específicamente para ser incorporado en productos o como componentes de seguridad debe conocer su sector, la naturaleza habitual de los usuarios del sistema de IA y el régimen jurídico que se aplica a estos productos. Es decir, es natural que el productor de estos productos o el proveedor de servicios de IA conozca si el sistema queda sujeto a la normativa específica de armonización de la Unión referida en el Anexo I. A partir de ello, podrá valorar si se cumplen las condiciones del artículo 6 para que se considere que es un SAR. Así, deberá tener en cuenta si la legislación armonizada del producto o componente de seguridad del producto contempla la evaluación de conformidad por parte de terceros. Si es así, el proveedor de servicios tendrá que desarrollar el sistema de IA como de alto riesgo cumpliendo todas las exigencias.

El proveedor de un sistema de IA que no está específicamente desarrollado para ser incorporado en este tipo de productos también debe valorar si razonable y potencialmente su sistema de IA puede por sí solo ser considerado un producto cubierto por la legislación de armonización del Anexo I. En términos del RIA, debe valorar cuál es la «finalidad prevista» (art. 3. 1º. 12º).¹⁴ Si es así, en términos similares a lo que sucede respecto de la posibilidad de usar un sistema para fines del Anexo III, debe conocer si la concreta legislación exige una evaluación de conformidad por terceros o lo regula específicamente como SAR. Si es el caso, se tratará de un SAR con todas las consecuencias que ello conlleva.

Asimismo, un proveedor debe valorar si razonable y potencialmente su sistema de IA puede integrarse en un producto o ser utilizado como componente de seguridad de un producto sometido a la legislación armonizada del Anexo I. Si puede prever razonablemente que sí es así, y también en razón de su estrategia de puesta a disposición en el mercado, el proveedor deberá considerar que su sistema de IA es un SAR a los efectos del cumplimiento normativo de las obligaciones del RIA. Esto será imprescindible para poder introducirlo en el mercado o ponerlo a disposición como tal. En todo caso, de cara a los potenciales clientes o usuarios del sistema que desarrolle, deberá diseñar un marco jurídico en el que se determine si el sistema de IA puede o no ser incorporado a productos o como componente de seguridad. De este modo, su entidad habrá cumplido con las exigencias del RIA y el usuario habrá de cumplir con las suyas.

14. «12) “finalidad prevista”: el uso para el que un proveedor concibe un sistema de IA, incluidos el contexto y las condiciones de uso concretos, según la información facilitada por el proveedor en las instrucciones de uso, los materiales y las declaraciones de promoción y venta, y la documentación técnica;».

IV. LOS SISTEMAS DE ALTO RIESGO QUE PERSIGUEN LOS FINES DEL ANEXO III

1. SISTEMAS QUE TENGAN UNA INFLUENCIA SUSTANCIAL EN LA TOMA DE DECISIONES PARA LAS FINALIDADES DEL ANEXO III

En principio, «se considerarán de alto riesgo los sistemas de IA contemplados en el anexo III» (art. 6. 2º RIA). El Anexo III sobre «Sistemas de IA de alto riesgo a que se refiere el artículo 6, apartado 2» agrupa en ocho apartados los tipos de SAR: 1. Biometría; 2. Infraestructuras críticas; 3. Educación y formación profesional; 4. Empleo, gestión de los trabajadores y acceso al autoempleo; 5. Acceso a servicios privados esenciales y a servicios y prestaciones públicos esenciales y disfrute de estos servicios y prestaciones; 6. Aplicación de la ley; 7. Migración, asilo y gestión del control fronterizo; 8. Administración de justicia y procesos democráticos. En estos ocho apartados del Anexo III se contienen en veinticinco letras con otros tantos tipos de sistemas de IA en razón de su finalidad. Se utiliza en veinticinco ocasiones la fórmula de «Sistemas de IA destinados a ser utilizados...».¹⁵

Así pues, la «finalidad prevista» (art. 3. 12 RIA)¹⁶ para la que está concebido el sistema IA y «la capacidad de un sistema de IA para alcanzar su finalidad prevista», esto es, su «funcionamiento» (art. 3. 18 RIA),¹⁷ es el elemento determinante del anexo III. Por ello, y como se concretará, si el sistema de IA desarrollado tiene por finalidad prevista una de las del anexo III y la capacidad para lograrlo, se ha de presumir que sí que es SAR y que sólo excepcionalmente no lo será.

Ahora bien, es importante destacar cómo el RIA ha cambiado en su redacción respecto a este punto hasta su versión final. Inicialmente, había un automatismo: el sistema IA era SAR si estaba destinado a las finalidades descritas en el anexo III. Más tarde, en la versión del Consejo de la UE de diciembre de 2022, se añadió la excepción de que sería SAR «salvo que la información de salida del sistema sea *meramente accesoria* respecto de la acción o decisión pertinente que deba adoptarse» (art. 6. 3º RIA Consejo de la UE de diciembre 2022). Sin embargo, en la versión final, ya no se da ese automatismo, sino que además de las finalidades del anexo III, es preciso que se cumplan unos requisitos para que se considere de alto riesgo: el sistema IA con las

15. En puridad, en 24 ocasiones la fórmula se utiliza para describir la finalidad que define el alto riesgo. No obstante, la primera ocasión (Anexo III 1. a) se trata de un conjunto «1. Biometría, en la medida en que su uso esté permitido por el Derecho de la Unión o nacional aplicable». Este conjunto incluye tres supuestos. Las letras b y c sí que siguen la fórmula de «Sistemas de IA destinados a ser utilizados», sin embargo y como excepción a todo el anexo, en la letra a) la fórmula «los sistemas de IA destinados a ser utilizados» se utiliza como excepción respecto de los sistemas de identificación biométrica remota, esto es, «Quedan excluidos los sistemas de IA destinados a ser utilizados con fines de verificación biométrica cuya única finalidad sea confirmar que una persona física concreta es la persona que afirma ser».

16. «12) “finalidad prevista”: el uso para el que un proveedor concibe un sistema de IA, incluidos el contexto y las condiciones de uso concretos, según la información facilitada por el proveedor en las instrucciones de uso, los materiales y las declaraciones de promoción y venta, y la documentación técnica».

17. «18) “funcionamiento de un sistema de IA”: la capacidad de un sistema de IA para alcanzar su finalidad prevista».

finalidades del Anexo III ha de generar efectivamente un riesgo y, sobre todo, «influir sustancialmente en el resultado de la toma de decisiones» (art. 6. 3º RIA).

Si se me permite, es *sustancial* este cambio por el que el sistema de IA ha de influir *sustancialmente* en la decisión que se adopte. De hecho, posiblemente al lector avezado le llevará *automáticamente* a pensar en la regulación de las decisiones *automatizadas* del artículo 22 RGPD¹⁸ o del artículo 9. 1º del Convenio 108 del Consejo de Europa en su versión de 2018. Cabe recordar que las especiales garantías que brinda el artículo 22 RGPD se dan respecto de «una decisión basada *únicamente* en el tratamiento automatizado», mientras que será SAR del anexo III siempre que influya sustancialmente en la decisión que se adopte.

En el ámbito de protección de datos, la cuestión ha sido importante ya que, en principio, las decisiones que no son únicamente automatizadas no gozan de las garantías del artículo 22 RGPD. No obstante, la interpretación por las autoridades y los jueces ha ido en una dirección claramente garantista de proteger también las decisiones aparentemente humanas pero basadas sustancialmente en el sistema automatizado. Para el Grupo del Artículo 29 (CEPD), sí que está en el ámbito del artículo 22 RGPD «si alguien aplica de forma rutinaria perfiles generados automáticamente a personas sin que ello [la revisión humana] tenga influencia real alguna en el resultado, esto seguiría siendo una decisión basada únicamente en el tratamiento automatizado». Para que no rija este derecho, la intervención humana ha de ser «significativa, en vez de ser únicamente un gesto simbólico» y llevada a cabo por «persona autorizada y competente».¹⁹

Especialmente destacable es la sentencia del STJUE del 7 de diciembre de 2023,²⁰ la primera en abordar centralmente el artículo 22 RGPD. La misma supone un «corrimiento del velo» de las decisiones «únicamente» automatizadas. Así, las garantías de este precepto también se darán si los resultados del sistema automatizado, el perfilado o la ponderación automatizada de datos (o con inteligencia artificial) se conectan materialmente con la decisión finalmente adoptada por quien tiene que adoptarla respecto del afectado por dicha decisión, a pesar de que pueda darse una mediación o intervención humana. En esta línea, ya se iban dando algunos

18. He analizado en particular este precepto en «Derechos y garantías ante el uso público y privado de inteligencia artificial, robótica y big data», en Bauzá, M. (dir.), *El Derecho de las TIC en Iberoamérica*, Obra Colectiva de FIADI (Federación Iberoamericana de Asociaciones de Derecho e Informática), La Ley — Thompson-Reuters, Montevideo, 2019, pp. 917-952, acceso en <http://links.uv.es/BmO8AU7> En cualquier caso, por todos, Palma Ortigosa, A., *Decisiones automatizadas y protección de datos personales. Especial atención a los sistemas de inteligencia artificial*, Dykinson, 2022 y Roig I Batalla, A., *Las garantías frente a las decisiones automatizadas del Reglamento general de Protección de Datos a la gobernanza algorítmica*, J.M. Bosch, Barcelona, 2021.

19. G29-UE, *Directrices sobre decisiones ...* cit. p. 23.

20. El primer estudio sobre la misma Cotino Hueso, L. «La primera sentencia del Tribunal de Justicia de la Unión Europea sobre decisiones automatizadas y sus implicaciones para la protección de datos y el Reglamento de Inteligencia artificial», *Diario La Ley*, enero de 2024. <https://ir.uv.es/V14YNLI> Acceso a la sentencia en <https://curia.europa.eu/juris/document/document.jsf?text=&docid=280426&pageIndex=0&doclang=ES&mode=lst&dir=&occ=first&part=1&cid=10472490>
<https://eur-lex.europa.eu/legal-content/es/TXT/?uri=CELEX:62021CJ0634>

pasos por las autoridades.²¹ También las normas más recientes van reconociendo las garantías a las decisiones parcial o semi-automatizadas.²² Debe tenerse en cuenta que las decisiones automatizadas —o con IA— en muchas ocasiones se integran en una cadena o ecosistema de acciones en los que sí que hay intervención humana. Así pues, la existencia de dicha intervención humana no debe excluir *automáticamente* la aplicación de las particulares garantías que para las decisiones automatizadas confiere el artículo 22. Pues bien, y por lo que aquí interesa, el RIA deja atrás la necesidad de que la decisión que se adopte sea únicamente automatizada y expresamente delimita una serie de criterios para considerar si el sistema IA influye sustancialmente en la decisión.

Los criterios para considerar que el sistema inteligencia artificial influye sustancialmente en la decisión

La premisa es que el sistema «plantea un riesgo importante de causar un perjuicio a la salud, la seguridad o los derechos fundamentales de las personas físicas», pero ese riesgo debe darse, «en particular», por «influir sustancialmente en el resultado de la toma de decisiones» (art. 6.3 RIA). Así pues, no se considera SAR «un sistema de IA que no afecta al fondo, ni por consiguiente al resultado, de la toma de decisiones, ya sea humana o automatizada» (Cons. 53). Como se detalla a continuación, no serán SAR los sistemas utilizados para los fines del anexo III, pero con una tarea de «procedimiento limitado», o para «mejorar [...] una actividad humana previamente» realizada; cuando el sistema «no esté destinado a sustituir [...] ni a influir» en la toma de decisiones, sino a «detectar patrones» o, finalmente, cuando el uso del sistema IA es claramente accesorio.

Sin perjuicio de las incorporaciones a este respecto del Consejo de la UE en diciembre de 2022 y, especialmente, del Parlamento en junio de 2023, ha sido en la versión finalmente acordada donde se ha especificado que esta influencia sustancial no se dará «cuando se cumplan una o varias de las condiciones siguientes»:

a) «Que el sistema de IA tenga por objeto llevar a cabo una tarea de procedimiento limitada». A este respecto, se concreta que no sería SAR «un sistema de IA que transforme datos no estructurados en datos estructurados, un sistema de IA que

21. La autoridad de protección de datos de Portugal consideró como completamente automatizado un proceso que incluía intervención humana, dado que la persona encargada de supervisar los resultados del algoritmo no contaba con directrices ni criterios definidos para su interpretación. *Comissão Nacional de Proteção de Dados. Deliberação 622/2021*. Apartado 55. Resolución disponible en: <https://www.cnpd.pt/decisoes/deliberacoes/>

22. Así, el artículo 20 de la Ley orgánica de protección de datos personales de Ecuador (Registro Oficial Suplemento 459 de 26-may.-2021) extiende el derecho a «una decisión basada única o parcialmente en valoraciones automatizadas.» En Canadá cabe destacar la *Directive on Automated Decision-Making* que desde su primera versión de 2018 que define a un sistema de decisiones automatizado (automated decision system) como «Cualquier tecnología que ayude o reemplace el juicio de los tomadores de decisiones humanos.» (Apéndice A — Definiciones), por lo que en modo alguno las garantías de esta Directiva se limitan a las decisiones totalmente automatizadas.

En el caso de España, por ejemplo, la reciente Carta de Derechos Digitales contempla la necesidad de realizar una evaluación de impacto sobre los derechos digitales cuando se diseñan algoritmos para la toma de decisiones automatizadas o semi-automatizadas (apartado XVIII.7^o).

clasifique en categorías los documentos recibidos o un sistema de IA que se utilice para detectar duplicados entre un gran número de aplicaciones. La naturaleza de estas tareas es tan restringida y limitada que solo presentan riesgos limitados» (Cons. 52). Habrá que determinar casuísticamente cuándo la tarea es de «procedimiento limitado».

b) «Que el sistema de IA tenga por objeto mejorar el resultado de una actividad humana previamente realizada». En este punto, el Considerando 52 precisa que «el sistema de IA solo añade un nivel adicional a la actividad humana, entañando por consiguiente un riesgo menor. Esta condición se aplicaría, por ejemplo, a los sistemas de IA destinados a mejorar el lenguaje utilizado en documentos ya redactados, por ejemplo, en lo referente al empleo de un tono profesional o de un registro lingüístico académico o a la adaptación del texto a una determinada comunicación de marca». En este supuesto, la acción y los elementos básicos de la decisión a adoptar deben ser humanos y previos al uso de la IA. El elemento probatorio será esencial.

c) «Que el sistema de IA tenga por objeto detectar patrones de toma de decisiones o desviaciones con respecto a patrones de toma de decisiones anteriores y no esté destinado a sustituir la evaluación humana previamente realizada sin una revisión humana adecuada, ni a influir en ella». Al respecto de este supuesto, se concreta que «El riesgo sería menor debido a que el sistema de IA se utiliza tras una evaluación previamente efectuada por un humano y no pretende sustituirla o influir en ella sin una revisión adecuada por parte de un ser humano. Por ejemplo, entre los sistemas de IA de este tipo, se incluyen aquellos que pueden utilizarse para comprobar *a posteriori* si un profesor puede haberse desviado de su patrón de calificación determinado, a fin de llamar la atención sobre posibles incoherencias o anomalías» (Cons. 52). En este supuesto, el elemento sustancial parece ser que el sistema IA no está destinado ni a adoptar decisiones ni a influir en las mismas, sino a evaluar las decisiones adoptadas por los humanos.

d) «Que el sistema de IA tenga por objeto llevar a cabo una tarea preparatoria para una evaluación pertinente a efectos de los casos de uso enumerados en el anexo III». Resulta de nuevo interesante el Considerando 52, ya que precisa que «la posible repercusión del resultado del sistema sería muy escasa en términos de representar un riesgo para la subsiguiente evaluación. Esta condición abarca, entre otras cosas, soluciones inteligentes para la gestión de archivos, lo que incluye funciones diversas tales como la indexación, la búsqueda, el tratamiento de texto y del habla o la vinculación de datos a otras fuentes de datos, o bien los sistemas de IA utilizados para la traducción de los documentos iniciales». En este supuesto, el carácter claramente accesorio y alejado de la decisión a adoptar parece ser el elemento distintivo.

Así pues, bastaría que se considere que se da solo una de estas circunstancias para considerar que el sistema no es SAR del anexo III. Obviamente, si concurren varias de estas circunstancias, será más claro que no es SAR. Cabe recordar que un sistema de IA utilizado para fines del anexo III parte de una presunción de que es SAR, y que la excepción es que no lo sea, probando y justificando que se da alguna de estas circunstancias. Igualmente, el artículo 7 RIA, que se expone posteriormente, señala no pocos criterios que pueden ser utilizados para la interpretación y aplicación en cada caso concreto.

2. SERÁ SIEMPRE DE ALTO RIESGO LA ELABORACIÓN DE PERFILES CON INTELIGENCIA ARTIFICIAL PARA FINES DEL ANEXO III

Como premisa, «los sistemas de IA a que se refiere el anexo III siempre se considerarán de alto riesgo cuando el sistema de IA lleve a cabo la elaboración de perfiles de personas físicas» (art. 6. 3º). Según el Considerando 53, procede acudir a la definición del RGPD (art. 4.4) de «elaboración de perfiles». ²³ Así, siempre será SAR un sistema de IA que, para los fines del anexo III, evalúe, ²⁴ analice o prediga aspectos relativos al rendimiento profesional, situación económica, salud, preferencias personales, intereses, fiabilidad, comportamiento, ubicación o movimientos a partir de datos relativos a una persona física. Sin embargo, no sería suficiente que el sistema IA realice «una simple clasificación» si solo es «para obtener una visión global de estos [las personas] sin hacer predicciones ni sacar conclusiones sobre una persona». ²⁵ Siguiendo al CEPD, se trataría de utilizar un sistema automatizado que integre IA para hacer el tratamiento de datos personales para hacer ese tipo de evaluación o análisis, siempre con relación a las finalidades del anexo III. Es importante destacar que sí sería SAR un perfilado con IA que sea parcial y no total, pues vale «toda forma de tratamiento automatizado».

Si se dan estos requisitos, esta elaboración de perfiles con IA para los fines del anexo III sí que será SAR y, como consecuencia, se aplicarán las obligaciones del RIA. Además, claro está, también será aplicable a estos perfilados todo el régimen del RGPD que proceda (Considerando 10 RIA) ²⁶. Entre las normas de protección de datos, en muchas ocasiones —pero no siempre— habrá que aplicar las particularidades de las decisiones totalmente automatizadas artículo 22 RGPD. ²⁷

23. La definición es la misma en el artículo 4.4. Directiva (UE) 2016/680, de 27 de abril de 2016: «4) “elaboración de perfiles”: toda forma de tratamiento automatizado de datos personales consistente en utilizar datos personales para evaluar determinados aspectos personales de una persona física, en particular para analizar o predecir aspectos relativos al rendimiento profesional, situación económica, salud, preferencias personales, intereses, fiabilidad, comportamiento, ubicación o movimientos de dicha persona física».
24. Recuerda el CEPD que «El uso de la palabra “evaluar” sugiere que la elaboración de perfiles implica algún tipo de evaluación o juicio sobre una persona.» Grupo del Artículo 29, *Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679*, 3 de octubre de 2017, versión final 6 de febrero de 2018, Doc WP251rev.01, <https://www.aepd.es/documento/wp251rev01-es.pdf> p. 6-7.
25. *Ídem*.
26. Cons 10. [...] También conviene aclarar que los interesados siguen disfrutando de todos los derechos y garantías que les confiere dicho Derecho de la Unión, incluidos los derechos relacionados con las decisiones individuales totalmente automatizadas, como la elaboración de perfiles.
27. A este respecto G29-UE, *Directrices sobre decisiones ...* cit. p. 8-9 recuerda el CEPD que las decisiones automatizadas tienen un ámbito de aplicación distinto y pueden solaparse o derivarse parcialmente de la elaboración de perfiles. Se recuerda que las decisiones automatizadas pueden llevarse a cabo con o sin elaboración de perfiles; la elaboración de perfiles puede darse sin realizar decisiones automatizadas. No obstante, ambas no son necesariamente actividades independientes. Algo que empieza como un simple proceso de decisiones automatizadas puede convertirse en un proceso basado en la elaboración de perfiles, dependiendo del uso que se dé a los datos. En esta dirección señala como

Finalmente, respecto de la elaboración con perfiles con IA cabe tener en cuenta que pueden estar dentro de las prohibiciones específicas (art. 5.1. d) RIA, Cons. 42)²⁸. También, hay que tener en cuenta las especialidades de los sistemas IA para la aplicación de la ley (Anexo III 6. d y e)²⁹.

3. LAS FINALIDADES DE LOS SISTEMAS DE ALTO RIESGO DEL ANEXO III

En otros apartados de esta obra se analizan de forma más detallada muchos de los supuestos que se consideran SAR en virtud del Anexo III y a ellos cabe remitir en general. No obstante, cabe recordar cuáles son las finalidades y supuestos principales de los SAR del anexo III.

Biometría que no esté prohibida, infraestructuras, educación y trabajo

En primer término, son SAR los sistemas IA de «biometría». Cabe recordar que serán SAR siempre que no estén *totalmente* prohibidos por el artículo 5. Y se afirma que *totalmente* en tanto en cuanto especialmente, en el supuesto de identificación biométrica remota «en tiempo real» en espacios de acceso público con fines de garantía del cumplimiento del Derecho (art. 5. 1º h), están previstas las posibles excepciones legales y bajo sistema de autorización. De modo que, si el sistema no está prohibido, será en todo caso SAR. Así sucede también respecto de los casos no prohibidos bajo la prohibición general de los sistemas para «inferir las emociones» (art. 5. 1º f). Igualmente, serían SAR por no estar prohibidos los «sistemas de categorización

ejemplo la imposición de multas por exceso de velocidad únicamente sobre la base de las pruebas de los radares de velocidad es un proceso de decisiones automatizadas que no implica necesariamente la elaboración de perfiles. Se indica que sería una decisión basada en la elaboración de perfiles si los hábitos de conducción de la persona se supervisan a lo largo del tiempo y, por ejemplo, la cuantía de la multa impuesta es el resultado de una evaluación que implique otros factores, como si el exceso de velocidad es un caso de reincidencia o si el conductor ha cometido otras infracciones de tráfico recientemente. También se señala que las decisiones que no están basadas únicamente en el tratamiento automatizado pueden incluir también la elaboración de perfiles. Por ejemplo, antes de conceder una hipoteca, un banco puede tener en cuenta la calificación crediticia del prestatario, y pueden producirse otras intervenciones humanas significativas adicionales antes de que se tome ninguna decisión sobre la persona.

28. Se trataría de «evaluaciones de riesgos [...] con el fin de evaluar o predecir la probabilidad de que una persona física cometa una infracción penal basándose únicamente en la elaboración del perfil». Y el Considerando 42 indica que «Las personas físicas nunca deben ser juzgadas a partir de comportamientos predichos por una IA basados únicamente en la elaboración de sus perfiles».
29. En principio no sería de alto riesgo hacer un perfilado de datos con IA. El Anexo III 6. D) indica que es de alto riesgo el uso de IA «para evaluar la probabilidad de que una persona física cometa una infracción o reincida en la comisión de una infracción atendiendo no solo a la elaboración de perfiles de personas físicas mencionada en el artículo 3, punto 4, de la Directiva (UE) 2016/680 o para evaluar rasgos y características de la personalidad o comportamientos delictivos pasados de personas físicas o grupos».

Por su parte, según la letra d) es de alto riesgo todo perfilado con IA «durante la detección, la investigación o el enjuiciamiento de infracciones penales». Y en virtud de la letra e) serían de alto riesgo los sistemas IA «para elaborar perfiles de personas físicas [...] durante la detección, la investigación o el enjuiciamiento de infracciones penales».

biométrica que clasifiquen individualmente a las personas» pero que no sea con las finalidades sensibles del artículo 5. 1º g.

En segundo término, dentro de la «biometría» del Anexo III se especifica que son SAR los «sistemas de identificación biométrica remota» (a). Especialmente cabe significar que no serán SAR los sistemas de identificación biométrica uno a uno, es decir, «cuya única finalidad sea confirmar que una persona física concreta es la persona que afirma ser». Ello es relevante porque, pese a que el CEPD³⁰ y la AEPD³¹ consideran recientemente que el tratamiento de datos para esta finalidad de identificación uno a uno sí que es un tratamiento de categoría de datos sensibles del artículo 9 RGPD, no sería un SAR a efectos del RIA por la expresa exclusión del Anexo III. También sería «biometría» los sistemas «de categorización biométrica en función de atributos o características sensibles» o «para el reconocimiento de emociones», en los términos concretados en este apartado 1º.

Por cuanto a los sistemas de IA en «infraestructuras críticas»,³² hay cierta proximidad con el Anexo I por cuanto se maneja el concepto de «componentes de seguridad» (2.a), respecto del cual cabe remitir a lo ya expuesto. En cualquier caso, se especifica que «son sistemas utilizados para proteger directamente la integridad física de las infraestructuras críticas o la salud y la seguridad de las personas y los bienes, pero que no son necesarios para el funcionamiento del sistema». Se pone como ejemplo a «los sistemas de control de la presión del agua o los sistemas de control de las alarmas contra incendios en los centros de computación en la nube». Y es relevante que «los componentes destinados a ser utilizados exclusivamente con fines de ciberseguridad no deben considerarse componentes de seguridad» (Consid. 55).

Respecto de cuáles son las infraestructuras críticas, las mismas se definen en general en el artículo 3. 62º RIA, que remite a su vez al artículo 2, punto 4, considerando «infraestructuras digitales críticas»³³ las «que se enumeran en el anexo I, punto 8, de la Directiva (UE) 2022/2557» (Consid. 55). Asimismo, son infraestructuras críticas las relativas al «tráfico rodado o del suministro de agua, gas, calefacción o electricidad» (2.a). En principio, solo estaría incluido el «tráfico rodado» y no, por tanto, el aéreo o ferroviario.³⁴ La explicación es obvia: «un fallo o un defecto de funcionamiento de estos componentes puede poner en peligro la vida y la salud de las personas a gran escala y alterar de manera considerable el desarrollo habitual de las actividades

30. Así, el CEPD en mayo de 2022 inició este cambio en la primera versión de los CEPD, *Guidelines 05/2022 on the use of facial recognition technology in the area of law enforcement* y en 26 de abril de 2023 se actualizó y reforzó este criterio en la *Version 2.0* https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-052022-use-facial-recognition-technology-area_en

31. La AEPD ha exteriorizado este cambio de criterio en AEPD, *Guía Tratamientos de control de presencia mediante sistemas biométricos* de 23 de noviembre de 2023. <https://www.aepd.es/guias/guia-control-presencia-biometrico.pdf>

32. En el proceso regulatorio tanto la Comisión Europea en 2021 como el Parlamento Europeo en 2023 hablaban de la «Gestión y explotación» de las infraestructuras.

33. En la tramitación del RIA, la mención a las infraestructuras digitales críticas aparece en la versión del Consejo de la UE en diciembre de 2022.

34. Cabe apuntar que en la versión final se hace referencia al «tráfico rodado», en la versión del Consejo de diciembre 2022 se menciona sólo el «tráfico por carretera» y el Parlamento añadió «ferroviario y aéreo», pero no se adoptó en la versión final.

sociales y económicas» (Consid. 55). En cualquier caso, cabe tener en cuenta la legislación específica y la posibilidad de que se trate de un SAR del Anexo II.

En el ámbito de educación (Anexo III. 3^o), se parte de lo positivo que puede ser «fomentar una educación y formación digitales de alta calidad» (Consid. 56). No obstante, algunos usos concretos se consideran SAR por cuanto «pueden invadir especialmente y violar el derecho a la educación y la formación, y el derecho a no sufrir discriminación, además de perpetuar patrones históricos de discriminación» (Consid. 56). A lo largo del proceso regulatorio se fueron delimitando y ampliando estos usos.³⁵ Así, se concreta como SAR los sistemas que determinen el «acceso o la admisión» o «para distribuir a las personas físicas» entre los centros educativos (3.a). También los sistemas IA para la evaluación de los resultados del aprendizaje, en este sentido se subraya que «en particular cuando dichos resultados se utilicen para orientar el proceso de aprendizaje de las personas» (3.b). Asimismo, a partir de la propuesta del Parlamento UE, es SAR la IA «para evaluar el nivel de educación adecuado que recibirá una persona o al que podrá acceder» (3.c). Finalmente, también a propuesta del Parlamento se consideran SAR los sistemas «para el seguimiento y la detección de comportamientos prohibidos [...] durante los exámenes».

El ámbito laboral ha ido ganando presencia a lo largo de todo el proceso de aprobación del RIA. Finalmente, los sistemas de IA para el «Empleo, gestión de los trabajadores y acceso al autoempleo» (Anexo III. 4^o) serán SAR por cuanto «pueden afectar de un modo considerable a las futuras perspectivas laborales, a los medios de subsistencia de dichas personas y a los derechos de los trabajadores» y «pueden perpetuar patrones históricos de discriminación [...] durante todo el proceso de contratación y en la evaluación, promoción o retención de personas en las relaciones contractuales de índole laboral» (Consid. 57).

Más concretamente, serán SAR si se utilizan en concreto «para la contratación o la selección de personas físicas, en particular para publicar anuncios de empleo específicos, analizar y filtrar las solicitudes de empleo y evaluar a los candidatos» (4.a).³⁶ También para la toma de «decisiones que afecten a las condiciones de las relaciones de índole laboral o a la promoción o rescisión de relaciones contractuales de índole laboral, para la asignación de tareas a partir de comportamientos individuales o rasgos o características personales o para supervisar y evaluar el rendimiento y el comportamiento» (4.b).

Servicios y prestaciones esenciales: Administración, emergencias, seguros y banca

Otro grupo de SAR son los relativos al «acceso» y «disfrute» de «servicios privados esenciales y a servicios y prestaciones públicos esenciales» (Anexo III. 5). La consideración de «esenciales» se debe a la adición del Consejo de la UE en su versión de diciembre de 2022. Llama la atención que se aborde conjuntamente el uso público y privado de sistemas IA. Ello se justifica porque se trata de «servicios y prestaciones esenciales [...] necesarios para que las personas puedan participar

35. Así puede apreciarse desde la versión del Consejo de la UE, afinada por el Parlamento, que es esencialmente la finalmente adoptada.

36. En el proceso regulatorio se ha omitido la finalidad de «anunciar vacantes» por la de «publicar anuncios».

plenamente en la sociedad o mejorar su nivel de vida, y el disfrute de dichos servicios y prestaciones» (Consid. 58).

Respecto del uso de IA por la Administración y el sector público cabe remitir al apartado correspondiente en esta obra. Únicamente me permito señalar ahora que, aunque están todos los que son, no son todos los que están; esto es, no cabe duda de que el uso público de sistemas biométricos, de aplicación de la ley, en el ámbito de educación, laboral o de infraestructuras críticas debe ser considerado SAR.³⁷ Sin embargo, parece que el uso de sistemas de IA por la Administración que impacta en los derechos únicamente se centra en «prestaciones y servicios esenciales de asistencia pública». ³⁸ Los mismos incluyen un muy amplio espectro de administraciones («sanitaria», «seguridad social», «servicios sociales» relacionados con «maternidad», «accidentes laborales, la dependencia o la vejez y la pérdida de empleo, asistencia social y ayudas a la vivienda»). Debe recordarse que el uso de sistemas IA por el sector público goza de un —desproporcionado— plazo de adecuación al reglamento de 6 años.

No obstante, parecen quedar al margen del alto riesgo muchos usos públicos de IA que impactan también en no pocos derechos fundamentales. Entre otros, me permito ahora destacar los cada vez más habituales sistemas para la persecución del fraude o fiscalidad,³⁹ que podrían considerarse para la «aplicación de la ley» (Anexo III. 6), pero que en muchos casos quedan al margen de las actuaciones penales de dicho apartado 6.⁴⁰ Esta voluntad queda expresamente recogida en el Considerando 59, que excluye expresamente algunos sistemas utilizados en «procesos administrativos por las autoridades fiscales y aduaneras y las unidades de inteligencia financiera».⁴¹ Me

37. Al respecto, entre otros estudios, me remito a mi estudio «Los usos de la IA en el sector público, su variable impacto y categorización jurídica» *Revista Canaria de Administración Pública*, n.º 1, 2023, pp. 211-242, acceso revista, acceso artículo.

38. En la versión inicial de la Comisión de 2021 o la del Consejo de diciembre de 2022 no se hacía referencia a ningún ámbito concreto. El Parlamento mencionó los «servicios sanitarios y los servicios esenciales, entre otros, vivienda, electricidad, calefacción/refrigeración e Internet», si bien finalmente se ha optado por no especificar en esta letra estos servicios en el ámbito del sector público.

39. Al respecto me remito a mi estudio «Hacia la transparencia 4.0: el uso de la inteligencia artificial y big data para la lucha contra el fraude y la corrupción y las (muchas) exigencias constitucionales», en Ramíó, C. (coord.), *Repensando la administración digital y la innovación pública*, Instituto Nacional de Administración Pública (INAP), Madrid, 2021. <https://links.uv.es/FUW2pz6> Para el ámbito tributario, Olivares, B. D., «Law and Artificial Intelligence in the Spanish Tax Administration: the Need for a Specific Regulation», *European Review of Digital Administration & Law-ERDAL* 1 (1-2), pp. 227-234.

40. Cabe señalar que expresamente se excluyen como de alto riesgo «los sistemas de IA utilizados al objeto de detectar fraudes financieros» (5.b). No obstante, según el Considerando 58 no parece que se piense en sistemas públicos de detección de fraudes. Se trataría, sin embargo, de «sistemas de IA previstos por el Derecho de la Unión con vistas a detectar fraudes en la oferta de servicios financieros y, a efectos prudenciales, para calcular los requisitos de capital de las entidades de crédito y las empresas de seguros no deben considerarse de alto riesgo en virtud del presente Reglamento.»

41. En concreto, «Los sistemas de IA destinados específicamente a ser utilizados en procesos administrativos por las autoridades fiscales y aduaneras y las unidades de inteligencia financiera que llevan a cabo tareas administrativas de análisis de infor-

permite también llamar la atención de que, de haberse seguido la Enmienda 738 del Parlamento, hubieran pasado a ser SAR los sistemas utilizados por cualquier órgano administrativo para «la investigación e interpretación de hechos y de la ley, así como en la aplicación de la ley a un conjunto concreto de hechos», como sucede en el caso del uso de sistemas de IA en justicia. Ello hubiera cambiado mucho la proyección del RIA al sector público, pero no es el caso.

Por cuanto al ámbito de *emergencias*, el apartado 5.d incluye sistemas de IA que serán utilizados por el sector público, como los relativos a «llamadas de emergencia [...] prioridades en [...] situaciones de emergencia [...] policía, bomberos y servicios de asistencia médica, y en sistemas de triaje de pacientes». La versión inicial de la Comisión hablaba de «emergencia, incluidos los bomberos y la ayuda médica», y las precisiones finales se introdujeron posteriormente por el Consejo y el Parlamento.

Más vinculados al sector privado, en este grupo del apartado 5º de servicios esenciales se incluyen los de solvencia crediticia (5.b). Las versiones anteriores excluían los sistemas IA «puestos en servicio por proveedores a pequeña escala para su propio uso» o «por proveedores que sean microempresas y pequeñas empresas». Sin embargo, no se excluyen finalmente, sin perjuicio de la aplicación del artículo 63. Sí que se excluyen, desde la versión del Parlamento en junio de 2023, los utilizados para la detección del fraude financiero. La consideración de SAR a los servicios de solvencia crediticia se justifica por cuanto implica el posible acceso a recursos y servicios esenciales como «vivienda, electricidad y servicios de telecomunicaciones» (Consid. 58).

Aunque no estaba en la versión inicial de la Comisión de 2021, en las diversas versiones del Consejo de la UE, aunque no en su propuesta de diciembre de 2022, aparecían y desaparecían los sistemas IA del ámbito del seguro. Desde el Parlamento en junio de 2023 se propuso como SAR los sistemas para «la elegibilidad de personas físicas para seguros de salud y de vida». En la versión final adoptada, son SAR los sistemas IA para la «evaluación de riesgos y la fijación de precios en relación con las personas físicas en el caso de los seguros de vida y de salud» (5.c). Ello se debe a que «pueden afectar de un modo considerable a los medios de subsistencia de las personas y [...] pueden vulnerar sus derechos fundamentales y tener graves consecuencias para la vida y la salud de las personas, como la exclusión financiera y la discriminación».

Aplicación de la ley, migración, asilo y gestión del control fronterizo, justicia y procesos democráticos

Por cuanto a los sistemas para la aplicación de la ley (Anexo III. 6), sin duda, uno de los más importantes del Anexo III, cabe remitir asimismo al capítulo de esta obra centrado en esta cuestión, recordando que están centrados en el ámbito de la actuación penal. Por afinidad, cabe proyectar lo ahí expuesto respecto de los sistemas IA SAR para la «Migración, asilo y gestión del control fronterizo» (Anexo III. 7). En este ámbito, se reitera que la posibilidad de uso de sistemas IA solo se dará si hay una

mación de conformidad con el Derecho de la Unión en materia de lucha contra el blanqueo de capitales no deben clasificarse como sistemas de IA de alto riesgo usados por las autoridades encargadas de la aplicación de la ley con el fin de prevenir, detectar, investigar y enjuiciar infracciones penales.»

habilitación y regulación específica.⁴² Además, cabe tener especialmente en cuenta los requisitos procedimentales establecidos por la normativa aplicable.⁴³

En el último apartado, se consideran SAR los sistemas IA para la «Administración de justicia y procesos democráticos» (Anexo III. 8).⁴⁴ En este ámbito se ha ido precisando el alcance desde la primera versión de 2021. En la versión final, son SAR los sistemas IA «destinados a ser utilizados por una autoridad judicial» (8.a) para «ayudar» en la investigación e interpretación de hechos y de la ley, así como en la aplicación de la ley a un conjunto concreto de hechos. También se incluyen los sistemas IA «utilizados de forma similar en una resolución alternativa de litigios» (8.a), a lo que se añade «cuando los resultados de los procedimientos de resolución alternativa de litigios surtan efectos jurídicos para las partes» (Consid. 61).

El RIA excluye como SAR «los sistemas de IA destinados a actividades administrativas meramente accesorias que no afectan a la administración de justicia propiamente dicha en casos concretos, como la anonimización o seudonimización de resoluciones judiciales, documentos o datos, la comunicación entre los miembros del personal o las tareas administrativas» (Consid. 61).

La delimitación del uso sustantivo jurisdiccional respecto del uso administrativo y no jurisdiccional es una cuestión que no siempre está preclara y puede generar problemas interpretativos futuros.⁴⁵ De hecho, entre otras cuestiones, puede determinar la autoridad que será competente para la supervisión de los sistemas IA de Administración de Justicia, siendo la AEPD, la AESIA o el CGPJ.⁴⁶

Finalmente, a propuesta del Parlamento (Enmienda 739), se incluyen como SAR los sistemas IA para «procesos democráticos», en concreto «para influir en el resultado de una elección o referéndum o en el comportamiento electoral de personas físicas que ejerzan su derecho de voto en elecciones o referendos». Se hace la prevención de que no serán SAR los sistemas «a cuya información de salida no estén directamente expuestas las personas físicas, como las herramientas utilizadas para organizar, optimizar o estructurar campañas políticas desde un punto de vista administrativo

42. Considerando 60: «en la medida en que su utilización esté permitida en virtud del Derecho de la Unión y nacional».

43. Como el Reglamento (CE) n.º 810/2009 del Parlamento Europeo y el Consejo o la Directiva 2013/32/UE del Parlamento Europeo y del Consejo. Como el Reglamento (CE) n.º 810/2009 del Parlamento Europeo y el Consejo o la Directiva 2013/32/UE del Parlamento Europeo y del Consejo.

44. Así, la versión de 2021 de la Comisión hablaba de sistemas para «asistir a una autoridad judicial en la investigación e interpretación de los hechos y del Derecho y en la aplicación del Derecho a un conjunto concreto de hechos.» A lo que se fue añadiendo que los sistemas puedan ser «utilizados por una autoridad judicial, o en su nombre» (desde Versión del Consejo de la UE).

45. Cabe recordar que a los efectos de la protección de datos y de la autoridad competencia, el artículo 236 bis LOPJ distingue entre tratamientos de datos personales realizados con fines jurisdiccionales y no jurisdiccionales. Se consideran jurisdiccionales los tratamientos de datos incorporados en procesos que tengan por finalidad el ejercicio de la actividad jurisdiccional.

46. Al respecto, mi estudio «El uso jurisdiccional de la inteligencia artificial: habilitación legal, garantías necesarias y la supervisión por el CGPJ», *Actualidad Jurídica Iberoamericana*, n.º 21, 2024, monográfico. <https://revista-aji.com/>

o logístico» (Anexo III 8.b) y Consid. 62).⁴⁷ Cabe recordar que finalmente no prosperó la Enmienda 740 del Parlamento por la que serían SAR sistemas de recomendación de grandes plataformas reguladas en la DSA.⁴⁸ Sobre este tema cabe remitir en bloque al capítulo de esta obra relativo al tratamiento de las grandes plataformas y sistemas de inteligencia artificial destinados a la influencia política, donde se tiene especialmente en cuenta la conexión con otras normas como la DSA, la DMA y, en especial, con el recientemente aprobado Reglamento (UE) 2024/900 del Parlamento Europeo y del Consejo, de 13 de marzo de 2024, sobre transparencia y segmentación en la publicidad política.

4. PRESUNCIÓN DE QUE EL SISTEMA INTELIGENCIA ARTIFICIAL QUE PERSIGUE FINES DEL ANEXO III SÍ QUE ES DE ALTO RIESGO. ESPECIALES OBLIGACIONES Y ACTUACIONES

Es posible que el proveedor que desarrolla un sistema de IA con finalidades del Anexo III considere que no es SAR. Como se ha insistido, si la finalidad prevista (art. 3.12 RIA) para la que está concebido el sistema de IA coincide con las finalidades del Anexo III, se debe presumir que sí es SAR y que sólo excepcionalmente no lo será.

Para estos supuestos, los proveedores deben actuar automáticamente y, además, existe un protocolo de actuación de la autoridad de vigilancia del mercado. Así, en estos casos, los proveedores deben documentar que su sistema no es SAR y, en todo caso, deben incluirlos en el registro (art. 6. 4º y Considerando 52).⁴⁹ Resulta un deber llamativo que se impone a los proveedores de IA que consideran que su sistema no es SAR, pero que se presume que sí lo es. Por otro lado, el artículo 80 establece que, si la autoridad sospecha que podría ser un SAR, debe evaluarlo. Si la evaluación revela que el sistema de IA es efectivamente SAR, la autoridad demandará al proveedor que tome las medidas necesarias para cumplir con el RIA y corrija el problema en un plazo fijado por la autoridad. Si el uso del sistema de IA va más allá del ámbito nacional, la autoridad de vigilancia debe informar a la Comisión Europea y a los otros Estados miembros sobre la evaluación y las medidas requeridas al proveedor.

47. En este caso el Considerando 62 no hace aportación alguna.

48. «a ter) sistemas de IA destinados a ser utilizados por plataformas de redes sociales designadas como plataformas en línea de muy gran tamaño en el sentido del artículo 33 del Reglamento (UE) 2022/2065, en sus sistemas de recomendación para recomendar al destinatario del servicio contenido generado por los usuarios disponible en la plataforma.»

49. Así, artículo 6.4º: «El proveedor que considere que un sistema de IA contemplado en el anexo III no es de alto riesgo documentará su evaluación antes de que dicho sistema sea introducido en el mercado o puesto en servicio. Dicho proveedor estará sujeto a la obligación de registro establecida en el artículo 49, apartado 2. A petición de las autoridades nacionales competentes, el proveedor facilitará la documentación de la evaluación.» Por su parte el Considerando 52 señala que «Para garantizar la trazabilidad y la transparencia, los proveedores que, basándose en estas condiciones, consideren que un sistema de IA no es de alto riesgo, deben elaborar la documentación de la evaluación previamente a la introducción en el mercado o la entrada en servicio de dicho sistema y facilitarla a las autoridades nacionales competentes cuando estas lo soliciten. Dichos proveedores deben tener la obligación de registrar el sistema en la base de datos de la UE creada en virtud del presente Reglamento.»

Considero que estas situaciones pueden darse especialmente en los casos en los que el proveedor desarrolla un sistema que *potencialmente* sirve a finalidades determinadas del Anexo III, pero que *no controla el uso concreto que hará el usuario o responsable de la implementación del sistema*. Es decir, el proveedor de un sistema de IA para la contratación, monitorización o despido laboral no es quien contrata, controla o despide a los trabajadores de la empresa. En estos casos, el proveedor debe valorar si las inferencias generadas por el sistema de IA pueden pasar a ser un elemento sustancial que sirva de soporte para las finalidades del Anexo III por el usuario de su sistema. Si así lo considera razonablemente, el proveedor debe partir de que su sistema sí es SAR y, en consecuencia, desarrollarlo siguiendo las exigencias del RIA. En estos casos, no es suficiente que el proveedor se limite a expresar en sus instrucciones de uso o el marco contractual con el usuario o implementador que no debe ser utilizado como elemento sustancial para la toma de decisiones relativas a las finalidades del Anexo III. El sistema será SAR si genera salidas que lo hacen idóneo para basar decisiones para estas finalidades. En todo caso, habrá que analizar el supuesto concreto y también la posibilidad de que el proveedor acuda al artículo 6. 4º RIA y documente y justifique que el sistema no es SAR. Se da la situación paradójica de que los proveedores tienen obligaciones ante las autoridades de vigilancia del mercado, y éstas mayormente actuarán si consideran que su sistema no es SAR. Mientras que, si el sistema sí es SAR, es muy posible que el proveedor *sólo* tenga que hacer una autoevaluación de conformidad y posiblemente no sea supervisado por la autoridad.

5. CUANDO EL IMPLEMENTADOR ALTERA UN SISTEMA Y PASA A TENER UNA FINALIDAD DE ALTO RIESGO

Otra situación puede darse cuando un proveedor desarrolla un sistema de IA que inicialmente no está destinado a las finalidades del Anexo III, pero que podría ser alterado en su sistema o finalidad para adaptarse a ellas. Esto se consideraría un «uso indebido razonablemente previsible» (art. 3.13),⁵⁰ esencialmente por parte del usuario o implantador que lo destina a las finalidades del Anexo III.

En este caso, se debe distinguir el improbable supuesto de que ese sistema de IA ya sea SAR. En tal caso, el proveedor tiene ciertas obligaciones. Debe gestionar los riesgos y evaluar la posibilidad de que ocurra este uso indebido para una finalidad del Anexo III, y prever medidas para mitigar estos riesgos y sus efectos (art. 9.2 b) RIA⁵¹ y Consid. 65). Además, si existe algún riesgo, debe informar al usuario o implantador del sistema en las instrucciones de uso. También debe prevenir este riesgo de uso para finalidades del Anexo III mediante control humano (art. 14. 2º

50. «uso indebido razonablemente previsible»: la utilización de un sistema de IA de un modo que no corresponde a su finalidad prevista, pero que puede derivarse de un comportamiento humano o una interacción con otros sistemas, incluidos otros sistemas de IA, razonablemente previsible.

51. Artículo 9. Sistema de gestión de riesgos [...] 2º b) «la estimación y la evaluación de los riesgos que podrían surgir cuando el sistema de IA de alto riesgo se utilice de conformidad con su finalidad prevista y cuando se le dé un uso indebido razonablemente previsible».

RIA).⁵² En cualquier caso, si el implantador realiza una «modificación sustancial» del SAR, por ejemplo, para que adopte directamente decisiones en el contexto de las finalidades del Anexo III, asumirá las obligaciones del proveedor (art. 25.1º b).⁵³

La situación es más compleja cuando el sistema de IA no es SAR ni persigue las finalidades del Anexo III. En estos casos, se trataría de sistemas fuera del marco general de obligaciones del RIA y quedarían al margen de la previsión del artículo 6. 4º RIA. Sin embargo, será claramente aplicable el artículo 25. 1º c) si el responsable del despliegue de un sistema de IA «modifica la finalidad prevista» y lo «convierte en un sistema de IA de alto riesgo de conformidad con el artículo 6».⁵⁴ En tal caso, se considerará proveedor y deberá cumplir sus obligaciones.

V. EL PAPEL DE LA COMISIÓN, CRITERIOS, ACTOS DELEGADOS, ACTUALIZACIÓN Y MODIFICACIÓN DE LOS SISTEMAS DE ALTO RIESGO, EN ESPECIAL, DEL ANEXO III

Aunque el RIA ha fijado criterios para la consideración de los SAR del Anexo III, éstos quedan claramente sujetos a la casuística e interpretación. Consciente de ello, se atribuye un poder importante a la Comisión para delimitar más estos supuestos tanto a través de directrices como de actos delegados. Además, el propio RIA concreta a la Comisión los parámetros que deben guiar su actuación.

Así, según el art. 6.5 RIA, «la Comisión [...] dará directrices que especifiquen la aplicación práctica del presente artículo en consonancia con el artículo 96, junto con una lista exhaustiva de ejemplos prácticos de casos de uso de sistemas de IA que sean de alto riesgo y que no sean de alto riesgo». Estas directrices deben ubicarse en el ámbito del artículo 96 RIA, por lo que la Comisión tendrá especialmente en cuenta a las PYMES o autoridades públicas locales, así como el estado de la técnica, normas armonizadas o especificaciones técnicas. Asimismo, la Comisión «adoptará actos delegados» para «modificar» e incluso añadir «nuevas condiciones» de los criterios SAR vinculados al Anexo III del art. 6.3.1.

Es importante señalar que el punto de partida en estas actuaciones de la Comisión es que si un sistema de IA persigue los fines del Anexo III es SAR y sólo excepcionalmente no lo será (Consid. 52: «no se consideran, con carácter excepcional, de alto riesgo»). En consecuencia, la actuación de la Comisión a la hora de determinar criterios y actos delegados habrá de estar bien justificada, especialmente cuando se

52. Artículo 14 Vigilancia humana. [...] 2º «El objetivo de la vigilancia humana será prevenir o reducir al mínimo los riesgos para la salud, la seguridad o los derechos fundamentales que pueden surgir cuando se utiliza un sistema de IA de alto riesgo conforme a su finalidad prevista o cuando se le da un uso indebido razonablemente previsible, en particular cuando dichos riesgos persistan a pesar de la aplicación de otros requisitos establecidos en la presente sección.»

53. «b) cuando modifique sustancialmente un sistema de IA de alto riesgo que ya haya sido introducido en el mercado o puesto en servicio de tal manera que siga siendo un sistema de IA de alto riesgo con arreglo al artículo 6;».

54. «c) cuando modifique la finalidad prevista de un sistema de IA, incluido un sistema de IA de uso general, que no haya sido considerado de alto riesgo y ya haya sido introducido en el mercado o puesto en servicio, de tal manera que el sistema de IA en cuestión se convierta en un sistema de IA de alto riesgo de conformidad con el artículo 6.»

trate de no considerar SAR a sistemas utilizados para fines del Anexo III. En esta dirección, el RIA contiene algunos criterios que ha de seguir la Comisión respecto de los actos delegados (art. 6.6):

— Podrá suprimir o modificar las condiciones para que se considere SAR «únicamente cuando existan pruebas concretas y fiables de la existencia de sistemas de IA que entren en el ámbito de aplicación del Anexo III, pero que no planteen un riesgo importante de causar un perjuicio a la salud, la seguridad o los derechos fundamentales».

— Asimismo, «ninguna modificación reducirá el nivel global de protección».

— «Tendrá en cuenta la evolución tecnológica y del mercado».

Además, el artículo 7 regula la «adición o modificación» de casos de uso del Anexo III a través de estos actos delegados. Para esta adicción, los ámbitos deben ser los ya incluidos en el Anexo III y debe realizarse una evaluación de riesgos en la que este artículo señala muchos criterios que la Comisión ha de tener en cuenta, tales como: la finalidad concreta del sistema, el grado de uso de IA, «la naturaleza y la cantidad de los datos tratados, en particular si se tratan categorías especiales de datos personales», «el grado de autonomía del sistema de IA y la posibilidad de que un ser humano anule una decisión o recomendación», impactos efectivos en salud, seguridad o derechos fundamentales o la probabilidad de que se den según informes o alegaciones documentadas, alcance de este perjuicio, «gran número de personas o afectación desproporcionada a un grupo determinado de personas», dependencia que puedan tener estas personas del resultado de ese uso de IA, «desequilibrio de poder y posición de vulnerabilidad», situación, autoridad, conocimientos, circunstancias económicas o sociales, o edad del afectado. También, la posibilidad de «corregir o revertir el resultado» y la probabilidad de que el despliegue del sistema de IA resulte beneficioso, así como «la medida en que el Derecho de la Unión vigente establezca medidas de compensación efectivas para prevenir o reducir notablemente esos riesgos».

Como se ha adelantado, cabe entender que estos criterios que ha de seguir la Comisión pueden ser también criterios en los que apoyarse para la evaluación de si un sistema es SAR en razón del artículo 6.

Además de los criterios para los actos delegados de la Comisión, el RIA también contiene algunos mandatos operativos para que anualmente elabore «la lista de sistemas de IA de alto riesgo». Así, «reviste especial importancia que la Comisión lleve a cabo las consultas oportunas durante la fase preparatoria, en particular con expertos» (Consid. 173). La lista SAR deberá evaluarse una vez al año (Consid. 173). Esto es relevante dado que la revisión general del Reglamento debería darse a los cinco años y luego cada cuatro años, y el Anexo III debería revisarse cada cuatro años (y a los dos años de aplicarse el RIA, Consid. 174).

VI. RECAPITULACIÓN Y CONCLUSIONES

La UE ha adoptado un modelo basado en riesgos para la regulación de la IA, y el concepto de SAR ha pasado a ser clave. A mi juicio, son de verdadero alto riesgo los que el RIA define, y su impacto en las personas, los derechos y el sistema democrático es incuestionable. Según se ha analizado, la consideración de «alto riesgo» es el pilar

central del RIA, y buena parte de la normativa gira en torno a este concepto. Se han analizado los elementos esenciales de lo que es un SAR, subrayando la importancia de que la regulación de un sistema de IA como SAR por el RIA no implica su habilitación legal en términos de límites legales de derechos fundamentales, protección de datos, justicia u otros ámbitos. Esto significa que la clasificación como SAR no exige a los desarrolladores y usuarios de contar con una base legal específica para el uso de tales sistemas.

El RIA establece un sistema dual para calificar un SAR. Por un lado, los sistemas de IA que son productos y componentes de seguridad o productos del Anexo I, específicamente aquellos que requieren una evaluación de conformidad por parte de un tercero independiente. Estos sistemas de IA son SAR especialmente por los peligros a la integridad, seguridad y fiabilidad. Por otro lado, y posiblemente más importantes, el Anexo III del RIA detalla hasta veinticinco finalidades que califican a un sistema de IA como SAR. Estas incluyen aplicaciones de IA en biometría, infraestructuras críticas, educación, ámbito laboral, servicios esenciales, emergencias, aplicación de la ley, migración, asilo, control fronterizo, administración de justicia y procesos democráticos. Se ha descrito sumariamente estas finalidades, que en algunas ocasiones son objeto de mayor análisis en otros apartados de esta obra. El impacto en este caso está claramente ligado a numerosos derechos fundamentales.

Se ha expuesto que ya no hay un automatismo por el que si el sistema tiene una de las 25 finalidades del Anexo III pasa a ser automáticamente un SAR. En la evolución del RIA se ha incorporado el requisito de que el sistema de IA debe tener una influencia sustancial en la toma de decisiones. Así, se considera que un sistema de IA influye sustancialmente en la toma de decisiones si sus resultados afectan directamente decisiones importantes para las finalidades del Anexo III. Afortunadamente, este criterio se ha regulado de modo mucho más perfilado que las decisiones únicamente automatizadas del artículo 22 RGPD. Este criterio es esencial para determinar la clasificación de alto riesgo y, por tanto, si aplican las obligaciones correspondientes.

Se ha insistido en que existe una presunción de que cualquier sistema de IA que persiga fines incluidos en el Anexo III es SAR, lo que conlleva obligaciones y actuaciones especiales por parte de los proveedores y usuarios. Esta presunción garantiza un nivel de precaución, obligaciones para los proveedores y controles adicionales, lo que podría generar problemas en el futuro. Podría resultar más *rentable* para un proveedor no hacer nada, antes que negar que su sistema es un SAR. También se ha especificado que siempre será SAR la elaboración de perfiles con IA para los fines del Anexo III.

Finalmente, se ha podido observar cómo la Comisión Europea va a jugar un papel esencial en la actualización y modificación de los elementos que definen un SAR. Esto se llevará a cabo a través de criterios y actos delegados, para asegurar que el RIA se mantenga actualizado y eficaz en respuesta a los avances tecnológicos y cambios en el mercado. Sólo el tiempo dirá si la UE se ha excedido al considerar SAR y, por tanto, imponer toda una serie de obligaciones.

La regulación de los sistemas policiales predictivos en el Reglamento Inteligencia Artificial

FERNANDO MIRÓ LLINARES Y MARIO SANTISTEBAN GALARZA

Fernando Miró Llinares. Catedrático de Derecho Penal y Director Centro CRIMINA. Universidad Miguel Hernández de Elche¹

MARIO SANTISTEBAN GALARZA

Universidad del País Vasco

I. INTRODUCCIÓN

Antes de que la IA fuese una realidad ya se había instaurado en el imaginario colectivo algunas ideas, generalmente distópicas, sobre el uso policial de estas tecnologías. Bien en forma de policía robótico o bien de sistemas que predicen el delito antes de que acontezca como los imaginados por Philipp K. Dick en el relato «El informe en minoría» (*minority report*), la cultura había explicitado con anterioridad al desarrollo tecnológico algunas de las promesas y de los riesgos que esta tecnología podía traer consigo. Es comprensible, por tanto, que uno de los primeros usos de sistemas de IA que preocupase socialmente fuese el policial, más cuando se comenzaba a saber que ya existían sistemas algorítmicos (unos basados en IA otros no) enmarcados dentro del concepto general de «policía predictiva». Lo cierto es que, fundamentalmente a través de la legislación de protección de datos, se apreciaba la tendencia de limitar ciertos usos policiales de los sistemas algorítmicos. Altos Tribunales han limitado el tratamiento masivo de datos sobre infracciones penales por la ausencia de garantías oportunas², o por la opacidad

1. La publicación de este trabajo es parte del proyecto Ius-Machina, sobre las bases normativas y el impacto real de la utilización de algoritmos predictivos en los ámbitos judicial y penitenciario TED2021— 129356B-I00, financiado por MCIN/AEI/10.13039/501100011033 y por la Unión Europea «NextGenerationEU»/PRTR. Este trabajo se enmarca en el proyecto GODAS*: Proyecto PID2022-137140OB-I00, financiado por el MCIN/AEI/10.13039/501100011033/FEDER, UE.
2. Es el caso del Tribunal Constitucional Alemán, véase en este sentido Cotino Hueso, L., «Una regulación legal y de calidad para los análisis automatizados de datos o con inteligencia artificial. Los altos estándares que exigen el Tribunal constitucional alemán y otros tribunales, que no se cumplen ni de lejos en España», *Revista General de Derecho Administrativo*, (2023).

de ciertos sistemas algorítmicos³. El Tribunal de Justicia de la Unión Europea ha determinado la ilicitud de la recogida sistemática de datos biométricos y genéticos en el marco del proceso penal⁴. El Tribunal Europeo de Derechos Humanos⁵ ha limitado el uso de la tecnología de reconocimiento facial por parte de autoridades policiales, supeditándolo a que sea necesario en una sociedad democrática. También la Directiva 2016/680 de 27 de abril de 2016 relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por parte de las autoridades competentes para fines de prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de sanciones penales, establece garantías frente el procesamiento de categorías especiales de datos en el marco de la aplicación de la ley penal, y frente a las decisiones automatizadas. No obstante, su naturaleza de Directiva y la amplitud con la que se acerca a la cuestión de la automatización de las decisiones aconsejan respuestas específicas a los problemas de la IA en el ámbito de aplicación de la ley penal⁶.

De ahí, por tanto, que ante la decisión de la Unión Europea de regular la IA, uno de los ámbitos objeto de regulación «preferencial» fuera el uso policial de estos sistemas y, particularmente, de los predictivos. El presente trabajo aborda la regulación por parte del Reglamento General de Inteligencia Artificial de los sistemas predictivos de IA usados, o potencialmente utilizables, en el ámbito policial y de justicia penal. Lo hace con el propósito de comprender las disposiciones regulatorias, de delimitar su ámbito de aplicación y aportar criterios interpretativos para la misma, pero, también, de hacerlo en relación con el uso actual y potencial policial de los mismos. La intención es comprender no sólo qué usos se declaran prohibidos y cuáles son considerados de alto riesgo y qué es lo que esto implica, sino enmarcar todo ello dentro del papel que tales sistemas, y otros similares, tienen actualmente dentro de la actuación policial. Así, trataremos de entender no sólo las implicaciones normativas desde la perspectiva de qué derechos ciudadanos no se van a poder ver afectados debido a la prohibición de algunos sistemas o a la imposición de obligaciones para otros sistemas considerados de alto riesgo, sino lo que ello conlleva desde la propia realidad policial que, en los últimos años, estaba inmersa en una fuerte tendencia de tecnificación y de reorientación de sus funciones hacia la prevención y la predicción. Por ello, comenzaremos por encuadrar el concepto de policía predictiva, proceder a

3. Es conocido como en el Holanda el sistema SyRI, diseñado para detectar fraudes en determinados aspectos del sistema de seguridad social fue declarado contrario al art. 8 del Convenio Europeo de Derechos Humanos, teniendo especial importancia la falta de transparencia de los parámetros del sistema en la decisión. Véase a este respecto Appelman, N., Ó Fathaigh, R., y van Hoboken, J., «Social Welfare, Risk Profiling and Fundamental Rights: The Case of SyRI in the Netherlands», *JIPITEC* 257 12 (2021).
4. «El mero hecho de que se investigue a una persona por la comisión de un delito público doloso no puede considerarse un dato que permita presumir, por sí solo, que la recogida de sus datos biométricos y genéticos es estrictamente necesaria a la vista de los fines que persigue y habida cuenta de las vulneraciones de los derechos fundamentales, en particular, de los derechos al respeto de la vida privada y de protección de los datos personales garantizados por los artículos 7 y 8 de la Carta, que de ella se derivan» STJUE de 26 de enero de 2023 (asunto C-205/21), apdo. 130.
5. STEDH de 4 de julio de 2023 (GLUKHIN v. RUSIA, demanda n. 11519/20).
6. Simón Castellano, P., «Inteligencia artificial y Administración de Justicia: ¿Quo vadis, justitia?», *IDP. Revista de Internet, Derecho y Política*, (2021), n.º 33.

una simple categorización de los sistemas algorítmicos que entran dentro del mismo y a una explicación de cuáles de ellos usan IA y qué singularidades aporta la misma, para que cuando procedamos al análisis del texto legal podamos enmarcarlo en relación con la realidad que está regulando y comprender mejor sus implicaciones. El tercer apartado lo dedicaremos a la evolución del texto reglamentario para situar la regulación final en su contexto, y en el cuarto trataremos de hacer una recapitulación, esbozando unas breves conclusiones.

II. SISTEMAS POLICIALES PREDICTIVOS E INTELIGENCIA ARTIFICIAL

1. ORGANIZACIÓN POLICIAL EN TIEMPOS DE DIGITALIZACIÓN Y LA «MAL LLAMADA» POLICÍA PREDICTIVA

La actividad policial ha cambiado enormemente a lo largo del tiempo. Como «organización pública permanente encargada del mantenimiento de la seguridad y el orden a través del ejercicio cuasi monopolístico de potestades estatales (básicamente la coerción)»⁷ la policía, tal y como la conocemos en la actualidad, nace con la revolución industrial y no se explica sin la aparición del Estado de derecho y con el desarrollo de las ciudades, pero sus funciones y el modo de ejercerlas han ido cambiando debido a múltiples factores relacionados con la visión política y social sobre cuál debe ser el ejercicio del control o con la evolución del tipo de problemas que ha ido afrontando. Otro de los motores de cambio de la institución y la actividad policial ha sido, precisamente, la tecnología⁸. Quizás porque ha determinado las posibilidades de la acción policial y, por tanto, configurado la visión de la misma de la ciudadanía y de los poderes políticos, lo cierto es que los cambios tecnológicos no sólo han modificado el modo de ejercer la acción policial, sino que han acabado afectando a cuáles son sus funciones. La aparición de la radio y del automóvil determinaron una policía más reactiva que la anterior, menos ocupada de la investigación del crimen y más centrada en reaccionar y controlar el mismo; la digitalización y el proceso de «datificación» en el que estamos dirigió la función policial hacia la gestión y prevención policial del crimen⁹, mediante el uso de la información relacionada con el delito y la gestión de los recursos para usar el control social tratando de evitar que el delito acontezca. Dentro de esa etapa, la aparición del Big Data y de los sistemas de IA parece conducir la acción policial hacia la automatización de la labor preventiva bajo la idea de «predicción».

Efectivamente, el tipo de práctica policial que se encuadra bajo la denominación de «policía predictiva» no se entendería si no es en este contexto de digitalización, de «cientificación» y de burocratización y automatización del trabajo policial que se ha producido en las últimas décadas. De hecho se atribuye el primer uso del término a William Bratton, el padre de Compstat, el sistema estadístico de criminalidad que, desde las bases del Problem Oriented Policing de Goldstein y de la informatización de

-
7. Guillén Lasiera, F., *Modelos de policía. Hacia un modelo de seguridad plural*, Barcelona, Bosch, (2016).
 8. Deflem, M., y Chicoine, S. «History of technology in policing. *J Psychopharmacol*», 24 (2) (1988), pp. 141-145.
 9. González-Álvarez, J., Santos Hermoso, J., y Camacho-Collados, M., «Policía predictiva en España. Aplicación y retos futuros», *Behavior & Law journal*, 6(1), (2020).

la actividad policial comenzada en los 60 y 70, desarrolló a finales de los 80 un sistema estadístico de integración informatizada de datos policiales¹⁰ que, con el tiempo, como ha señalado Wilson, aceleró la recopilación y el procesamiento de los datos, determinó la integración de la informatización en el trabajo rutinario de las patrullas a una escala sin precedentes, incrementó significativamente la utilización de técnicas cartográficas de mapeo de la delincuencia para analizar la distribución geográfica de las patrullas y situar las «horas del día» y los «picos de delincuencia» en los conocidos «hot spots» y, con el apoyo del desarrollo académico de la denominada «criminología ambiental», dio lugar más adelante al desarrollo de técnicas de análisis delictivo como el Intelligence Led policing o el Evidence Based Policing que defienden la toma de decisiones basada en datos y la resolución estratégica de problemas para la gestión policial, la asignación de recursos y el control de la delincuencia. Pero ¿es esto la policía predictiva o es algo más? y ¿cómo se relaciona todo ello con la IA?

El término policía predictiva comenzó a generalizarse a principios de la década de 2010 en el ámbito académico para referirse a un conjunto de «técnicas analíticas —particularmente técnicas cuantitativas— que, mediante la realización de predicciones estadísticas, tratan bien de identificar posibles objetivos de intervención policial y prevenir el crimen o, bien, resolver delitos pasados»¹¹. Similar es la definición que propone Ratcliffe quien, además de preferir el término «crime forecasting»¹², lo define como «el uso de datos históricos para crear pronósticos de áreas de criminalidad o puntos críticos de criminalidad, o perfiles característicos de delincuentes de alto riesgo que serán un componente de las decisiones de asignación de recursos policiales»¹³. En los últimos años el término predictive policing ha comenzado a ser abandonado por sus principales usuarios, los departamentos de policía de EEUU¹⁴. Pero se usa una nueva marca, «data driven policing», o acción policial basada en datos, tras la cual sigue vigente el uso de algunas de estas herramientas. Quizás porque las promesas de incremento de la eficiencia en la toma de decisiones, de reducción de sesgos y subjetividad en las mismas, etc, y de solucionismo tecnológico no sólo siguen vigentes, sino que se ven incrementadas por la generalización de los sistemas de IA¹⁵.

-
10. Miró Llinares, F., «Predictive Policing: Utopia or Dystopia? On attitudes towards the use of Big Data algorithms for law enforcement», *IDP. Revista de Internet, Derecho y Política*, n.º 30., (2020).
 11. Walter, P., McInnis, B., Price, C., Smith, S., y and Hollywood, J., *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*, Santa Monica, CA: RAND Corporation, (2013).
 12. Ratcliffe, J. «Predictive policing», en Weisburd, D y Braga, A.A. (EDS.), *Police innovation. Contrasting perspectives*. 2nd. edition. Cambridge: Cambridge University Press (2019).
 13. *Ibíd.*
 14. Especialmente a partir de que el Departamento de Policía de Los ángeles en la primavera de 2020 (al que siguieron muchos otros después) dejara de usar el sistema «predpol» por las críticas de discriminación racial que se hacía a estas herramientas y que aumentaron con el movimiento BLM. Davis, J., Purves, D., Gilbert, J., & Sturm, S. (2022). Five ethical challenges facing data-driven policing. *AI and Ethics*, 2(1), 185-198.
 15. López Riba, JM., «Inteligencia artificial y control policial. Cuestiones para un debate criminológico frente al hype», en prensa, (2024).

Más allá de que tenga razón Wilson cuando señala que dentro del amplio término de policía predictiva nos encontramos con un «popurrí de metodologías y filosofías policiales contemporáneas» que enlazan la digitalización y la tendencia a la automatización en la toma de decisiones con una idea de predicción¹⁶, existe acuerdo en destacar dos tipos de técnicas de pronóstico delictivo, los pronósticos delictivos enfocados en el lugar en el que se va a cometer el crimen, y los pronósticos delictivos centrados en las personas que lo van a perpetrar o van a ser víctimas del mismo. Ambas técnicas se pueden llevar a cabo utilizando o no sistemas de IA, algo sobre lo que volveremos más adelante cuando tratemos la cuestión de los riesgos éticos asociados a las mismas. Ambas técnicas, además, se caracterizan por aprovechar los datos históricos sobre el crimen para identificar objetivos de interés policial con el propósito de prevenir el delito, reducir su riesgo o causar una disrupción sobre la actividad criminal¹⁷, pero se diferencian entre sí por el tipo de input del que se alimentan y el output que generan. Las técnicas de policía predictiva que se basan en el lugar (Place Based Predictive Policing), se centran en la determinación del dónde y cuándo se perpetran los delitos de cada tipo y extraen patrones de riesgo en tales entornos para mejorar la toma de decisiones sobre dónde y cuándo intervenir preventivamente. Las técnicas de policía predictiva que se basan en las características de las personas (Person Based Predictive Policing)¹⁸ que los perpetran (agresores)¹⁹ o de quienes las sufren (víctimas), establecen patrones o perfiles generales de los mismos y estiman quiénes y cuándo tienen más posibilidades de volver a perpetrar delitos o de ser víctimas de estos²⁰. Unas y otras técnicas tienen elementos comunes: a) buscan estimar la posibilidad de que se perpetre un crimen, b) utilizan datos de los crímenes ya cometidos y, por tanto, información de criminales y de víctimas, c) son herramientas informáticas de apoyo a la toma de decisiones y, d) son algoritmos predefinidos y pueden usar, o no, aunque la mayoría no lo hace, algoritmos de aprendizaje automático. Pero también tienen características distintivas muy relevantes: 1) las bases científicas de cada uno de ellos son muy diferentes (la criminología ambiental en el caso de las PlaceBPP²¹ y, generalmente, la valoración del riesgo de violencia en el caso de las de PersonBPP²²; 2) las herramientas basadas en

16. Wilson, D. «Predictive policing management: A brief history of patrol automation», *New formations*, 98(98), (2019). 139-155.
17. Ratcliffe, J. «Predictive policing», en Weisburd, D y Braga, A.A. (EDS.), *Police innovation. Contrasting perspectives*. 2nd. edition. Cambridge: Cambridge University Press (2019).
18. También han sido denominadas Ofender Focused Crime Forecasting tools, aunque en ese caso hacen referencia exclusivamente a la estimación de las posibilidades de perpetración delictiva no incluyendo las de victimización.
19. Sobre las heat list y otros sistemas similares de cálculo de riesgo de personas individuales como Beware, véase Degeling, M. y Berendt, B., ¿«What is wrong about Robocops as consultants? A technology-centric critique of predictive policing AI and Society», 33 (3) (2018). pp. 347-356.
20. Es el caso de Viogen, Presno Linera, M. A.,. «Policía predictiva y prevención de la violencia de género: el sistema VioGén» IDP. *Revista de Internet, Derecho y Política*, 2023, n.º 39, pp. 1-13.
21. Véase por todos, FELSON, M. «Routine activities and crime prevention in the developing metropolis», *Criminology*, vol. 25, (1987).
22. González-Álvarez, J., Santos Hermoso, J., y Camacho-Collados, M., «Policía predictiva en España. Aplicación y retos futuros», *Behavior & Law journal*, 6(1), (2020).

el lugar sólo tienen en cuenta los datos del criminal o de la víctima que se relacionan con el tipo de infracción y con el lugar en el que se ha perpetrado pero no realizan un perfil de posibles víctimas o agresores a partir de la generalización de quienes perpetran (o padecen) tales delitos; 3) dan lugar a actuaciones policiales de muy diferente naturaleza.

Junto a estas dos técnicas que pueden considerarse propiamente de pronóstico policial sobre el delito o de, mal llamada, policía predictiva, hay un tercer y último conjunto de técnicas policiales que en ocasiones se dan por incluidas dentro de ese cajón de sastre y son todas aquellas técnicas basadas en la vigilancia por medio de imágenes y que, a partir de técnicas de reconocimiento facial, de movimientos, lectura de matrículas, etc., se combinarían con algoritmos para, identificar a sujetos sospechosos y «predecir» posibles acciones delictivas²³. Las organizaciones policiales cada vez usan con más frecuencia software de reconocimiento facial que se nutre de las cámaras de tráfico, de cámaras corporales llevadas por los propios policías o de cámaras lectoras de matrículas y similares que producen datos digitales que pueden combinarse para identificar y rastrear a las personas en aras de la seguridad y la protección²⁴. Aunque se dice que estas técnicas podrían servir para predecir la realización de conductas ilícitas²⁵, lo cierto es que las mismas son más bien técnicas de investigación de crímenes ya cometidos o que están en proceso de perpetración, que de estimación de la probabilidad de que se cometan delitos. Su consideración como policía predictiva deviene de la consideración de que ese amplio cajón de sastre también incluya la utilización de técnicas para «predecir» la identidad de los autores de delitos²⁶. En todo caso el análisis de estas herramientas, que más bien entran en la lógica de la investigación del crimen que en la de la prevención/predicción del mismo, quedará al margen de este capítulo al estar íntimamente relacionadas con las técnicas de reconocimiento facial que han recibido un tratamiento especial por parte del RIA puesto que los riesgos éticos que plantean son diferentes.

2. LOS RIESGOS ÉTICOS DE LA POLICÍA PREDICTIVA (Y LOS QUE AÑADE EL USO DE INTELIGENCIA ARTIFICIAL)

En ese proceso de burocratización y digitalización de la policía, la IA se presenta como el último exponente de la mejora de la eficacia y de la eficiencia (e, incluso, de reducción de la subjetividad) en el ejercicio de sus tareas, en general, y en la prevención del crimen en particular. Con el Big Data y la aparición de nuevas técnicas computacionales como el machine learning se abre la posibilidad de mejorar las estimaciones que antes se hacían sin tantos datos o sin estas técnicas. Al emplear

23. Miró Llinares, F., «Predictive Policing: Utopia or Dystopia? On attitudes towards the use of Big Data algorithms for law enforcement», *IDP. Revista de Internet, Derecho y Política*, n.º 30., (2020).

24. Hannah-Moffat, K. «Algorithmic risk governance: Big data analytics, race and information activism in criminal justice debates», *Theoretical Criminology*, 23(4), (2019), 453-470.

25. Ferguson, A. G. *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. NY: NYU Press, (2017).

26. Brayne, S., Rosenblat, A., y Boyd, D., *Predictive policing. Data & civil rights: a new era of policing and justice*, (2015). https://www.datacivilrights.org/pubs/2015-1027/Predictive_Policing.pdf

IA se mejoraría la precisión en las estimaciones, se optimizaría la distribución de los recursos y, por tanto, se reducirían los costes de la acción policial²⁷ e incluso se garantizaría una actuación más equitativa²⁸. De hecho, estas ideas parecen ir en línea con lo establecido. Ideas que parecen ir en línea por el propio RIA, que en el considerando 3 señala que «al mejorar la predicción, optimizar las operaciones y la asignación de recursos y personalizar las soluciones digitales disponibles para particulares y organizaciones, el uso de la inteligencia artificial puede proporcionar ventajas competitivas clave a las empresas y respaldar resultados beneficiosos» desde muchos puntos de vista de entre los cuales cita expresamente el de la seguridad. Pero el mismo reglamento, en el considerando 4 reconoce que los sistemas de IA pueden «generar riesgos y causar daños a los intereses públicos y a los derechos fundamentales protegidos por el Derecho de la Unión». Es necesario, por tanto, identificar cuáles pueden ser estos riesgos para poder valorar el enfoque de la regulación europea, también en el ámbito del uso de sistemas policiales predictivos.

Pues bien, a nuestro parecer hay dos planos que hay que tener en cuenta de forma diferenciada, primero, y conectada, después, a la hora de asignar posibles riesgos éticos por el uso policial de sistemas denominados predictivos: por un lado, estaría el plano de la técnica de pronóstico (bien de lugar, bien de persona) que está detrás de la herramienta, y por otro el de la técnica algorítmica (bien clásica o bien automatizada o de IA) que se usa a nivel computacional. Cada uno de esas dimensiones o planos permite identificar técnicas diferentes, pero, también, riesgos a derechos y garantías propios y distintivos que, además, pueden combinarse de distintas formas entre sí.

Atendiendo, en primer lugar, al plano de la técnica de pronóstico utilizada por la herramienta (bien de lugar, bien de persona), aunque habrá problemas o riesgos éticos compartidos otros serán diferentes o propios sólo de algunas de ellas por a) la distinta naturaleza de los datos de los que se nutren unas y otras, b), por su diferente lógica científica, o c), por la distinta naturaleza del impacto que conlleva la actuación policial predictiva en uno o en otro caso. En cuanto a la cuestión de los datos, su calidad, y los posibles sesgos que pueden producir, los sistemas policiales que pretenden pronosticar en qué lugares se perpetrarán los crímenes se suelen nutrir de datos de denuncias, aunque también pueden nutrirse de los datos de los delitos investigados por la policía. Este tipo de datos pueden verse afectados por diferentes sesgos relacionados con la mayor presencia policial en esas zonas, con el modo en que se recogen los datos por la policía, con los hechos delictivos que son más denunciados, entre otros muchos²⁹. En estos casos, y veremos que esto es importante de cara al RIA, no podemos decir que se estén llevando a cabo patrones de personas

27. Brayne, S., Rosenblat, A., y Boyd, D., *Predictive policing. Data & civil rights: a new era of policing and justice*, (2015). https://www.datacivilrights.org/pubs/2015-1027/Predictive_Policing.pdf

28. A esta narrativa que conforma una táctica tecnopolítica usada por organismos policiales la denomina Saphiro «policía predictiva para la reforma», un intento erróneo de racionalizar las patrullas policiales mediante una reestructuración algorítmicas de su actuar. Saphiro, A., «Predictive Policing for Reform? Indeterminacy and Intervention in Big Data Policing», *Surveillance & Society*, 17(3/4), 2019, pp. 456-472.

29. Véase sobre todo esto, con múltiples referencias y detalles, López Riba, JM., «Inteligencia artificial y control policial. Cuestiones para un debate criminológico frente al hype», en prensa, 2024.

atendiendo a las características genéricas asociadas a las mismas puesto que son los lugares los que son objeto de análisis. Pero eso no significa que los posibles sesgos no acaben afectando a las personas de forma indirecta. Así, el hecho de que se patrulle más en determinadas zonas puede llevar a que se lleven a cabo un mayor número de intervenciones y paradas policiales y, por tanto, determinar un mayor número de detenciones y de infracciones penales que acaben afectando a las personas que viven en esas zonas al verse más expuestas a la vigilancia policial por los algoritmos contruidos sobre tal información.

En el caso de los sistemas policiales que tratan de pronosticar la posibilidad de que una persona cometa un delito o sea víctima del mismo o de mejorar la toma de decisiones relacionadas con ello, los datos de los que se nutren tales algoritmos si incluyen múltiples variables relacionadas con características personales relacionadas con la edad, los antecedentes penales, el género, y otros factores similares que se relacionan con la perpetración de crímenes o con la victimización por los mismos.

Además de la cuestión de los datos, que la predicción sea relativa al lugar del crimen o a las personas implicadas en él incide en otros problemas éticos. Uno de ellos tiene que ver con la validez y fiabilidad de los pronósticos, no porque sean mejores generalmente los de un caso o los del otro, sino porque las bases científicas de las que parten unos y otros sistemas son diferentes. Las herramientas de PlaceBPP se basan en las premisas de la geografía del crimen y en las técnicas del análisis geográfico del delito y han mostrado gran fiabilidad en algunos experimentos³⁰ si bien es más discutido que sean capaces de prevenir y reducir la delincuencia más allá de simplemente pronosticar la intervención policial³¹. Hay otros problemas éticos, además, relacionados con la aplicación de estas herramientas, como que cambien el modo en que actúan los policías y dejen de realizar funciones comunitarias esenciales para pasar a ser detectives de lugares de riesgo³², o que confundan pronóstico con intervención³³. En el caso de las PersonBPP la valoración del riesgo de violencia, como método alternativo al diagnóstico de peligrosidad para la predicción de la violencia³⁴, que se apoya en el conocimiento de los factores de riesgo asociados a la violencia identificado las causas que explican y los factores que se relacionan con la conducta violenta (factores de riesgo) y también aquellos otros que influyen en la reducción o abandono de la actividad violenta y/o delictiva (factores de protección), que pueden ser comunes o específicos a diferentes formas de violencia que, a su vez, puede estar más o menos relacionada con la *conducta delictiva*³⁵. En el ámbito policial, pero también

30. Ratcliffe, J. «Predictive policing», en Weisburd, D y Braga, A.A. (EDS.), *Police innovation. Contrasting perspectives*. 2nd. edition. Cambridge: Cambridge University Press (2019).

31. Miró Llinares, F., «Predictive Policing: Utopia or Dystopia? On attitudes towards the use of Big Data algorithms for law enforcement», *IDP. Revista de Internet, Derecho y Política*, n.º 30., (2020).

También, López Riba, JM., «Inteligencia artificial y control policial. Cuestiones para un debate criminológico frente al hype», en prensa, (2024).

32. Ferguson, A. G. *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. NY: NYU Press, (2017).

33. *Ibíd.*

34. Andres Pueyo, A. y Illescas, S. R.. «Predicción de la violencia: entre la peligrosidad y la valoración del riesgo de violencia». *Papeles del psicólogo*, 28(3), (2007), 157-173.

35. *Ibíd.*

en el judicial para la toma de decisiones relacionada con la estimación del riesgo de reincidencia o de quebrantamiento de condena, ha aumentado significativamente el uso de estos instrumentos actuariales y de juicio clínico estructurado de base actuarial a partir de la consideración de su alta capacidad predictiva³⁶ que, sin embargo, puede y debe ser matizada. Lo ha hecho Martínez Garay poniendo de manifiesto que estas herramientas alcanzan unos niveles de precisión probablemente satisfactorios en la estimación del riesgo relativo de reincidencia, pero son más imprecisos en la estimación del riesgo absoluto de reincidencia, y producen una sobreestimación del riesgo al aplicarse a fenómenos con bajas tasas de prevalencia como lo es la delincuencia violenta³⁷. Siendo esto así, lo relevante es comprender que no se trata de dicotomizar si estas herramientas predicen o no predicen, sino en comprender cuáles de las herramientas existentes disponibles dan mejores criterios de forma comparada para la toma de decisiones judicial sobre la peligrosidad³⁸, y, también, cuáles son las limitaciones reales de cada sistema para realizar cada concreta valoración en cada tipo de comportamiento³⁹.

Y esto se debe ligar con el otro elemento fundamental que debe tenerse en consideración a la hora de valorar los riesgos éticos asociados a uno u otro tipo de pronóstico: las consecuencias normativas de la actuación policial en uno y otro caso. Mientras que las técnicas de predictive policing basadas en el lugar sirven esencialmente para tomar decisiones consistentes en la asignación de recursos policiales y, por tanto, sus consecuencias consistirán en que habrá zonas geográficas más o menos vigiladas y con mayor o menor presencia policial, fundamentalmente, las técnicas de pronóstico personal pueden tener consecuencias directamente relacionadas con los derechos fundamentales. Como ha señalado Martínez Garay generar expectativas acerca de las posibilidades reales de predicción de conductas violentas es más peligroso allí donde las consecuencias es la posible afectación de la libertad de los ciudadanos⁴⁰. Esto también nos puede llevar a diferenciar entre los problemas éticos asociados a las herramientas que estiman la comisión de delitos frente a las que estiman la victimización. Si bien puede ser razonable de inicio pensar que estas últimas plantean menos problemas lo cierto es que tanto unas como otras están, en realidad, generalmente basadas en la valoración del riesgo de violencia

36. Andres Pueyo, A. y Illescas, S. R. «Predicción de la violencia: entre la peligrosidad y la valoración del riesgo de violencia». *Papeles del psicólogo*, 28(3), (2007), 157-173, también, Skeem, J. L., & Monahan, J. «Current directions in violence risk assessment. *Current Directions in Psychological Science*», 20(1), (2011), pp. 38-42.

37. Martínez Garay, L. «Errores conceptuales en la estimación de riesgo de reincidencia: La importancia de diferenciar sensibilidad y valor predictivo, y estimaciones de riesgo absolutas y relativas». *Revista Española de Investigación Criminológica: REIC*, (14), (2016).

38. Miró Llinares, F. y Castro Toledo, F. J., «¿Correlación no implica causalidad? El valor de las predicciones algorítmicas en el sistema penal a propósito del debate epistemológico sobre el fin de la teoría'» en Demetrio, E., de la Cuerda Martín, M. y García de la Torre García, F., *Derecho penal y comportamiento humano. Avances desde la neurociencia y la inteligencia artificial*, Tirant lo Blanch, 2022.

39. Martínez Garay, L., & Montes Suay, F., «El uso de valoraciones del riesgo de violencia en Derecho Penal: algunas cautelas necesarias», *InDret: Revista para el análisis del derecho*, 2018, N. 2/2018, (2018), 1-46.

40. *Ibíd.*

atendiendo a la conducta de un potencial agresor, por lo que habrá que ponderar si las medidas de unas y otras estimaciones son diferentes o si, por el hecho de ser aplicadas a la víctima para su protección las medidas van a tener también un impacto en el potencial agresor.

Más allá del tipo de pronóstico llevado a cabo por cada herramienta o técnica hay un segundo plano que hay que tener en cuenta a la hora de evaluar los riesgos éticos que plantean los sistemas que, demasiado genéricamente, suelen englobarse de policía predictiva. Se trata del que atiende a la técnica computacional utilizada y, en particular, a si el sistema está o no basado en IA, entendiéndose por esta, tal y como establece el artículo 3 a) del Reglamento, «un sistema basado en máquinas diseñado para funcionar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación tras su despliegue y que, para objetivos explícitos o implícitos, infiere, a partir de la entrada que recibe, cómo generar salidas tales como predicciones, contenidos, recomendaciones o decisiones que pueden influir en entornos físicos o virtuales»⁴¹. Lo cierto es que los sistemas policiales predictivos nacieron antes de la popularización de IA y la mayoría de ellos se desarrollaron sin la utilización de técnicas de aprendizaje automatizado como el machine learning. Sólo a partir de la generalización del Big Data y de la posibilidad de acceder a grandes cantidades de información comenzó a plantearse la posibilidad de utilización de estas técnicas para mejorar las estimaciones tanto para sistemas de pronóstico del lugar cómo de las personas. Pero ¿por qué, en términos de riesgos éticos, es relevante que los sistemas predictivos utilicen o no técnicas computacionales de aprendizaje automático o similares que puedan decir que estamos ante IA al tener cierto grado de independencia de las acciones de la intervención humana? O, planteado de otro modo, ¿Qué es lo que, en términos de riesgo ético, añade el hecho de que se utilice IA? A nuestro parecer hay tres cuestiones fundamentales: la diferente lógica científica que está detrás de los algoritmos de IA; la cuestión de la trazabilidad y explicabilidad de las decisiones y; en relación con todo ello y de fondo, el problema de la autonomía de la IA.

Comenzando por la diferente lógica científica de unos y otros sistemas, los algoritmos predictivos clásicos surgidos en plena digitalización venían usando grandes conjuntos de datos oficiales. Pero estos habían sido deliberadamente configurados por investigadores en ciencias sociales, siguiendo directrices metodológicas y a partir de marcos teóricos concretos basados en una lógica científica

41. Como precisa el considerando 6 del citado reglamento, esto implica no incluir dentro de estos sistemas otros softwares tradicionales más sencillos y tampoco sistemas que se basen únicamente en reglas definidas únicamente por personas físicas para ejecutar operaciones automáticamente. Señala así mismo el considerando que «Las técnicas que permiten la inferencia al construir un sistema de IA incluyen enfoques de aprendizaje automático que aprenden a partir de los datos cómo alcanzar determinados objetivos; y enfoques basados en la lógica y el conocimiento que infieren a partir del conocimiento codificado o la representación simbólica de la tarea que debe resolverse», y también que el que los sistemas de IA estén diseñados para funcionar con distintos niveles de autonomía «significa que tienen cierto grado de independencia de las acciones de la intervención humana y de capacidades para funcionar sin intervención humana», entre lo cual incluye las «capacidades de autoaprendizaje, que permiten al sistema cambiar mientras se utiliza». Sobre esto volveremos más adelante.

causal para estimar el riesgo de violencia o los patrones del crimen, según la modalidad de herramienta⁴². La llegada de la analítica de Big data supuso un cambio radical, no sólo por las grandes cantidades de información a las que se puede acceder sino porque permite, incluso demanda, el empleo de nuevas técnicas que están pensadas desde una lógica distinta en la que la inferencia causal no es relevante y la correlación entre factores lo es todo⁴³. A diferencia de los algoritmos predictivos tradicionales, estos nuevos algoritmos parten de grandes cantidades de macrodatos; son capaces de funcionar con datos en tiempo real y de adaptarse gracias al aprendizaje automático; no están limitadas ni en los datos que se recogen ni en los resultados que producen por marcos teóricos predeterminados y, en realidad, tampoco tienen por qué estar pensadas implícitamente para predecir el lugar en el que se va a producir el delito o si alguien va a perpetrar el crimen, pero son capaces de contrastar múltiples variables sobre individuos, lugares y sociedades para realizar pronósticos supuestamente más precisos sobre muy distintos elementos⁴⁴. Los resultados de estos algoritmos no tratan de explicar por qué alguien cometerá un delito o dónde se producirá este, sino de estimarlo independientemente de cuáles sean las variables que lleven al pronóstico y el sentido causal de las mismas. Estos algoritmos, por tanto, podrían tener menos problemas de producir sesgos relacionados con la selección de las variables por parte de los investigadores sociales que tengan una determinada visión del problema, pero pueden conllevar el riesgo de no ser capaces de explicar el sentido de las estimaciones.

De hecho, íntimamente relacionado con esto, se ha dicho que la IA conlleva más problemas de trazabilidad de las decisiones por dos motivos. El primero por el tema de la opacidad de los algoritmos, aunque el denominado problema de la «caja negra» no es propio exclusivamente de la IA y también puede venir asociada a instrumentos actuariales o de predicción de lugares tradicionales. Muchos de estos algoritmos no son accesibles al público y tienen derechos de propiedad intelectual que impide el acceso a los mismos, por lo que las razones de las decisiones no podrían ser seguidas planteando problemas de falta de transparencia y, en particular, el enorme riesgo de que el derecho a la defensa no se pueda ejercer adecuadamente. Pero, además, y, en segundo lugar, los algoritmos de IA, incluso aunque fueran trazables, son fluidos y transformativos, cambiando de forma impredecible a partir de los nuevos datos que van introduciendo⁴⁵ y, en el caso de aquellas herramientas que utilicen Deep learning, dando lugar a soluciones difíciles de explicar en términos causales.

42. Hannah-Moffat, K., «Algorithmic risk governance: Big data analytics, race and information activism in criminal justice debates», *Theoretical Criminology*, 23(4), (2019), 453-470.

43. Miró Llinares, F. y Castro Toledo, F. J., «¿Correlación no implica causalidad? El valor de las predicciones algorítmicas en el sistema penal a propósito del debate epistemológico sobre “el fin de la teoría”» en Demetrop, E., de la Cuerda Martín, M. y García de la Torre García, F., *Derecho penal y comportamiento humano. Avances desde la neurociencia y la inteligencia artificial*, Tirant lo Blanch, (2022).

44. Hannah-Moffat, K., «Algorithmic risk governance: Big data analytics, race and information activism in criminal justice debates», *Theoretical Criminology*, 23(4), (2019), 453-470.

45. Danaher, J., Hogan, M. J., Noone, C., Kennedy, R., Behan, A., De Paor, A., Felzmann, H., Haklay, M., Khoo, S.-M., Morison, J., Murphy, M. H., O’Brolchain, N., Schafer, B., & Shankar, K. «Algorithmic governance: Developing a research agenda through the power of collective intelligence», *Big Data & Society*, 4(2), (2017).

Y esto está relacionado, a su vez, con la última de las características que añade, en términos de riesgo, singularidad, a los algoritmos predictivos que usan IA: la posibilidad de que los algoritmos actúen autónomamente, aunque sea en el aprendizaje de los datos. Al funcionamiento de algunos sistemas de aprendizaje automatizado, por ejemplo, los de Deep learning, en los que la ausencia de entrenamiento y la dificultad de determinación de las variables implicadas ya hacen difícil explicar (comprender) el origen de las decisiones se une el que algunos de ellos puedan «autónomamente» decidir cómo seleccionar unas u otras variables sin que ello sea supervisado y, por tanto, pueda ser evitado por un humano⁴⁶. Obviamente los riesgos que esto supone son muy distintos. Los sistemas de pronóstico tradicionales están pensados como guías o herramientas de apoyo al servicio de los profesionales que no les sustituyen a estos en las tomas de decisiones, sino que les informan, y se supone que quienes las aplican están formados para comprender la lógica de las recomendaciones. En los sistemas de IA el hecho de que la recomendación no sea del todo explicable puede producir una automatización de su aplicación por parte del sujeto que conlleva riesgos especialmente relevantes.

III. LA REGULACIÓN DE LA POLICÍA PREDICTIVA EN EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL

1. EVOLUCIÓN DE LA REGULACIÓN DEL USO DE SISTEMAS POLICIALES PREDICTIVOS EN EL PROCESO LEGISLATIVO DEL REGLAMENTO DE INTELIGENCIA ARTIFICIAL

La regulación de la policía predictiva ha experimentado cambios sustanciales en el procedimiento legislativo, evidenciando por un lado la complejidad de la tarea a

46. Muy relacionado con lo anterior el TJUE se ha mostrado reacio a aceptar el uso del machine learning en ciertos usos policiales. Es el caso de la utilización de datos del registro de nombres de los pasajeros de aerolíneas para detectar delitos de terrorismo y delitos graves facilitado por la Directiva 2016/681, que establece, en palabras del TJUE «un régimen de vigilancia continuo, no selectivo y sistemático, que incluye la evaluación automatizada de datos de carácter personal de todas las personas que utilizan los servicios de transporte aéreo». Pues bien, la comparación de datos PNR con las bases de datos pertinentes se realiza en base a criterios determinados, lo que se opone «a la utilización de las tecnologías de inteligencia artificial en el marco de sistemas de autoaprendizaje (machine learning), que puedan alterar, sin intervención y sin control humanos, el proceso de evaluación y, en particular, los criterios de evaluación en los que se basa el resultado de la aplicación del proceso, así como la ponderación de dichos criterios». En este sentido, el TJUE aclara que «el recurso a estas tecnologías entrañaría el riesgo de privar de efecto útil a la revisión individualizada de los resultados positivos y al control de licitud exigido por las disposiciones de la Directiva PNR. En efecto, como hace constar, en esencia, el Abogado General en el punto 228 de sus conclusiones, habida cuenta de la opacidad que caracteriza el funcionamiento de las tecnologías de inteligencia artificial, puede resultar imposible comprender la razón por la cual un determinado programa ha alcanzado una concordancia positiva. En estas circunstancias, el uso de esas tecnologías también podría privar a las personas afectadas de su derecho a la tutela judicial efectiva, que consagra el artículo 47 de la Carta y que la Directiva PNR pretende, a tenor de su considerando 28, garantizar en un nivel elevado, en particular para cuestionar el carácter no discriminatorio de los resultados obtenidos».

la que se enfrentaba el legislador comunitario y, por otro, la naturaleza especialmente controvertida de esta materia.

Inicialmente, la Propuesta de Reglamento del Parlamento Europeo y del Consejo de la Comisión Europea, presentada en abril del año 2021, no proscribió el uso de las herramientas de policía predictiva. Entre las prácticas prohibidas mencionadas en el Título II, salvo la identificación biométrica remota en tiempo real, no objeto de este capítulo, ninguna de ellas abarcaba el uso de sistemas de IA para valorar el riesgo de comisión de delitos. En cambio, y por remisión del art. 6.2 del texto, el uso de algunas de las herramientas que hemos descrito podía encajar en las prácticas del Anexo III, y constituir, consiguientemente, un uso de IA de alto riesgo. El citado anexo comprendía ciertas aplicaciones «relacionados con la aplicación de la ley» (Law enforcement). Entre ellas, el apartado 6 a) mencionaba «sistemas de IA destinados a utilizarse por parte de las autoridades encargadas de la aplicación de la ley para llevar a cabo evaluaciones de riesgos individuales de personas físicas con el objetivo de determinar el riesgo de que cometan infracciones penales o reincidan en su comisión, así como el riesgo para las potenciales víctimas de delitos». Esto abarcaría determinados sistemas de policía predictiva ya descritos, con independencia de la fase del procedimiento de aplicación de la ley penal y que informen tanto decisiones policiales o judiciales.

Además de ello, el anexo III incluía también como IA de alto riesgo aquellos «sistemas de IA destinados a utilizarse por parte de las autoridades encargadas de la aplicación de la ley para predecir la frecuencia o reiteración de una infracción penal real o potencial con base en la elaboración de perfiles de personas físicas, de conformidad con lo dispuesto en el artículo 3, apartado 4, de la Directiva (UE) 2016/680, o en la evaluación de rasgos y características de la personalidad o conductas delictivas pasadas de personas físicas o grupos» (6. e)). Volveremos sobre esta definición más abajo pues parece ser clave para determinar lo que se encuentra prohibido en el texto final. Por el momento, lo que nos interesa resaltar es la similitud entre ambas prácticas descritas (letras a) y e)), siendo muy difícil diferenciar «la predicción de la frecuencia o reiteración de una infracción penal» (letra 6 e)), de la evaluación del riesgo de la comisión del delito o la recidiva criminal (letra a)). Muy ligado a lo interior, la realización de «evaluaciones de riesgos individuales de personas físicas con el objetivo de determinar el riesgo de que cometan infracciones penales», supone, en la práctica, la elaboración de perfiles de los presuntos delincuentes en los términos fijados en la mencionada Directiva (UE) 2016/680, lo que de nuevo enturbia la diferenciación de los supuestos de hecho a los que alude la norma en ambos apartados. Seguramente la dificultad de diferenciar entre ambos supuestos de hecho ha motivado que en versiones posteriores se haya prescindido de esta división.

Finalmente, en el apartado 6 g), se introducía un uso de alto riesgo con un supuesto de hecho, en este caso, claramente diferenciado: «sistemas de IA destinados a utilizarse para llevar a cabo análisis sobre infracciones penales en relación con personas físicas que permitan a las autoridades encargadas de la aplicación de la ley examinar grandes conjuntos de datos complejos vinculados y no vinculados, disponibles en diferentes fuentes o formatos, para detectar modelos desconocidos o descubrir relaciones ocultas en los datos». En este caso, la propuesta de la Comisión parecía abarcar las técnicas de PlaceBPP ya que el proceso de evaluación no tiene como misión valorar el riesgo que comporta un sujeto, con independencia de que

esta evaluación se nutra de datos relativos a infracciones penales que sí contengan datos personales⁴⁷.

Algunos cambios relevantes se produjeron ya con la posición Común del Consejo⁴⁸. En particular introdujo ciertas modificaciones relevantes en la definición de estas prácticas y, también, en el listado de sistemas de alto riesgo contemplado en el anexo III. En lo que aquí nos interesa resaltar se eliminó la letra g) del apartado 6, esto es, lo que la posición común del Consejo identificaba con sistemas para realizar análisis del crimen (*crime analytics*), y que ya hemos señalado que puede abarcar sistemas de policía predictiva basados en el lugar. Por su lado, las letras a) y e) se mantuvieron si bien se clarificó que estas herramientas se consideraban de alto riesgo cuando se utilizasen por las autoridades encargadas de aplicación de la ley, pero también por otros sujetos en delegación de estas autoridades.

El cambio sustancial en el régimen jurídico de estas herramientas lo encontramos en las enmiendas del Parlamento Europeo. Ciertamente, el legislador ya había mostrado su recelo ante el uso de estas herramientas, manifestando en una resolución del año 2021, «que si bien la actuación policial predictiva puede analizar los conjuntos de datos necesarios para la determinación de patrones y correlaciones, no puede responder a la cuestión de la causalidad y no puede hacer predicciones fiables del comportamiento individual, por lo que no puede constituir la única base de una intervención»⁴⁹. Pues bien, ya en la citada resolución se oponía «al uso de la IA por parte de las autoridades policiales para hacer predicciones conductuales relativas a individuos o grupos sobre la base de datos históricos y comportamientos pasados, pertenencia a un grupo, ubicación o cualquier otra característica de este tipo, para tratar así de identificar a personas que probablemente vayan a cometer un delito».

Esta desconfianza ante la policía predictiva se plasmó de forma tajante en las enmiendas presentadas⁵⁰. El cambio más relevante respecto al texto de la Comisión

47. Debemos hacer mención, por otro lado, a otros sistemas que, no pudiendo ser considerados en puridad como «policía predictiva», si entran dentro de lo que se considera «predictive» o «evidence based» *sentencing*, y que eran considerados como un sistema de alto riesgo por el texto de la Comisión. El artículo 8.A se refería en particular a aquellos sistemas de IA destinados «a ayudar a una autoridad judicial en la investigación e interpretación de hechos y de la ley, así como en la aplicación de la ley a un conjunto concreto de hechos» (8. A)). Consiguientemente, todas las formas de *evidence based sentencing* entran, en cualquier caso, y en la medida en que no participen de las categorías anteriores, dentro del grupo de sistemas de alto riesgo.

48. Disponible en: <https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf>

49. Resolución del Parlamento Europeo, de 6 de octubre de 2021, sobre la inteligencia artificial en el Derecho penal y su utilización por las autoridades policiales y judiciales en asuntos penales (2020/2016(INI)).

50. Seguramente contribuyó también el posicionamiento del CEPD-SEPD en su Dictamen conjunto 5/2021.

Sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial), de 18 de junio de 2021. Sobre los sistemas de *policing* señalaron lo siguiente: «La determinación o clasificación por un ordenador de la conducta futura con independencia de la voluntad propia también afecta a la dignidad humana. Los sistemas de IA destinados a ser utilizados por las autoridades encargadas

consistía en la inclusión de en los usos prohibidos del art. 5 de algunas técnicas que encajarían dentro de la amplia caja de la policía predictiva, pasando del uso de alto riesgo a la prohibición. Así, la enmienda 224 del Parlamento elevó a la categoría de uso prohibido la siguiente práctica de IA:

la introducción el mercado, la puesta en servicio o la utilización de sistemas de IA para llevar a cabo evaluaciones de riesgo de personas físicas o grupos de personas físicas con el objetivo de determinar el riesgo de que estas personas cometan delitos o infracciones o reincidan en su comisión, o para predecir la comisión o reiteración de un delito o infracción administrativa reales o potenciales, mediante la elaboración del perfil de personas físicas o la evaluación de rasgos de personalidad y características, en particular la ubicación de la persona o las conductas delictivas pasadas de personas físicas o grupos de personas físicas (nuevo art. 5. 1 d bis).

La práctica prohibida proscribe sistemas de valoración del riesgo, con independencia de la fase del proceso en la que esta tiene lugar, y que suponga la elaboración de perfiles (profiling) de la persona. Parece, de hecho, que las enmiendas del Parlamento siguieron de cerca a la normativa de protección de datos, que establece limitaciones relevantes a los tratamientos que constituyen elaboración de perfiles. Esto es lógico en atención a los «potenciales efectos discriminatorios en las personas físicas por motivos de raza u origen étnico, opiniones políticas, religión o creencias, afiliación sindical, condición genética o estado de salud u orientación sexual»⁵¹ que pueden provocar las decisiones automatizadas basadas en el perfilado⁵². Así, los considerandos del texto enmendado por el parlamento aducen que estos sistemas «entrañan un riesgo particular de discriminación contra determinadas personas o grupos de personas, ya que vulneran la dignidad humana, así como el principio jurídico clave de presunción de inocencia» (considerando 26 bis, enmienda 50).

Por otro lado, en lo que se refiere a los sistemas de alto riesgo, en contraposición a la posición común del Consejo, se mantiene la inclusión de letra g), debiendo de considerarse que aquellos sistemas de PlaceBPP se mantienen en esta categoría.

de la aplicación de la ley para llevar a cabo evaluaciones de riesgos individuales de personas físicas con el objetivo de determinar el riesgo de que cometan infracciones penales [véase el anexo III, apartado 6, letra a)], o para predecir la frecuencia o reiteración de una infracción penal real o potencial con base en la elaboración de perfiles de personas físicas o en la evaluación de rasgos y características de la personalidad o conductas delictivas pasadas [véase el anexo III, apartado 6, letra e)] utilizados según su finalidad prevista conducirán a la dominación fundamental de la toma de decisiones policiales y judiciales, con la consiguiente cosificación de la persona afectada. Dichos sistemas de IA, que afectan a la esencia del derecho a la dignidad humana, deberán prohibirse en virtud del artículo 5».

51. STUE de 7 de diciembre de 2023, asunto C-634/21, (OQ y Land Hessen).

52. También la Recomendación CM/Rec (2010)13 del Comité de Ministros a los Estados miembros sobre la protección de las personas con respecto al tratamiento automatizado de datos de carácter personal en el contexto de la creación de perfiles señaló que «dicha creación de perfiles puede exponer a las personas a riesgos particularmente elevados de discriminación y de atentados contra sus derechos personales y su dignidad».

El texto definitivo sigue de cerca el planteamiento dual del Parlamento Europeo, calificando algunos usos de las herramientas de policía predictiva como prácticas prohibidas y otras como usos de alto riesgo. No obstante, seguramente por la divergencia entre pareceres entre el Consejo y el Parlamento, manifestados en las diferencias notables entre la posición común del Consejo y las enmiendas del Parlamento, el texto final también ha suavizado su respuesta frente a este uso de la IA.

El exponente más claro de ese «trade off» entre el parlamento y el Consejo es la descripción de la práctica prohibida referente al PersonBPP. Lo analizaremos a continuación distinguiendo entre las dos principales medidas que adopta finalmente el Reglamento en relación con los sistemas predictivos: a) la prohibición de algunos de ellos; b) la consideración de otras técnicas de pronóstico policial del delito como sistemas de alto riesgo.

2. SISTEMAS PREDICTIVOS POLICIALES PROHIBIDOS EN EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL

El artículo 5.1 d del RIA prohíbe «la introducción en el mercado, la puesta en servicio para este fin específico o el uso de un sistema de IA para realizar evaluaciones de riesgos de personas físicas con el fin de evaluar o predecir la probabilidad de que una persona física cometa una infracción penal basándose únicamente en la elaboración del perfil de una persona física o en la evaluación de los rasgos y características de su personalidad». La prohibición respeta en líneas generales el texto del Parlamento, pero con un matiz relevante como es que la evaluación del riesgo se base únicamente en el proceso de perfilado.

También es interesante apuntar que el objeto de la evaluación se ha acotado, refiriéndose exclusivamente a una infracción penal, excluyéndose infracciones administrativas. Por el contrario, la redacción actual admite que la evaluación tenga lugar en distintas fases del proceso y por distintos órganos, lo cual parece admitir supuestos de predictive justice o sentencing. Asimismo, se elimina la referencia a «la ubicación de la persona o las conductas delictivas pasadas de personas físicas o grupos de personas físicas» como ejemplos de las variables que se tienen en cuenta en la predicción. Esto no es problemático pues estas características pueden tenerse en cuenta en la elaboración de perfiles, que, concurriendo en el proceso predictivo, determina la aplicación de la prohibición; téngase que en cuenta que la elaboración de perfiles y la evaluación de sus rasgos y características de personalidad son requisitos alternativos y no cumulativos.

Como decíamos la prohibición abarca aquellas evaluaciones que se basan «únicamente» en la elaboración de perfiles. El RIA realiza una remisión a los conceptos de la normativa general de protección de datos para configurar la prohibición. Esta define la elaboración de perfiles como «toda forma de tratamiento automatizado de datos personales consistente en utilizar datos personales para evaluar determinados aspectos personales de una persona física, en particular para analizar o predecir aspectos relativos al rendimiento profesional, situación económica, salud, preferencias personales, intereses, fiabilidad, comportamiento, ubicación o movimientos de dicha persona física» (art. 4. 4) del RGPD). Tal como indica el Grupo de trabajo del artículo 29 deben darse tres requisitos para que

se dé la elaboración de perfiles: debe ser una forma automatizada de tratamiento; debe llevarse a cabo respecto a datos personales; y el objetivo de la elaboración de perfiles debe ser evaluar aspectos personales sobre una persona física⁵³. Teniendo en cuenta el anterior puede apreciarse la amplitud de la prohibición. Difícilmente puede alegarse que aquellos sistemas predictivos basados en las características de los sujetos (PersonBPP) no suponen la elaboración de perfiles: implican un tratamiento automatizado, se refieren a información de una persona física identificable, y su finalidad es evaluar aspectos personales del interesado (en este caso su peligrosidad)⁵⁴.

Una posible justificación de que el Reglamento abarque evaluaciones que se basen exclusivamente en la elaboración de perfiles la podemos encontrar en los considerandos: «en consonancia con la presunción de inocencia, las personas físicas en la UE deben ser juzgadas siempre por su comportamiento real». Pues bien «las personas físicas nunca deben ser juzgadas por su comportamiento previsto por la IA basado únicamente en su perfil, rasgos de personalidad o características, como nacionalidad, lugar de nacimiento, lugar de residencia, número de hijos, deudas, su tipo de coche, sin una sospecha razonable de que esa persona esté implicada en una actividad delictiva basada en hechos objetivos verificables y sin una evaluación humana de la misma» (considerando 42). Efectivamente, la elaboración de perfiles comporta el riesgo de que se realicen «correlaciones genéricas que pueden no ser correctas para todas las personas», tratándose a una persona como miembro de un grupo antes que como un individuo⁵⁵. En realidad, es inherente al perfilado una generalización en cuanto a que su misión es adscribir una persona a un perfil en cuya configuración nunca se encontrarán representados todos los rasgos relevantes del individuo⁵⁶. El perfilado presenta, por el contrario, beneficios sociales, en el sentido de que permite automatizar y aligerar nuestra comprensión del mundo, existiendo paralelismos entre la automatización algorítmica y una automatización biológica⁵⁷. Ahora bien, en ámbitos especialmente sensibles, donde es precisa una individualización de las decisiones, es lógico que el legislador vea su implementación. De hecho, en otros supuestos, también en la órbita de la aplicación de la ley penal, el legislador ha prohibido que el criterio único para adoptar decisiones vinculantes sea la adscripción a un perfil⁵⁸. El RIA se suma a ellos, entendiendo que la

53. Grupo de trabajo sobre protección de datos del artículo 29, *Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679*, 6 de febrero de 2018.

54. O., Lynskey, «Criminal justice profiling and EU data protection law: precarious protection from predictive policing», *International Journal of Law in Context*. 15(2), (2019), pp. 162-176.

55. FRA, *Guía para prevenir la elaboración ilícita de perfiles en la actualidad y en el futuro*, 2019, Disponible en: https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-preventing-unlawful-profiling-guide_es.pdf

56. Palma Ortigosa, A., *Decisiones automatizadas y protección de datos. Especial atención a los sistemas de inteligencia artificial*, Dykinson, Madrid, (2022).

57. Hildebrandt, M., «Defining Profiling: A New Type of Knowledge?», en Hildebrandt, M., y Gutwirth S., *Profiling the European citizen*, Springer, (2008).

58. Es el caso del Real Decreto 190/1996, de 9 de febrero, por el que se aprueba el Reglamento Penitenciario, que en su art. 6. 1 señala «Ninguna decisión de la Administración penitenciaria que implique la apreciación del comportamiento humano de los

valoración del riesgo de la comisión de una infracción penal realizada exclusivamente de forma automatizada y basada en el perfilado comporta riesgos inasumibles.

Aclarada la *ratio legis*, la segunda cuestión a analizar es cuándo debe entenderse que la evaluación se basa exclusivamente en el perfilado, presupuesto de la prohibición. De nuevo, parece indispensable acudir a la normativa en materia de protección de datos pues tanto el RGPD como la Directiva 2016/680 utilizan un concepto muy similar al proscribir aquellas decisiones «basadas únicamente» en un tratamiento automatizado de datos personales (arts. 22.1⁵⁹, y 11.1 respectivamente). Así, el RGPD postula que «todo interesado tendrá derecho a no ser objeto de una decisión basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzca efectos jurídicos en él o le afecte significativamente de modo similar». El alcance del art. 22 es ciertamente controvertido, siendo temerario postular una interpretación conclusiva del mismo. En este punto, simplemente nos acogemos a las directrices del grupo de trabajo del artículo 29, que en interpretación del art. 22 ha señalado que el hecho que una decisión esté basada únicamente en un tratamiento automatizado «quiere decir que no hay participación humana en el proceso de decisión». A ello añade que «para ser considerada como participación humana, el responsable del tratamiento debe garantizar que cualquier supervisión de la decisión sea significativa, en vez de ser únicamente un gesto simbólico» y «debe llevarse a cabo por parte de una persona autorizada y competente para modificar la decisión»⁶⁰.

Por tanto, de entre las diferentes formas de introducir la supervisión humana en el ciclo de vida del algoritmo, el RGPD en su art. 22.1 establece, en el corazón de la prohibición, que un sujeto con capacidad real para decidir pueda modificar la decisión automatizada (human in the loop)⁶¹. De forma muy similar, se acompaña al art. 5.1 d) del RIA una excepción a la prohibición de la policía predictiva, que se refiere a que «la prohibición no se aplicará a los sistemas de IA utilizados para apoyar la evaluación humana de la implicación de una persona en una actividad delictiva, que ya se basa en hechos objetivos y verificables directamente relacionados con una actividad delictiva». En realidad, este inciso segundo puede interpretarse como una aclaración más que una excepción. Nos indica cuando una decisión no se adopte exclusivamente teniendo en cuenta el perfilado: cuando la función del sistema sea apoyar la decisión humana y no sustituirla. Volviendo a la terminología del art. 22.1 del RGPD parece que lo que el RIA está proscribiendo son aquellas decisiones totalmente automatizadas sobre la evaluación del riesgo, mientras que en principio las decisiones parcialmente automatizadas son admisibles, a salvo de lo que se dirá a continuación.

reclusos podrá fundamentarse, exclusivamente, en un tratamiento automatizado de datos o informaciones que ofrezcan una definición del perfil o de la personalidad del interno».

59. En detalle, sobre el art. 22, la reciente STJUE de 7 de diciembre de 2023 (asunto C-634/21).

60. Grupo De Trabajo Sobre Protección De Datos Del Artículo 29, op. cit, p. 23.

61. En detalle Lazcoz, G., de Hert, P., «Humans in the GDPR and RIA governance of automated and algorithmic systems. Essential pre-requisites against abdicating responsibilities», *Computer Law & Security Review*, Volume 50, (2023).

3. POLICÍA PREDICTIVA «DE ALTO RIESGO» Y SUS IMPLICACIONES

En el Anexo III, se incluye en su apartado 6. d), a los «sistemas de IA destinados a ser utilizados por las autoridades relacionadas con la aplicación de la ley penal (law enforcement) o en su nombre o por las instituciones, agencias, oficinas u organismos de la Unión en apoyo de las autoridades policiales para evaluar el riesgo de una persona física de delinquir o reincidir que no se basen únicamente en la elaboración de perfiles de personas físicas a que se refiere el artículo 3, apartado 4, de la Directiva (UE) 2016/680 o para evaluar rasgos y características de la personalidad o comportamientos delictivos anteriores de personas físicas o grupos». La diferencia sustancial entre los supuestos de hecho a los que alude la norma es que, en el caso del uso de alto riesgo la evaluación «no se basa únicamente» en la elaboración de perfiles. Como puede entreverse, esta práctica se configura como una modalidad atenuada en términos de riesgo del uso prohibido: el perfilado se encuentra presente pero no es determinante en la evaluación. Eso no quiere decir que el sistema deje de ser influente (pensemos en el sesgo de automatización) y por tanto sus deficiencias sigan presentando riesgos que deban ser mitigados.

Aquí podemos apreciar una evolución respecto a la posición maniquea del Parlamento, en la que los riesgos en materia de PersonBPP no eran gestionados sino prohibidos. El texto final establece una respuesta proporcional al riesgo. Y es que, como adelantábamos, la presencia de un uso de alto riesgo con características muy similares dota de contenido al supuesto de hecho del uso prohibido: solo se encuentran proscritos aquellos usos en los que la evaluación no se ejecuta teniendo en cuenta exclusivamente medios automatizados.

La cuestión, *a priori* nada sencilla, será dilucidar en cada caso concreto si la implementación de un sistema de IA entra dentro de una u otra categoría. Y es que, en los términos defendidos, si el elemento clave es la función que ostenta la IA en la evaluación del riesgo, si esta es auxiliar o determinante, la *praxis* influirá notablemente en la calificación jurídica. Corresponderá al productor en la evaluación de riesgos (art. 9) contemplar específicamente que los implementadores del sistema lo utilicen en un sentido distinto al original, operando una delegación de funciones que, de haberse planteado en inicio, pudiera haber determinado la calificación de este como práctica prohibida.

Asimismo, la norma establece otros supuestos de IA de alto riesgo que pueden implicar actividades de policía predictiva. En primer lugar, «sistemas de IA destinados a ser utilizados por las autoridades policiales y judiciales o en su nombre, o por las instituciones, agencias, oficinas u organismos de la Unión en apoyo de las autoridades policiales y judiciales o en su nombre para evaluar el riesgo de que una persona física sea víctima de delitos». Por otro lado, «sistemas de IA destinados a ser utilizados por una autoridad judicial o en su nombre para asistir a una autoridad judicial en la investigación e interpretación de hechos y de la ley y en la aplicación de la ley a un conjunto concreto de hechos o utilizados de forma similar en la resolución alternativa de litigios». Esto lógicamente podría afectar a los sistemas de valoración del riesgo no encuadrables en el mencionado apartado 6. d), en ocasiones aplicándose a sistemas no estrictamente policiales sino judiciales (efectivamente, pensamos en el predictive sentencing).

Como adelantábamos, al utilizar el RIA un concepto tan amplio como el de elaboración de perfiles, es difícil que la mayoría de los usos de la policía predictiva, a menos que se llegue a la conclusión de que no son IA, escapen a los requisitos de la regulación. La notable excepción lo constituyen las herramientas predictivas basadas en el lugar. Estas se encontraban presentes tanto en la propuesta de la Comisión como en las enmiendas del Parlamento, calificadas como un uso de alto riesgo. Por el contrario, la posición común del Consejo las apartaba de su versión del Anexo III, suprimiendo el mencionado apartado 6. g): «sistemas de IA destinados a utilizarse para llevar a cabo análisis sobre infracciones penales en relación con personas físicas que permitan a las autoridades encargadas de la aplicación de la ley examinar grandes conjuntos de datos complejos vinculados y no vinculados, disponibles en diferentes fuentes o formatos, para detectar modelos desconocidos o descubrir relaciones ocultas en los datos». El texto final se alinea con el del Consejo, y suprime la letra g) del apartado 6.

Por tanto, las herramientas de Place BPP, configuradas para detectar en qué lugares y cuando un crimen tiene más posibilidades de llevarse a cabo no son consideradas sistemas de alto riesgo conforme al apartado 6 d). Y es que, según tal precepto el objeto de la evaluación debe recaer sobre personas físicas, característica que, por lo menos directamente, no se dan en estas herramientas. Ahora bien, es cierto que el mencionado apartado 6 d) parece mencionar dos usos distintos de la IA. Por un lado «para evaluar el riesgo de una persona física de delinquir o reincidir» y por otro, para «evaluar rasgos y características de la personalidad o comportamientos delictivos anteriores de personas físicas o grupos». La cuestión será determinar en qué medida un sistema que evalúa la posibilidad de que se den comportamientos delictivos en un punto (hot spot), está evaluando los rasgos y características de los sujetos o grupos que transitan esos lugares⁶².

4. CONCLUSIONES

En términos generales, el RIA ofrece una respuesta proporcional a los riesgos que presentan los sistemas de policía predictiva. No obstante, para fijar la peligrosidad de estos sistemas utiliza conceptos extraídos de la normativa de protección de datos como, el de la elaboración de perfiles. Esto conlleva varios problemas.

El primero es la exclusión de las herramientas encuadradas en el Place Based Predictive Policing. Como hemos apuntado previamente, estos sistemas también presentan riesgos éticos que, en el caso de un mal diseño o uso de los mismos, podría llevar a la posible afectación de derechos fundamentales. Y, sin embargo, quedan fuera de la regulación del Reglamento y lo más probable es que también de la normativa de protección de datos. Esto podría dejar a los ciudadanos europeos desprotegidos frente a estos sistemas, teniendo en cuenta además los límites que se impondrán a los Estados miembros en lo que sigue para regular la IA.

62. El FRA se mostró contrario a esta posibilidad, pues las predicciones realizadas por estos sistemas no incluyen datos personales sino «estadísticas agregadas». FRA, *Bias in algorithms artificial intelligence and discrimination*, 2022, Disponible en: https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf

El segundo problema es que el criterio para determinar la frontera entre una práctica prohibida y otra de alto riesgo parece enormemente influenciado por la *praxis*, y no puede determinarse en el momento del diseño del sistema. Efectivamente, lo que es adoptar un pronóstico de riesgo sobre la peligrosidad criminal basado «solamente» en la elaboración de perfiles, depende del papel del operador jurídico en la toma de la decisión. El análisis del contexto en el que se inserta el sistema no solo nos ayudará a determinar si «formalmente» se acepta la decisión del sistema como vinculante, sino también si, de facto, existe un normal proceder en el que el sistema no auxilia al implementador de la IA, sino que sustituye su decisión.

Finalmente, otra cuestión relevante es que el RIA no atiende, al menos indiciariamente, a la técnica computacional utilizada para determinar el nivel del riesgo (riesgo inadmisibles o alto riesgo). En otras palabras, si estamos ante una técnica actuarial clásica o un sistema basado en el aprendizaje automático parece resultar irrelevante. Como decimos, esto es problemático, en cuanto que los sistemas que participan de la segunda categoría conllevan riesgos éticos adicionales. Parece que, en este caso, el RIA vuelve a seguir demasiado de cerca al art. 22 del RGPD, cuyo supuesto de hecho se refiere a decisiones automatizadas, pero no necesariamente a las tomadas por IA⁶³.

Ahora bien, podría argumentarse por el contrario que este elemento, si bien no se introduce en la definición de las prácticas prohibidas, se deduce de la propia definición de Inteligencia Artificial que adopta el Reglamento. Es decir, solo quedarían sujetos al RIA aquellos sistemas que tengan capacidades de «inferencia», entre los cuáles los considerandos citan los sistemas de aprendizaje automático y los basados en enfoques de lógica y conocimiento, excluyéndose el «software tradicional basado en reglas definidas solo por humanos y que ejecuta automática operaciones» (considerando 12). Obviamente la relevancia de esta cuestión podría ser fundamental para resolver este problema interpretativo, puesto que de seguirse el concepto restringido al que parecen aludir algunos de los considerandos se estaría limitando notablemente las herramientas abarcadas y por tanto los objetos de prohibición y de regulación como de alto riesgo. Si pensamos en la práctica española no existe ningún sistema de Person Based Predictive Policing que haya sido entrenado a través del machine learning, lo cual no ha precluido ni que estos sistemas hayan sido usados y se hayan incluso institucionalizado, ni tampoco que no hayan sido analizados críticamente por los potenciales riesgos asociados a su uso⁶⁴. De modo que el operador jurídico se encontraría ante una decisión salomónica: o bien prescindir totalmente de la técnica computacional seguida para valorar el riesgo o, por el contrario, convertirla en el parámetro más relevante hasta el punto de que determine la sujeción global al Reglamento o la ausencia total de garantías frente a un sistema. Obviamente no nos compete a nosotros realizar un análisis exhaustivo de lo que es o no IA, pero sí nos

63. Palma Ortigosa, A., *Decisiones automatizadas y protección de datos. Especial atención a los sistemas de inteligencia artificial*, Dykinson, Madrid, (2022).

64. Véase López Riba, JM., «Inteligencia artificial y control policial. Cuestiones para un debate criminológico frente al hype», en prensa, (2024), también Martínez Garay, L., «Evidence-based sentencing y evidencia científica», en Miró Llinares, F., y Fuentes Ossorio, J. L., *El Derecho penal ante lo empírico. Sobre el acercamiento del Derecho penal y la Política Criminal a la realidad empírica*, Marcial Pons, Madrid, (2022).

podríamos atrever a realizar tres consideraciones que deben ser tomadas en cuenta a la hora de informar la interpretación para su aplicación a estos supuestos.

La primera, que el RIA tiene que interpretarse tomando en consideración todos los intereses en juego relacionados con el uso de estos sistemas, y no sólo los relativos a la innovación y al funcionamiento del mercado sino muy especialmente otros relacionados con la protección de los derechos fundamentales. La segunda, que no sólo aquellos sistemas que, de una forma más clara, entran dentro de la idea de lo que es IA pueden conllevar riesgos éticos y deben estar sometidos a evaluación. La tercera, que antes de regular y de interpretar los términos normativos de lo regulado es imprescindible comprender qué consecuencias conlleva ello en la práctica. Optar por una concepción amplia nos puede llevar a prohibir el uso de sistemas cuyos riesgos éticos en absoluto están proporcionalmente asociados con lo que supondría tal «sanción», con lo ilógico que resultaría que se prohibieran algunos sistemas que pueden requerir supervisión y control pero que, *per se*, no son peores, en términos no sólo de eficacia sino también de trazabilidad, transparencia y garantías, que modos clásicos de realizar la actividad policial. Pero usar una concepción demasiado restrictiva de lo que es IA podría desatender los riesgos asociados a tecnologías no tan desarrolladas computacionalmente pero igual o más peligrosas en otros aspectos. Si se sigue esa concepción deberíamos, entonces, ser capaces de encontrar otros modos de exigir normativamente a estos otros sistemas policiales predictivos aquello que, basado en el riesgo asociado al uso de tales sistemas, se está demandando también a lo que usa «machine learning».

El Reglamento ha avanzado notablemente en la regulación de la policía predictiva pero aún quedan en el tintero cuestiones interpretativas que determinarán en la práctica lo que es o no prohibido en el sector. Y para resolver esas cuestiones no podemos ser solo dependientes de lógicas previas, como la de protección de datos. El conocimiento empírico sobre los riesgos existentes deberá informar también nuestra interpretación de lo que está prohibido y no en el marco del policing, y porque no decirlo también: el «nebuloso» concepto de IA.

La aplicabilidad del Reglamento de inteligencia artificial al ámbito salud y especialidades respecto de su cumplimiento

INIGO DE MIGUEL BERIAIN

Ikerbasque research profesor. Investigador Universidad del País Vasco/Euskal Herriko Unibertsitatea. Miembro del Comité de Bioética de España

I. INTRODUCCIÓN

La aprobación del RIA¹ constituye un acontecimiento de especial relevancia en todo lo que se refiere a la regulación de los sistemas de este tipo en el ámbito de la UE, que vendrá a introducir seguridad jurídica en campos en los que, hasta ahora, apenas había referentes al respecto. No obstante, hay que recordar que el RIA ha sido diseñado con la vocación de constituir una normativa básica que, en determinados casos, ha de interpretarse de acuerdo con la regulación propia de sectores concretos. Esto es palmariamente cierto en los sistemas de IA que se utilicen con fines relacionados con la salud humana, a los que resulta aplicable con carácter general la normativa europea productos sanitarios a la que el RIA se remite expresamente. Dicha normativa incluye, al menos, el Reglamento 2017/745 sobre los productos sanitarios (MDR, abreviatura de sus siglas en inglés, en adelante) y el Reglamento 2017/746 sobre los productos sanitarios para diagnóstico in vitro (IVDR, en adelante), ambas normas considerablemente complejas.

En el presente capítulo analizaremos el marco jurídico que regulará las herramientas de IA con fines sanitarios, y más particularmente su calificación como sistemas de alto riesgo partiendo de lo dispuesto en la normativa que acabamos de citar. En todo caso, esperamos, al menos, ser capaces de ofrecer una descripción precisa de la clasificación de las herramientas de IA utilizadas en salud humana de acuerdo con el esquema basado en el riesgo implementado por el RIA, así como de las posibles dificultades que pueden llegar a presentar las diferencias de enfoque entre

1. El autor desea agradecer el apoyo recibido del Gobierno Vasco a través de las ayudas a grupos de investigación (GI CISJANT, ref. IT1541-22). Este trabajo se enmarca en el proyecto GODAS (Proyecto PID2022-137140OB-I00 financiado por el MCIN/AEI/10.13039/501100011033/FEDER, UE) y el Convenio entre la entidad pública RED.ES y un consorcio formado por seis entidades, incluyendo a la UPV/EHU, para impulsar la implementación de la carta de derechos digitales en el ámbito de entornos específicos.

este reglamento y los propios de los productos sanitarios que hemos citado. Para ello, y antes que nada, empezaremos por exponer las disposiciones al respecto del RIA.

II. REGULACIÓN DE LOS DISPOSITIVOS SANITARIOS EN EL REGLAMENTO Y SU CONSIDERACIÓN DE ALTO RIESGO POR EL ANEXO I O III

1. ANÁLISIS PRELIMINAR: LA REGULACIÓN DE LOS DISPOSITIVOS SANITARIOS QUE INCORPORAN INTELIGENCIA ARTIFICIAL EN EL REGLAMENTO

Como se ha dicho ya en otros capítulos de esta obra, el RIA se fundamenta sobre el concepto del riesgo: el grado de riesgo que implica el uso de un sistema será lo que determine los aspectos fundamentales de su estatuto jurídico. Entre otras cosas, dictaminará una cuestión capital: los requerimientos que los distintos agentes relacionados con el sistema (proveedores, distribuidores, importadores, etc.) habrán de cumplir antes y después de su introducción en la práctica sanitaria, así como el proceso de supervisión asociado a su aprobación. Por tanto, el problema esencial a dilucidar en este texto es el de cómo decidir si una IA que se asociará a fines sanitarios constituye o no un sistema de alto riesgo.

La respuesta a esta cuestión ha de partir del artículo 6 del RIA, que especifica las reglas de clasificación para los sistemas de IA: se los considerará de alto riesgo cuando por sus características sean susceptibles de ser englobados en la descripción que figura en el Anexo III de la norma o cuando reúnan dos condiciones: que vayan a utilizarse como componente de seguridad de uno de los productos contemplados en la legislación de armonización de la Unión que se indica en el Anexo I del RIA, o sean en sí mismos uno de dichos productos y que deban someterse a una evaluación de la conformidad realizada por un organismo independiente para su introducción en el mercado o puesta en servicio, de acuerdo con la normativa sectorial. En los siguientes apartados analizaremos ambas vías por separado.

2. SISTEMAS QUE SON DE ALTO RIESGO DE ACUERDO CON LO DISPUESTO EN EL ANEXO III

Acabamos de explicar que hay dos grandes vías por las que clasificar un sistema como de alto riesgo en el caso de los que se utilicen para fines sanitarios. Empezamos por analizar lo que se refiere a los sistemas que figuran en el Anexo III, punto 5, del Reglamento, que son los que la norma describe sin acudir a la remisión a la normativa sanitaria. Pues bien, de acuerdo con la propuesta formulada originalmente por la Comisión, serían productos de alto riesgo dos grandes clases de sistemas. En primer lugar, los «Sistemas de IA destinados a ser utilizados para la evaluación y la clasificación de las llamadas de emergencia realizadas por personas físicas o para el envío o el establecimiento de prioridades en el envío de servicios de primera intervención en situaciones de emergencia, por ejemplo, policía, bomberos y servicios de asistencia médica, y en sistemas de triaje de pacientes en el contexto de la asistencia sanitaria de urgencia» (Anexo III, punto 6, letra d). El Considerando 58 del texto consolidado explica bien la razón de esta decisión, diciendo, textualmente, que

«los sistemas de IA empleados para evaluar y clasificar llamadas de emergencia de personas físicas o el envío o el establecimiento de prioridades en el envío de servicios de primera intervención en situaciones de emergencia, en particular policía, bomberos y servicios de asistencia médica, así como sistemas de triaje de pacientes para la asistencia sanitaria de emergencia, también deben considerarse de alto riesgo, dado que adoptan decisiones en situaciones sumamente críticas para la vida y la salud de las personas y de sus bienes.» A este primer tipo de sistemas habría que añadir otro: los «Sistemas de IA destinados a ser utilizados por las autoridades públicas o en su nombre para evaluar la admisibilidad de las personas físicas para beneficiarse de servicios y prestaciones esenciales de asistencia pública, incluidos los servicios de asistencia sanitaria, así como para conceder, reducir o retirar dichos servicios y prestaciones o reclamar su devolución» (Anexo III, en este caso, punto 5.a).

La postura inicial de la Comisión fue objeto de respuestas alternativas en las versiones del Parlamento y del Consejo. En concreto, el Parlamento optó por acotar un tanto el rango de sistemas afectados, limitándolo a los sistemas que suponían un riesgo o daño significativo para la salud, la seguridad o los derechos fundamentales de las personas a la vez que, de otro lado, abogaba por incluir en la categoría de alto riesgo aquellos que suponían un riesgo o daño significativo para el medio ambiente. El Consejo, por su parte, propuso una redacción alternativa al artículo 6.2, que obviaba toda referencia al Anexo III. Ambas posiciones se desestimaron durante la negociación, quedando intacta la propuesta original de la Comisión, aunque las matizaciones del Parlamento se vieron de alguna forma reflejadas en la excepción al régimen general de los sistemas descritos en el Anexo III que analizaremos posteriormente.

No obstante, donde sí se introdujeron cambios durante la tramitación de la norma fue en el propio Anexo III. Frente a la redacción original de la Comisión, el Parlamento quiso introducir una letra adicional (ba). En su virtud, se considerarían de alto riesgo los sistemas de IA destinados a ser utilizados para tomar decisiones o influir materialmente en las decisiones elegibilidad de las personas físicas para los seguros de enfermedad y de vida. La versión final no recoge esta propuesta, sino una razonablemente similar presentada por el Consejo, conforme a la que serán de alto riesgo los sistemas de IA destinados a ser utilizados para la evaluación de riesgos y la fijación de precios en relación con las personas físicas y en los casos de los seguros de vida y de enfermedad (letra c bis)². La razón de la inclusión en la categoría alto riesgo de estos sistemas figura en el Considerando 58 de la versión final: «los sistemas de IA destinados a ser utilizados para la evaluación de riesgos y la fijación de precios en relación con las personas físicas en el caso de los seguros de vida y de salud también pueden afectar de un modo considerable a los medios de subsistencia de las personas y, si no se diseñan, desarrollan y utilizan debidamente, pueden vulnerar sus derechos fundamentales y pueden tener graves consecuencias para la vida y la salud de las personas, como la exclusión financiera y la discriminación».

A su vez, y en lo que respecta a la calificación de los sistemas de IA destinados a ser utilizados por las autoridades públicas o por un tercero en su nombre para

2. Quizás merezca la pena reseñar que la propuesta del Consejo incluía una excepción a esta norma general para sistemas desarrollados para su propio uso por proveedores que fueran pequeñas empresas que no prosperó.

evaluar la admisibilidad de las personas físicas para acceder a prestaciones y servicios de asistencia pública, así como para conceder, reducir, retirar o recuperar dichas prestaciones y servicios como de alto riesgo, la versión final aprobada recoge una redacción alternativa con una novedad importante: se reduce el rango de los sistemas de alto riesgo a los que se asocian a prestaciones y servicios de asistencia pública *esenciales*, lo que es, a nuestro juicio, una alternativa sensata, presente tanto en la versión del Parlamento como en la del Consejo, que evitará que ciertos sistemas que no alteran en exceso los bienes y derechos que se pretende proteger tengan que someterse a los requerimientos de los de alto riesgo. De otro lado, se ha mejorado la propuesta inicial manifestando explícitamente que en las prestaciones y servicios de asistencia pública se incluye la atención sanitaria, despejando así cualquier posible confusión al respecto.

Por fin, y con respecto al uso de sistemas de IA en el caso de situaciones de emergencia, hay que reseñar dos novedades importantes en el texto final del documento, respecto a lo que evidenciaba la propuesta de la Comisión. En primer lugar, se amplía el ámbito material de los sistemas calificados de alto riesgo, incluyéndose en ellos los destinados a evaluar y clasificar las llamadas de emergencia de personas físicas. Se incorpora así al texto una enmienda proveniente del Parlamento que está claramente enfocada a evitar que herramientas que pueden tomar decisiones de vital importancia tengan un sistema de supervisión adecuado. La segunda novedad es que se extiende también la previsión a los sistemas de triaje de pacientes. Esta previsión, de nuevo introducida por el Parlamento, se enfoca a las propuestas realizadas por algunos autores, de cara a introducir sistemas de IA en los triajes de emergencia³. En nuestra opinión, el fondo de la cuestión es más que razonable, si bien cabría objetar que probablemente no era necesario introducir esta coletilla, ya que parece obvio que tales sistemas ya pertenecen claramente a la categoría de los que se utilizarán para establecer prioridad en la provisión de cuidados y medicinas.

3. SISTEMAS QUE PUEDEN SER O NO DE ALTO RIESGO SEGÚN LO DISPUESTO EN EL ANEXO I

La segunda vía para considerar que un sistema de IA es de alto riesgo es la que incluye una remisión explícita a la legislación de armonización de la Unión listada en el Anexo I del Reglamento. El Consejo quiso simplificar la norma suprimiendo la remisión al Anexo e incluyendo, en cambio, una frase en la que aludía al requisito de tener que someterse a una declaración de conformidad como clave para decidir sobre el nivel de riesgo. Esta modificación no figura en el texto final, que recoge básicamente la propuesta de la Comisión. Hay, por tanto, que acudir al Anexo I, Sección A, del Reglamento para determinar qué sistemas de IA aplicados a la salud serán de alto riesgo. Pues bien, este incluye en su punto 11 una remisión expresa al

3. de Miguel Beriain, I. *The Ethical, Legal and Social Issues of Pandemics: An Analysis from the EU Perspective*. Springer, 2022; Weisberg EM, Chu LC, Fishman EK. The first use of artificial intelligence (AI) in the ER: triage not diagnosis. *Emerg Radiol*. 2020 Aug;27(4):361-366; Townsend BA, Plant KL, Hodge VJ, Ashaolu O, Calinescu R. Medical practitioner perspectives on AI in emergency triage. *Front Digit Health*. 2023 Dec 6;5:1297073.

«Reglamento (UE) 2017/745 del Parlamento Europeo y del Consejo, de 5 de abril de 2017, sobre los productos sanitarios, por el que se modifican la Directiva 2001/83/CE, el Reglamento (CE) n.º 178/2002 y el Reglamento (CE) n.º 1223/2009 y por el que se derogan las Directivas 90/385/CEE y 93/42/CEE del Consejo (DO L 117 de 5.5.2017, p. 1)» (MDR). Por su parte, el punto 12 del mismo Anexo I cita el Reglamento (UE) 2017/746 del Parlamento Europeo y del Consejo, de 5 de abril de 2017, sobre los productos sanitarios para diagnóstico in vitro y por el que se derogan la Directiva 98/79/CE y la Decisión 2010/227/UE de la Comisión (IVDR).

Hay que tener presente que esta remisión es crucial, por cuanto será en estos Reglamentos en los que encontremos respuesta a la cuestión de si un sistema de IA tendrá que someterse a una evaluación de la conformidad realizada por un organismo independiente para su introducción en el mercado o puesta en servicio, que es la verdaderamente capital para definir la calificación del sistema. De ahí que, de cara a establecer el estatuto jurídico de un sistema de IA sea necesario exponer el régimen creado por el MDR y por el IVDR, que es, precisamente, la tarea que abordaremos en el siguiente epígrafe.

No obstante, antes de entrar en ese análisis, se hace necesario abordar otro preliminar: la de qué sistemas de IA han de considerarse como productos sanitarios y cuáles no, ya que sólo si un sistema de IA es, efectivamente, producto sanitario, o constituye un componente de seguridad de uno de tales productos cobrará sentido determinar si ha de someterse a una evaluación de la conformidad realizada por un organismo independiente de acuerdo con los reglamentos sanitarios. En cambio, si consideramos que no es un producto sanitario, el análisis sobre el riesgo inherente a la herramienta de IA deberá efectuarse por otras vías, que quedan ahora fuera del ámbito de este capítulo. Explicada la trascendencia de este punto en concreto, procederemos inmediatamente a explicar el marco conceptual del MDR y el IVDR.

III. LA REGULACIÓN DE LOS PRODUCTOS SANITARIOS: LAS ESTIPULACIONES DEL MDR Y EL IVDR

1. LOS PRODUCTOS SANITARIOS. UNA CARACTERIZACIÓN

De acuerdo con el artículo 2(1) del MDR, se considera producto sanitario todo instrumento, dispositivo, equipo, programa informático, implante, reactivo, material u otro artículo destinado por el fabricante a ser aplicado en personas, por separado o en combinación con un producto sanitario, que no ejerce su acción principal prevista en el interior o en la superficie del cuerpo humano por mecanismos farmacológicos, inmunológicos ni metabólicos (pero a cuya función puedan contribuir tales mecanismos), y que se utilice con fines médicos específicos⁴. Lo primero de todo, por tanto, es que estemos —entre otras cosas— ante un programa informático o similar.

4. Entre los que el propio artículo describe los siguientes:
 — diagnóstico, prevención, seguimiento, predicción, pronóstico, tratamiento o alivio de una enfermedad,
 — diagnóstico, seguimiento, tratamiento, alivio o compensación de una lesión o de una discapacidad,
 — investigación, sustitución o modificación de la anatomía o de un proceso o estado fisiológico o patológico,

A su vez, el IVDR califica como un dispositivo de diagnóstico *in vitro* a «cualquier producto sanitario» que consista en un reactivo, producto reactivo, calibrador, material de control, kit, instrumento, aparato, pieza de equipo, programa informático o sistema, utilizado solo o en combinación, destinado por el fabricante a ser utilizado *in vitro* para el estudio de muestras procedentes del cuerpo humano, incluidas las donaciones de sangre y tejidos, única o principalmente con el fin de proporcionar información sobre uno o varios de los elementos siguientes:

- a) relativa a un proceso o estado fisiológico o patológico;
- b) relativa a deficiencias físicas o mentales congénitas;
- c) relativa a la predisposición a una dolencia o enfermedad;
- d) para determinar la seguridad y compatibilidad con posibles receptores;
- e) para predecir la respuesta o reacción al tratamiento;
- f) para establecer o supervisar las medidas terapéuticas.

Además, el IVDR aclara que como tales productos han de considerarse también los recipientes para muestras

Juntando las definiciones de ambos reglamentos tenemos, en suma, un catálogo bastante amplio de lo que son los productos sanitarios. No obstante, la lectura del MDR deja pendiente la delimitación de, al menos, dos cuestiones conceptuales. La primera es la que tiene que ver con la caracterización de los sistemas de IA como programas informáticos. La segunda, la que atañe a la noción de «fines médicos específicos».

En lo que se refiere a la primera, hay que recordar que los sistemas de IA son, de acuerdo con el artículo 3(1) del RIA, programas informáticos, es obvio que, si se utilizan para las finalidades que acabamos de precisar, han de considerarse como productos sanitarios y quedan, por consiguiente, sujetos a las disposiciones del MDR⁵. Por tanto, esta norma regirá sobre los sistemas de IA con independencia de si se trata de un programa ejecutable, un sitio web interactivo, un servicio web, un script o una simple macro en una hoja de cálculo. Además, la calificación como producto sanitario será independiente de si el procesamiento es simple o complejo, del riesgo que plantee el programa informático para el paciente o el usuario, de si lo utiliza un profesional sanitario o un profano, y de la plataforma informática en la que funcione, mientras esté destinado a utilizarse su uso con seres humanos o sus datos con fines médicos⁶.

Una vez aclarado este primer punto, centrémonos en el segundo: ¿qué significa exactamente «fin médico específico»? Aquí hay un intervalo e incertidumbre un poco mayor, por cuanto no todo sistema de IA que se emplee en el ámbito de la atención sanitaria será considerado producto sanitario. Así lo había señalado ya el TJUE en una sentencia referente a la Directiva 93/42, ahora derogada: «*Así pues,*

— obtención de información mediante el examen *in vitro* de muestras procedentes del cuerpo humano, incluyendo donaciones de órganos, sangre y tejidos.

5. Kiseleva, A. (2020). AI as a medical device: is it enough to ensure performance transparency and accountability?. *EPLR*, 4, 5.
6. Beckers, R., Kwade, Z., & Zanca, F. (2021). The EU medical device regulation: Implications for artificial intelligence-based medical device software in medical physics. *Physica Medica*, 83, 1-8.

el legislador ha dejado claro que, en relación con los programas informáticos, para que éstos queden comprendidos en el ámbito de aplicación de la Directiva 93/42, no basta con que se utilicen en un contexto sanitario, sino que es necesario que su finalidad, definida por su fabricante, sea específicamente médica»⁷. Siguiendo este criterio, el Grupo de Coordinación de Productos Sanitarios (MDGC)⁸ ha interpretado que no se deberían considerar productos sanitarios los programas informáticos que se limitan a tareas de almacenaje, archivado, comunicación o búsqueda, si no tienen finalidad médica. Esto incluye, por ejemplo, un programa dedicado a alteración de la representación de los datos para mejorar la calidad de su presentación o su compatibilidad. Tampoco, por supuesto, las herramientas que se utilicen para generar facturas u organizar a los trabajadores sanitarios. En cambio, sí deberían ser considerados productos sanitarios un programa que busque en una imagen hallazgos que apoyen una hipótesis clínica en cuanto al diagnóstico o la evolución de la terapia o que amplíe localmente el contraste del hallazgo en una pantalla de imagen para que sirva de apoyo a la decisión o sugiera una acción a realizar por el usuario⁹. También lo serán los productos de control o apoyo a la concepción o los productos destinados específicamente a la limpieza, desinfección o esterilización de los productos que se utilizan para los fines médicos descritos en el artículo 2(1) del MDR o incluidos en su Anexo XVI¹⁰.

Es importante, por fin, reseñar que sólo estaremos ante un producto sanitario si su propósito es beneficiar a pacientes concretos. Si, por el contrario, estamos ante programas que se destinan únicamente a agregar datos de población, proporcionar vías genéricas de diagnóstico o tratamiento genéricos (no dirigidos a pacientes individuales), mejorar la literatura científica, atlas médicos, o los modelos y plantillas, o de software destinado únicamente a estudios epidemiológicos o registros, no serán dispositivos sanitarios y quedarán, por tanto, fuera del marco del MDR.

2. CLASES DE PRODUCTOS SANITARIOS Y EXIGENCIAS DE SUPERVISIÓN INHERENTES A CADA TIPO SEGÚN EL MDR

Una vez expuestos los criterios generales que explicitan cuándo considerar un sistema de IA como un producto sanitario, es hora de centrarnos en la clase de producto del que se trata, lo que traerá importantes consecuencias de cara al proceso de aprobación y supervisión del dispositivo. En este apartado nos ocuparemos de los productos sanitarios,

7. Sentencia del Tribunal de Justicia (Sala Tercera) de 22 de noviembre de 2012, *Brain Products GmbH contra BioSemi VOF y otros*, Asunto C-219/11, par. 17.

8. El Grupo de Coordinación de Productos Sanitarios (MDGC) fue establecido por el artículo 103 del MDR. El MDGC está compuesto por representantes de todos los Estados miembros y está presidido por un representante de la Comisión Europea. Sus documentos no son documentos de la Comisión Europea y no pueden considerarse como reflejo de la posición oficial de la Comisión Europea. Tampoco son jurídicamente vinculantes (sólo el Tribunal de Justicia de la Unión Europea puede dar interpretaciones vinculantes del Derecho de la Unión), pero pueden servir como bases desde las que construir una aproximación a los conceptos contenidos en el Reglamento que creó el Grupo.

9. Grupo de Coordinación de Productos Sanitarios (MDGC), «Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 — MDR and Regulation (EU) 2017/746 — IVDR (MDGC 2019-11)», en:

10. Véase lo dispuesto en el artículo 2(1) del MDR.

dejando para el siguiente el análisis de lo que atañe al diagnóstico in vitro. Pues bien, la normativa sobre productos sanitarios establece que un programa informático sanitario puede ser clasificado en varias clases diferentes: I, IIa, IIb y III. Para determinar la clase a la que pertenecen los programas informáticos independientes de cualquier otro producto se ha de aplicar la Regla 11 del Anexo VIII del MDR, que señala lo siguiente:

Los programas informáticos destinados a proporcionar información que se utiliza para tomar decisiones con fines terapéuticos o de diagnóstico se clasifican en la clase IIa, salvo si estas decisiones tienen un impacto que pueda causar:

— *la muerte o un deterioro irreversible del estado de salud de una persona, en cuyo caso se clasifican en la clase III, o*

— *un deterioro grave del estado de salud de una persona o una intervención quirúrgica, en cuyo caso se clasifican en la clase IIb*

Los programas informáticos destinados a observar procesos fisiológicos se clasifican en la clase IIa, salvo si se destinan a observar parámetros fisiológicos vitales, cuando la índole de las variaciones de dichos parámetros sea tal que pudiera dar lugar a un peligro inmediato para el paciente, en cuyo caso se clasifican en la clase IIb.

Todos los demás programas informáticos se clasifican en la clase I.

Siguiendo esta norma, los programas informáticos utilizados para calcular dosis de fármacos de alta toxicidad, sugerir un diagnóstico o ayudar en la planificación de terapias o radiaciones pertenecerán a la clase III, ya que un error podría causar la muerte. Si es muy improbable que un error pueda causar la muerte, podría pertenecer a la clase IIb, mientras que sólo pueden pertenecer a la clase IIa aquellos en los que no puede preverse que un error cause un deterioro grave del estado de salud de una persona¹¹. Si el sistema tiene varios posibles usos, se considerará para su clasificación su utilización especificada más crítica¹².

La Guía del MDCG sobre calificación y clasificación de programas informáticos sanitarios, no obstante, parece atemperar un poco este marco. Así, por ejemplo, sugiere que el software destinado a clasificar sugerencias terapéuticas para un profesional sanitario basándose en el historial del paciente, los resultados de pruebas de imagen y las características del paciente debería clasificarse como clase IIa, aunque cabría interpretar que lo suyo sería incluirlo entre los de clase III, ya que un error podría causar la muerte del paciente. Algunos otros ejemplos que pueden ser de utilidad a la hora de interpretar el sistema establecido por el MDR¹³ son los siguientes:

— Un programa informático destinado a realizar diagnósticos mediante el análisis de imágenes para tomar decisiones de tratamiento de pacientes con ictus agudo debe clasificarse como clase III según la Regla 11(a).

11. Keutzer, L., & Simonsson, U. S. (2020). Medical device apps: an introduction to regulatory affairs for developers. *JMIR mHealth and uHealth*, 8(6), e17567.

12. AEMPS, Guía Para Fabricantes De Productos Sanitarios Clase I diciembre 2019, revisada en julio 2020, página 13, en: https://www.aemps.gob.es/productosSanitarios/docs/guia_fabricantes-ps.pdf

13. La traducción es de Guillermo Lazcoz Moratinos. Puede hallarse en: Lazcoz Moratinos, G. (2023). Gobernanza y supervisión humana de la toma de decisiones automatizada basada en la elaboración de perfiles, tesis doctoral, accesible en: <https://addi.ehu.es/handle/10810/61322>

— Un programa informático de diagnóstico destinado a puntuar la depresión basándose en los datos introducidos sobre los síntomas de un paciente (por ejemplo, el estado de ánimo, la ansiedad) deben clasificarse como clase IIb en virtud de la Regla 11(a).

— Un programa informático destinado a clasificar las sugerencias terapéuticas para un profesional de la salud sobre la base de la historia del paciente, los resultados de las pruebas de imagen y las características del paciente, por ejemplo, que enumera y clasifica todas las opciones de quimioterapia disponibles para las personas BRCA-positivas, debe clasificarse como clase IIa según la Regla 11(a).

— Una aplicación pretende ayudar a la concepción calculando el estado de fertilidad de la usuaria basándose en un algoritmo estadístico validado. La usuaria introduce datos de salud, como la temperatura corporal basal (TB) y los días de menstruación, para seguir y predecir la ovulación. El estado de fertilidad del día actual se refleja en uno de los tres indicadores luminosos: rojo (fértil), verde (infértil) o amarillo (fase de fluctuación del ciclo). Esta aplicación debe clasificarse como clase I según la Regla 11(c).

El tipo de calificación que una herramienta de IA obtenga dentro de esta clasificación, como hemos dicho, será el que determine, de acuerdo con el esquema del MDR, el tipo de evaluación clínica requerido para la certificación del producto (CE) o el seguimiento poscomercialización al que deberá someterse¹⁴, así como lo que obligue a la intervención de un tercero en el proceso. Y es que, como explica el Considerando 60 del MDR, *«los procedimientos de evaluación de la conformidad para los productos de la clase I deben llevarse a cabo, generalmente, bajo la exclusiva responsabilidad de los fabricantes, dado el bajo grado de vulnerabilidad asociado a estos productos. En el caso de los productos de las clases IIa, IIb y III, debe ser obligatorio un nivel apropiado de intervención de un organismo notificado»*.

No obstante, lo que ahora nos interesa es, exclusivamente, si esa intervención de un tercero es obligado o no, lo que sucede en los casos de los sistemas de IIa, IIb y III. También, por cierto, incluso en los de clase I, si tales productos se introducen en el mercado en condiciones estériles, tienen funciones de medición o son instrumentos quirúrgicos reutilizables¹⁵, aunque la participación del organismo notificado en tales casos se circunscribirá a verificar aspectos muy concretos de dichos productos. En este sentido, el MDR cambió sustancialmente el marco trazado por la anterior Directiva¹⁶, que dictaminaba que la mayoría de los programas informáticos independientes, incluidas las aplicaciones, se clasificaran

14. Como tal ha de entenderse «todas las actividades realizadas por los fabricantes en cooperación con otros agentes económicos para instaurar y actualizar un procedimiento sistemático destinado a recopilar y examinar de forma proactiva la experiencia obtenida con productos que introducen en el mercado, comercializan o ponen en servicio, con objeto de detectar la posible necesidad de aplicar inmediatamente cualquier tipo de medida correctiva o preventiva» (Considerando 60 MDR).

15. Véase: AEMPS, GUÍA PARA FABRICANTES DE PRODUCTOS SANITARIOS CLASE I, diciembre 2019 julio 2020 rev.1, p. 6, en: https://www.aemps.gob.es/productos-Sanitarios/docs/guia_fabricantes-ps.pdf

16. Medical Device Directive (MDD) 93/42/EEC.

como clase I o no se designaran como productos sanitarios en absoluto¹⁷. Por fin, hay que subrayar que la Guía Grupo de Coordinación de Productos Sanitarios¹⁸ aclara que, en caso de que suceda cualquier cambio tanto en la finalidad prevista, como en el contexto/situación de la atención clínica en la que se utiliza ese mismo producto, podría alterarse la calificación, sustituyéndose la actual por una clase de riesgo diferente.

3. EL IVDR Y LOS DISPOSITIVOS DE DIAGNÓSTICO IN VITRO

¿Qué decir a su vez de los dispositivos de diagnóstico in vitro? En este caso es necesario referirse al IVDR, que utiliza un sistema relativamente parecido al del MDR, solo que en este caso los dispositivos vienen a resultar divididos en cuatro clases, A, B, C y D, que se establecen teniendo en cuenta la finalidad prevista de los productos y sus riesgos inherentes. La clasificación se llevará a cabo de conformidad con el Anexo VIII del Reglamento, que incluye siete reglas de calificación de los dispositivos, de cierta complejidad técnica. La regla general es que el orden de riesgo es incremental, asignándose la Clase A a dispositivos de bajo riesgo y la Clase D para aquellos dispositivos que representan el mayor riesgo. Pues bien, la aplicación del sistema de clasificación de riesgos del Reglamento IVDR obliga a la participación de un Organismo Notificado para la aprobación de todos los dispositivos no estériles excepto para los de Clase A. Teniendo esto presente, se estima que el 90% de los dispositivos IVD estarán sujetos a la revisión de un Organismo Notificado, en comparación con el 15% que debían cumplir este requisito con la directiva anterior¹⁹.

4. LAS EXCEPCIONES AL RÉGIMEN GENERAL DE LOS SISTEMAS INCLUIDOS EN EL ANEXO III

A partir de lo explicado en el apartado anterior parece inevitable concluir que habrá muchos productos sanitarios y dispositivos de diagnóstico in vitro susceptibles de ser incluidos en alguna de las categorías que necesitan de supervisión a través de un organismo notificado. Esto, teniendo en cuenta que es el mero hecho de la intervención de un organismo notificado lo esencial a la hora de determinar si un sistema de IA que se aplica en el ámbito sanitario es o no de alto riesgo, significaría que habría muchos de estos sistemas que se considerasen como de alto riesgo, ya que muy pocos de ellos son susceptibles de ser clasificados como de clase I en el esquema

17. Keutzer, L., & Simonsson, U. S. (2020). Medical device apps: an introduction to regulatory affairs for developers. *JMIR mHealth and uHealth*, 8(6), e17567.

18. Grupo de Coordinación de Productos Sanitarios (MDGC), «Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 — MDR and Regulation (EU) 2017/746 — IVDR (MDCG 2019-11)».

19. <https://www.tuvsud.com/es-es/industrias/asistencia-sanitaria-productos-sanitarios/diagnostico-in-vitro/aprobacion-certificacion-mercado/reglamento-ue-productos-sanitarios-diagnostico-in-vitro>

BSI: IVDR Conformity Assessment Routes Notified Body Assessments, en: <https://www.bsigroup.com/globalassets/meddev/localfiles/en-gb/documents/bsi-md-ivdr-conformity-assessment-routes-booklet-uk-en.pdf>

del MDR (e incluso puede haber excepciones en la clase I, como hemos indicado)²⁰ o de clase A en el del IVDR.

No obstante, hay una excepción a esta norma general, debido al resultado de la negociación entre las tres instituciones europeas. El cambio esencial operado en la versión final del Reglamento frente a lo que estipulaba la propuesta original de la Comisión es que contempla la posibilidad de que alguno de los sistemas incluidos en el Anexo III podrán evitar dicha calificación bajo dos condiciones: que el implementador sea un organismo de derecho público o un operador privado que preste servicios públicos; y que se cumpla la condición incluida ahora en el artículo 6.3: que el sistema no suponga un riesgo significativo para la salud, la seguridad o los derechos fundamentales de las personas, lo que incluye que no influyan materialmente en el resultado de la toma de decisiones²¹. Esto, aclara el artículo, ocurrirá cuando si concurre alguna de las siguientes circunstancias:

Que el sistema de IA está destinado a realizar una tarea procedimental limitada;

— que el sistema de IA tenga por objeto llevar a cabo una tarea de procedimiento limitada;

— que el sistema de IA tenga por objeto mejorar el resultado de una actividad humana previamente realizada;

— que el sistema de IA tenga por objeto detectar patrones de toma de decisiones o desviaciones con respecto a patrones de toma de decisiones anteriores y no esté destinado a sustituir la evaluación humana previamente realizada sin una revisión humana adecuada, ni a influir en ella; o d) que el sistema de IA tenga por objeto llevar a cabo una tarea preparatoria para una evaluación pertinente a efectos de los casos de uso enumerados en el Anexo III.

De acuerdo con el artículo 6.2.4, será el proveedor quien determine si concurre o no alguna de estas circunstancias. Y en caso de que como resultado de su evaluación considere que un sistema de IA de los contemplados en el Anexo III no es de alto riesgo deberá documentar su evaluación antes de que dicho sistema se comercialice o se ponga en servicio. A petición de las autoridades nacionales competentes, le facilitará la documentación de la evaluación.

Obviamente, el recurso escogido para evitar la calificación de alto riesgo —la inclusión de estos criterios concretos— posee inconvenientes fáciles de intuir: cabe que la tecnología cambie sustancialmente y los criterios enunciados no respondan a las necesidades del momento; es posible que la práctica revele que puede haber otros que deban añadirse a la lista; puede ser complejo entender qué realidades comprenden en la práctica, etc. De ahí que en el articulado final del Reglamento se hayan incluido una serie de previsiones encaminadas a afrontar estas cuestiones. Así, en primer lugar, de acuerdo con el artículo 6.2.6, la Comisión estará facultada para adoptar actos delegados con arreglo al artículo 97 de la norma a fin de modificar los

20. Grzybowski, A., & Brona, P. (2023). Approval and Certification of Ophthalmic AI Devices in the European Union. *Ophthalmology and Therapy*, 12(2), 633-638.

21. Este resultado responde adecuadamente a la voluntad, expresada en el Considerando 29, de limitar la calificación de alto riesgo a los sistemas de IA identificados que tengan realmente un impacto perjudicial en la salud, la seguridad y los derechos fundamentales de las personas en la Unión y dicha limitación minimice.

criterios que acabamos de reflejar, ya sea añadiendo otros nuevos o modificando los que ya existen, siempre que existan pruebas concretas y fiables de la existencia de sistemas de IA que entren en el ámbito de aplicación del Anexo III pero que no supongan un riesgo significativo de perjuicio para la salud, la seguridad y los derechos fundamentales. También podrá suprimir —siempre por actos delegados— alguno de los criterios establecidos en el apartado 3 del artículo 6 cuando existan pruebas concretas y fiables de que ello es necesario para mantener el nivel de protección de la salud, la seguridad y los derechos fundamentales en la Unión. En todo caso, será necesario garantizar que las modificaciones no causarán una disminución en el nivel global de protección de la salud, la seguridad y los derechos fundamentales en la Unión.

De otro lado, el artículo 6.2.5 incluye la previsión de que la Comisión, previa consulta al Comité Europeo de IA, y a más tardar dieciocho meses después de la entrada en vigor del RIA, proporcione directrices que especifiquen la aplicación práctica del artículo 6, incluyendo una lista exhaustiva de ejemplos prácticos de casos de uso de alto y no alto riesgo en sistemas de IA, de conformidad con el artículo 96. Es de esperar que en esta lista de casos se incluyan muchos relacionados con el uso de sistemas de IA en el contexto de la atención sanitaria o la salud pública.

Es de otro lado preciso hacer notar que el artículo 6 impone una serie de obligaciones para el proveedor. De este modo, si considera que un sistema de IA contemplado en el Anexo III no es de alto riesgo deberá documentar su evaluación antes de que dicho sistema se comercialice o se ponga en servicio. También estará sujeto a la obligación de registro establecida en el apartado 1 bis del artículo 49.2 del Reglamento. A petición de las autoridades nacionales competentes, el proveedor facilitará la documentación de la evaluación.

IV. RECAPITULACIÓN

La mejor forma de resumir todo lo dicho en este capítulo sería, seguramente, incidir en el hecho de que hay dos grandes grupos de sistemas de IA de uso sanitario que deberán considerarse de alto riesgo: los que son susceptibles de ser utilizados para algunos de los fines descritos en el Anexo III del Reglamento sin que el proveedor pueda aducir la concurrencia de alguna de las circunstancias descritas en el artículo 6.3 para evitar la calificación de «alto riesgo» o los que, siendo medicamentos, productos o dispositivos sanitarios *in vitro*, necesitan de una evaluación de la conformidad por terceros.

El sistema parece, en suma, exigente, en lo que se refiere a la determinación del riesgo, e incluso disfuncional, dadas las diferencias conceptuales entre las normas aplicables. Puede haber, de hecho, contraste entre la medición del riesgo del RIA y el de los reglamentos sanitarios. Puede suceder que una herramienta sea evaluada por un organismo notificado de los previstos en el MDR y el IVDR con un nivel de riesgo medio (clase IIa o B, por ejemplo) y, sin embargo, sea considerada de «alto riesgo» por el RIA, ya que, en el caso de esta última norma, el requisito para obtener tal calificación es precisamente, que deba someterse a esa supervisión. Por esto mismo, cobra perfecto sentido haber incluido en el Considerando 51 la idea de «que un sistema de IA se clasifique como de alto riesgo en virtud del presente Reglamento no significa necesariamente que el producto del que sea componente de seguridad,

o el propio sistema de IA como producto, se considere de “alto riesgo” con arreglo a los criterios establecidos en la correspondiente legislación de armonización de la Unión que se aplique al producto. Tal es el caso, en particular, del Reglamento (UE) 2017/745 del Parlamento Europeo y del Consejo y del Reglamento (UE) 2017/746 del Parlamento Europeo y del Consejo, que prevén que un organismo independiente realice una evaluación de la conformidad de los productos de riesgo medio y alto». El sentido de «alto riesgo» puede coincidir o no en el caso de unas normas y otras, luego un sistema de IA puede ser de alto riesgo en los términos del RIA y no serlo en los del MDR, por ejemplo.

La aplicabilidad del Reglamento de inteligencia artificial al ámbito de la Administración pública y servicios públicos y especialidades respecto de su cumplimiento: Especial atención a Anexo III y actuación administrativa y particularidades cumplimiento

GAL·LA BARRACHINA NAVARRO
Universitat de València

ANDRÉS BOIX PALOP
Profesor Titular Derecho Administrativo Universitat de València¹

I. INTRODUCCIÓN: EL PLANTEAMIENTO DEL REGLAMENTO Y LA PROYECCIÓN DE SUS CONTROLES Y GARANTÍAS SOBRE LA ACTUACIÓN DE LOS PODERES PÚBLICOS

1. PANORÁMICA GENERAL: ORIENTACIÓN BÁSICA Y APLICACIÓN A LA ACCIÓN DE LOS PODERES PÚBLICOS DEL REGLAMENTO

Como es sabido, y en esta obra se ha desarrollado y expuesto de manera más completa en otras partes de este comentario, el RIA no está pensando ni definido en términos jurídicos, de manera específica, para ser aplicado a la actuación de los poderes públicos (incluyendo al poder judicial) en general ni, más en concreto, de las Administraciones públicas, ni de la propia Unión Europea ni de sus Estados Miembros. Se trata de una norma que, aplicando las enseñanzas de décadas de control público sobre las exigencias de seguridad y control respecto de la puesta en el

1. El presente estudio es resultado de la investigación llevada a cabo en el marco de los siguientes proyectos: Proyecto MICINN «Registro público de algoritmos» (PDC2022-133890-I00); Proyecto «Algorithmical law» (Prometeo/2021/009, 2021-24 Generalitat Valenciana); y Convenio de Derechos Digitales-SEDIA Ámbito 5 (2023/C046/00228673).

mercado de productos (o, aunque con menos incidencia, de la prestación de servicios que puedan conllevar también problemas ambientales o de seguridad), establece una serie de protocolos y exigencias típicos de este ámbito. Así, muy rápidamente, junto al establecimiento de una serie de usos de la IA que se consideran prohibidos en todo caso en cuanto a los servicios o productos que podrían desplegarse con ciertas finalidades (Capítulo II, prácticas de inteligencia artificial prohibidas; art. 5 RIA), y que en todo caso cuentan siempre con ciertas excepciones (y que, aunque en este punto nos remitimos al comentario respecto de las políticas prohibidas, hay que señalar que suponen limitaciones tanto para el sector público como para el privado, como luego veremos, aunque la lógica sigue sin ser tanto poner el foco en el sector público como en los riesgos intrínsecos a ciertos usos de la IA), la norma pasa a delimitar aquellos usos que derivarán en una mayor exigencia de regulación y cumplimiento y por ello en un mayor control jurídico (art. 6 y Anexo III RIA), usos que en ningún caso son definidos en atención al empleo que puedan realizar los poderes públicos respecto de sistemas de IA, pues no es el foco, como se ha dicho del RIA.

Con todo, y como veremos posteriormente, hay algunos usos que, por supuesto, impactan de lleno en la esfera de actuación y posibles usos de IA que puedan hacer los poderes públicos, y estos usos van por ello a quedar regulados también por el Reglamento, desplegando obligaciones que van a impactar, también, sobre los poderes públicos. Un poco a la manera de lo ocurrido también con la normativa en materia de protección de datos (RGPD y la correspondiente normativa de transposición a cada país europeo), que aun no siendo normas específicamente prevista para disciplinar las actuaciones de los poderes públicos, sino esencialmente a los operadores económicos y agentes y sujetos que actúan en el mercado, han acabado también disciplinando a los poderes públicos².

Además, en el proceso de concreción y pulido del texto legal se han ido introduciendo en el Anexo III toda una serie de usos de la IA que se consideran de alto riesgo que impactan de modo más directo sobre la actividad de los poderes públicos (como veremos posteriormente, puede considerarse que en la actualidad cualquier actuación de cualquier poder público, administrativo o judicial, que se apoye en el uso de IA para la toma de decisiones, o ayudar o condicionar la misma, que impacte sobre la esfera de derechos y obligaciones de los ciudadanos, por defecto, va a tener siempre esta consideración, por mucho que no sea éste el objetivo esencial de la norma, sino una mera consecuencia, por lo demás afortunada, indirecta de la voluntad del RIA de establecer controles para los operadores privados y los usos de mercado de la IA³).

El modelo de regulación y cumplimiento de mayor relieve (y que más impacto tiene sobre el despliegue efectivo y la manera de hacerlo de herramientas de IA a

2. Palma Ortigosa, A., «Decisiones automatizadas en el RGPD. El uso de algoritmos en el contexto de la protección de datos», *Revista General de Derecho Administrativo*, 50, 2019.
3. Boix Palop, A., «Los algoritmos son reglamentos: la necesidad de extender las garantías propias de las normas reglamentarias a los programas empleados por la administración para la adopción de decisiones», *Revista de Derecho Público: Teoría y Método*, 1, 2020.

día de hoy y en el futuro) lo tenemos pues en el Capítulo III RIA sobre los sistemas de IA definidos como de alto riesgo, donde tras la referida delimitación (art. 6 y Anexo III RIA), su Sección Segunda pasa a establecer una serie de requisitos que han de cumplir los mismos, de tipo técnico (gestión de riesgos, de los datos, de llevanza de documentación, de información y transparencia, de robustez y seguridad de los sistemas). Como puede comprobarse, ninguna de estas exigencias, de nuevo, está pensada para los poderes públicos, pero acabarán debiendo ser cumplidas también por éstos cuando empleen IA para el ejercicio de funciones públicas. Significativamente, la Sección Tercera de este Capítulo Segundo define las obligaciones de los distintos proveedores e implantadores de estos sistemas, a efectos de garantizar el cumplimiento de las exigencias y requisitos jurídicos fijados por el RIA para poder poner en el mercado productos o proporcionar servicios que integren el empleo de IA, y lo hace tratando de abarcar toda la cadena de valor de modo que siempre haya un responsable por la puesta en el mercado o prestación del servicio en el ámbito europeo al que hacer responsable frente a posibles daños e incumplimientos, así como a efectos de definir las obligaciones concretas (y entre ellos) de cada uno de estos agentes.

En este sentido, de nuevo, es fácil proyectar sobre los poderes públicos estas reglas cuando estén en cada una de las diferentes posiciones definidas, aunque en la acción pública no vaya a ser frecuente que en la acción judicial o administrativa las herramientas de IA o productos o servicios públicos que la incorporen vayan a ser puestos luego en el mercado para que otros los puedan comercializar e integrar en cadenas productivas, por lo que su posición será normalmente siempre la de responsable en tanto que utilizador final (que además puede haber sido quien haya definido y desarrollado el uso o empleo de la IA concreta). En todo caso, no es llamativa, por ser la tónica habitual, de una previsión específica en el RIA para esta posición peculiar de los poderes públicos, pero sí conviene mencionarla una vez más: el tipo de obligaciones y exigencias será para los poderes públicos equivalentes a las de un agente privado que estuviera en una posición equivalente respecto de un concreto sistema de IA del que puedan estar aquéllos.

Una vez definido en estos términos el régimen de requisitos, exigencias y obligaciones, las Secciones Cuarta y Quinta del RIA trasladan a este sector un modelo de control propio de los sistemas de normas técnicas públicas obligatorias y de normalización industrial de tipo voluntario (en un juego explicado con carácter general por Álvarez García⁴ y que además se concreta también en el Capítulo X RIA con la regulación sobre códigos de conducta) pensados para la garantía de la seguridad industrial y protección de los consumidores, así como para facilitar la aparición de normas de autorregulación y de sistemas de armonización en los mercados, en ocasiones facilitada por los poderes públicos (autorregulación regulada⁵) que se apoya en la amplia experiencia previa. Así, la Sección Cuarta proyecta sobre el ámbito de la IA el típico modelo de control a partir de entes privados

4. Álvarez García, V., «Los instrumentos normativos reguladores de las especificaciones técnicas en la Unión Europea: un breve ensayo de identificación de nuevas fuentes del Derecho», *Revista General de Derecho Administrativo*, 64, 2023.
5. Darnaculleta i Gardella, M. M., «La autorregulación y sus fórmulas como instrumentos de regulación de la economía», *Revista General de Derecho Administrativo*, 20, 2019.

que serán los que esencialmente desarrollen las actuaciones de control, verificación y certificación del cumplimiento (organismos notificados), aunque obviamente cumpliendo siempre con una serie de exigencias y controles de tipo público que garanticen que realizan correctamente su labor y sobre los que reposa el grueso del modelo de control de cumplimiento, a fin de garantizar su rapidez, eficacia en términos de mercado y capacidad de adaptación. Junto a ellos, obviamente, aparecen unas autoridades, éstas sí, públicas, de notificación que han de velar por el cumplimiento de la normativa por parte de los agentes privados, verificar que cumplen correctamente con su cometido y que poseen la capacidad técnica para ello, así como intervenir en casos de detección de incumplimientos graves, en funciones más clásicas de control y verificación administrativa respecto de la actuación de agentes privados (típicas en cualquier mercado regulado y manifestación de la más básica policía administrativa).

A partir de este entramado, la Sección Quinta define en qué términos ha de realizarse la evaluación de conformidad, verificada en principio por esos organismos notificados, de los productos y servicios que incorporen IA, lo que lleva, como en cualquier mercado, a la obtención de unos certificados y etiquetas (que se concretan en el marco CE, art. 48 RIA) y su traslación a unos registros a efectos de control que son los que permitirán poner en el mercado los productos, pero, significativamente, esta evaluación de conformidad cuando sean sistemas empleados por poderes públicos será realizada internamente, sin necesidad de recurrir a controles externos. De nuevo, todo el sistema se define desde esta óptica tradicional de mercado y sin pensar en normas más exigentes o diferentes para los poderes públicos y las Administraciones públicas atendida su peculiar posición y su capacidad de poder lesionar en mayor medida los derechos de los ciudadanos y ciudadanas. Es más, cuando hay excepciones o singularidades, éstas lo son en términos de deferencia (el art. 111.2 RIA permite diferir la entrada en vigor de las exigencias del texto hasta seis años cuando son empleados por los poderes públicos, por ejemplo; como hemos señalado la evaluación de conformidad para estos casos es interna, por otro).

En todo caso, como es lógico, estas normas, en cuanto a su contenido material, sí deberán ser en todo caso también cumplidas por éstos y éstas cuando empleen en el futuro sistemas de IA para realizar sus funciones u ofrecer servicios, de modo que las herramientas de IA que empleen y que entren dentro del ámbito de aplicación definido por el art. 6 y el anexo III (que ya hemos dicho que serán prácticamente todas las que puedan usar los poderes públicos), al margen de controles de tipo públicos, europeos o estatales, que puedan definirse, deberán someterse a estos controles respecto de su cumplimiento, e incorporar el marcado CE o equivalente para el sector público, así como quedarán sometidos al control e inspección para casos que puedan suponer más riesgos por parte de las autoridades administrativas de control. Como, por otro lado, ocurre con cualquier administración pública cuando hace uso de un servicio o producto puesto en el mercado que ha de cumplir con normas técnicas y que, por supuesto, sólo podrá usar o integrar en su actuación y servicios si ha pasado por los correspondientes controles de forma satisfactoria.

Los Capítulos IV y V RIA, en la medida en que incorporan obligaciones específicas para ciertos usos de IA específicos, ya sean *chatbots*, ya modelos de IA de propósito general, no proyectan grandes cuestiones sobre el sector público, más allá de que, si se emplean por éste sistemas de IA de estos tipos, habrán de atenderse a estas

reglas, pero sin que haya ninguna especificidad digna de mención que se proyecte sobre el sector público. Mucha más incidencia sobre el sector público tienen los Capítulos VI, VII y VIII RIA, en la medida en que establecen respectivamente medidas de fomento (apoyo a la innovación en el sector), de gobernanza pública (con el despliegue de autoridades nacionales y europeas de control sobre el sector) que se han reforzado a medida que avanzaba la negociación del texto legal y sobre los registros públicos que, por supuesto, sí conforman núcleos claros de acción pública. Sin embargo, aunque en estos casos hay una acción administrativa clara sobre el sector, se alejan de nuestro objeto de interés, que no es tanto cómo han de actuar las Administraciones y poderes públicos para fomentar, controlar o garantizar el correcto funcionamiento de las normas de IA sobre el mercado de la IA o de los productos y servicios que la incorporan como respecto de las normas que han de cumplir los poderes públicos cuando son ellos los que la emplean. Lo mismo puede decirse del capítulo XI sobre funcionamiento de los comités o las medidas de organización de la potestad sancionadora del Capítulo XII.

Sí han de señalarse, en cambio, que algunas medidas del Capítulo IX sobre controles posteriores a la comercialización sí van a proyectarse también sobre los usos de IA realizados por administraciones y poderes públicos, por cuanto serán usos que normalmente se prolongarán en el tiempo. Estas obligaciones suponen la necesidad de controles sobre su uso y ejecución, que obligan a notificar incidentes graves y a evaluaciones de riesgo (art. 79 RIA) y notificación específica de problemas respecto de usos de IA de alto riesgo (art. 82 RIA) a los que por supuesto los poderes públicos van a quedar vinculados. Por supuesto, respecto de incumplimiento en este sentido, el Capítulo XII establece un régimen sancionador, de nuevo pensado para el mercado y las empresas más que para los poderes públicos (como muestra más clara de ello las sanciones se definen a partir de la cifra de negocios de la empresa considerada responsable), que habrá que ajustar, como se ha hecho en materia de protección de datos, a las especificidades de los poderes públicos.

En todo caso, y como podemos ver, estamos ante un régimen jurídico que no se ha definido ni desarrollado pensando en los poderes públicos, por lo que habrá de verse complementado con los sistemas de control nacionales ya existentes en la materia para la regulación de actuaciones administrativas automatizadas o que empleen algoritmos o IA de cada Estado Miembro, a la espera de que pueda haber alguna norma europea armonizadora (la experiencia en la materia en protección de datos permite anticipar que si éstas llegan no serán muy ambiciosas, para dejar espacio competencial a la autoorganización administrativa de los poderes públicos interna), de manera que en esta caracterización general del funcionamiento de estas reglas y para entender cómo se proyectarán hay que repasar someramente el modelo de control actual de Derecho interno en los distintos Estados Miembros y también en España. Adicionalmente, y como se ha indicado, el art. 111.2 RIA permite a los Estados miembros diferir hasta en 6 años la aplicabilidad de los requisitos y obligaciones desplegadas por la norma a los algoritmos y soluciones de IA que vayan a ser empleados por el sector público, lo que da la sensación de que, en última instancia (máxime dada la previsible evolución rápida del sector y de su marco normativo), convierte todas las normas que a continuación detallaremos más en una especie de directrices que posteriormente los Estados Miembros han de decidir o no cómo proyectar sobre sus poderes públicos que en normas imperativas. En definitiva,

y como venimos diciendo, es manifiesto que no nos encontramos ante un reglamento primariamente pensado para su aplicación a las soluciones de IA empleadas por los poderes públicos sino que sólo las afectarán de manera indirecta y de una manera muy dependiente del modo, la forma y los tiempos que los propios poderes públicos estimen que les resultan más convenientes.

2. INTEGRACIÓN DEL CONTROL DE LA ACTIVIDAD AUTOMATIZADA Y ALGORÍTMICA, Y DEL USO DE INTELIGENCIA ARTIFICIAL, POR PARTE DE LOS PODERES PÚBLICOS CON LAS NORMAS DE DERECHO INTERNO Y ALGUNOS DE SUS PROBLEMAS Y CARENCIAS

Tanto en nuestro sistema nacional, como en otros estados miembros, pueden los poderes públicos llevar a cabo parte de su actividad de forma automatizada y sin intervenir decisión humana, posibilitándose el recurso a la IA (art. 41 de la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público, RJSP). El empleo por parte de los poderes públicos de sistemas basados en IA puede aportar beneficios evidentes, relativos sobre todo a la eficacia y eficiencia del sistema en su funcionamiento, o a la adopción de decisiones discrecionales dotadas de cierta imparcialidad y objetividad, no obstante, dicho avance conlleva riesgos, posibles confrontaciones con los derechos fundamentales, libertades y garantías que la administración tiene en deber de proteger. Una normativa que proporcione seguridad jurídica y asegure que los administrados no verán afectas sus garantías es esencial, pues la optimización del sistema es deseable, pero la protección de los derechos es ineludible. Encontrar una regulación con un marco normativo que permita el avance, pero que a su vez asegure el estatuto jurídico del administrado es una tarea que requiere llegar a un delicado equilibrio, no muy fácil de conseguir cuando existen intereses contrapuestos.

La sustitución de la inteligencia humana por la artificial puede realizarse por dos vías: la decisión adoptada directamente por el algoritmo y sin intervenir la inteligencia humana, y aquella en la que media intervención humana, aunque de manera secundaria, al servicio de la IA⁶. Dentro de estos dos grupos, hay doctrina que considera esencial excluir la aplicación de la IA en determinados supuestos, como en la toma de decisiones discrecionales, siguiendo la solución alemana⁷ mientras que otros, en cambio sí abogan por su uso en aquellas decisiones con un «bajo nivel de discrecionalidad o cuando el ejercicio de la potestad discrecional supone el uso de criterios técnicos y no políticos»⁸. Esta cuestión, con independencia de las discusiones doctrinales, tiene consecuencias prácticas, por cuanto el uso de la IA será mucho más común y extendido si se adopta la segunda visión, por un lado. Además, las

6. Boix Palop, A., «Los algoritmos son reglamentos... cit.; Carloni, E., IA, algoritmos y Administración pública en Italia», *Revista de Internet, Derecho y Política*, 2020.

7. Ponce Solé, J., «Reserva de humanidad y supervisión humana de la Inteligencia artificial», *El Cronista del Estado Social y Democrático de Derecho*, 100, 2022, pp. 58-67; Cerrillo i Martínez, A., «¿Son fiables las decisiones de las Administraciones públicas adoptadas por algoritmos?», *European review of digital administration & law*, 2020.

8. Boix Palop, A., «Los algoritmos son reglamentos... cit.; Cerrillo i Martínez, A., El impacto de la inteligencia artificial en las Administraciones públicas: estado de la cuestión y una agenda», en Cerrillo i Martínez, Agustí et Peguera Poch, Miquel (Coords.), *Retos jurídicos de la inteligencia artificial*, Thomson-Reuters Aranzadi, 2020, p. 24.

garantías que requieren ambos sistemas difieren, puesto que un sistema donde la IA pueda determinar o ayudar a la delimitación de la toma de decisiones no regladas, que por otro lado es donde de verdad tendrá una verdadera utilidad diferencial, pues es donde la IA puede suponer mejoras (un algoritmo que automatice decisiones regladas, en realidad, no es ni siquiera un uso de IA a partir de la definición del art. 3 RIA⁹), pero en estos casos los posibles riesgos para los derechos de los administrados se disparan, no sólo por cuestiones de problemas de aplicación o corrección y de eficacia sino, por ejemplo, por las muy diversas y potencialmente graves afecciones a derechos fundamentales (Soriano Aranz los ha analizado detalladamente¹⁰; *vid.* e también su trabajo del año 2023¹¹). No obstante lo anterior, es un error considerar que la mera supervisión final de la persona física a la hora de emitir el dictamen o la resolución administrativa sea garantía suficiente para confiar en la no afectación a los derechos fundamentales, pues el propio sistema tiene un peligro inherente: su perfeccionamiento genera exceso de confianza, y el exceso de confianza mengua la propia percepción de necesidad de supervisión, incluso llegando a condicionar la opinión propia que quedara a expensas de la respuesta de la IA¹². Las soluciones no pueden pasar siempre por esta supervisión humana, además, porque en muchos casos la misma será redundante o disfuncional¹³, aunque como veremos ésta sí es una aproximación que ha adoptado el RIA para los sistemas de alto riesgo¹⁴.

El primer mecanismo de control que debe garantizar la no afectación de derechos fundamentales para el ciudadano, en el contacto con las decisiones de IA adoptadas en la administración pública, debe ser una norma consciente de las posibles vulneraciones que regule el buen uso, se tratará de la plasmación legal y proyección sobre el sector del principio de precaución¹⁵, por el que deben extremarse las precauciones, y eliminarse, por ejemplo, el riesgo inherente en la gestión documental aislando todos aquellos datos cuyo tratamiento pueda generar vulneración de derechos. La administración debe proteger en mayor medida los derechos afectados

9. Cotino Hueso, L., «Los usos de la IA en el sector público, su variable impacto y categorización jurídica», *Revista Canaria de Administración Pública*, n.º 1, 2023, pp. 213.124, crítico con esta definición restrictiva.
10. Soriano Aranz, A., *Data protection for the prevention of algorithmic discrimination. Protecting from discrimination and other harms caused by algorithms through privacy in the EU and US: possibilities, shortcomings and proposals*, Thomson-Reuters Aranzadi, 2021.
11. Soriano Aranz, A., «Creating non-discriminatory Artificial Intelligence systems: balancing the tensions between code granularity and the general nature of legal rules», *IDP: Revista de Internet, Derecho y Política*, 38, 2023.
12. Obregón Fernández, A. y Lazcoz Moratinos, G., «La supervisión humana de los sistemas de inteligencia artificial de alto riesgo», *Revista electrónica de estudios internacionales*, 42, 2021.
13. Lazcoz Moratinos, G., «Análisis de la propuesta de Reglamento sobre los principios éticos para el desarrollo, el despliegue y el uso de la inteligencia artificial, la robótica y las tecnologías conexas», *Ius et ciencia: revista electrónica de Derecho y Ciencia*, 6/2, 2020, pp. 26-41.
14. Cotino Hueso, L., «Los usos de la IA en el sector público...», *cit.*, pp. 221-223.
15. Boix Palop, A., «Los algoritmos son reglamentos...», pp.225 y ss, *cit.*; Cerrillo i Martínez, A., «El impacto de la inteligencia artificial... pp. 76 y ss, *cit.*; Cotino Hueso, L., Riesgos e impactos del Big Data, la inteligencia artificial y la robótica: enfoques, modelos y principios de la respuesta del derecho», *Revista General de Derecho Administrativo*, 50, 2019 pp. 225 y ss.

por la toma de decisiones en su ámbito con aplicación de la IA, pero ya adelantamos, acogerse a la norma que estamos analizando en toda su extensión generara una fricción con la naturaleza del ámbito público (que necesariamente requiere de cierta laxitud para proteger el sistema público), perjudicial en última instancia para los derechos del ciudadano. Los mayores peligros se generan, como es obvio, cuando el algoritmo puede convertirse en sustituto de ley¹⁶, lo que en ningún caso ocurre desde una perspectiva formal si el sistema está dotado de un régimen jurídico concreto y garantista, con adecuada cobertura legislativa concreta para el ámbito de la administración, pero puede ocurrir materialmente si la definición y concreción de decisiones con contenido discrecional que afectan a la esfera de derechos y deberes de los ciudadanos no están suficientemente perfiladas *ex ante* y no se logra un correcto acotamiento del funcionamiento de los modelos para adecuarlos a las finalidades públicas requeridas. Por ello uno de nosotros ha incidido, extensamente, en la necesidad, por el momento no atendida, y a la que el RIA tampoco presta excesiva atención (debido a esa preocupación orientada a una regulación más de mercado y sobre productos y servicios), de entender, comprender y ordenar y controlar debidamente el empleo de IA por parte del sector público que cumpla este efecto materialmente normativo a partir de la introducción de unas garantías y controles adicionales, propios del sector público, porque en estos casos la IA va a tener efectos materialmente normativos (si se quiere, en el ámbito de la Administración pública, materialmente reglamentarios¹⁷) al ser por medio de ésta como se concretará el efectivo ámbito de actuación del poder público en cada caso.

Como ya hemos señalado, nada de ello aparece en el RIA, que es un instrumento diseñado como modelo jurídico de intervención que trata de exigir transparencia en cuanto al acceso¹⁸, un importante control, auditorías externas, mecanismos propios del sector privado. En un futuro, en nuestra opinión, deberíamos aspirar a poder disponer de una regulación propia al uso de la IA para administraciones públicas, que se diferencie y vaya más allá del establecido para los entes privados en todo aquello que sea necesario para adaptar la naturaleza del sector público y la protección de los derechos fundamentales de los administrados con los que se trabaja diariamente. Una regulación que, dentro de este marco, proteja aún más al ciudadano, para asegurar que el proceso de sustitución decisoria por IA respeta todas las garantías. Algo que consideramos necesario por cuanto las posibilidades de afección a los derechos de los ciudadanos y a su estatuto jurídico básico por parte de los poderes públicos, y en concreto su capacidad de dañarlo, son muy superiores a las del sector privado, algo que nos parece básico que el Derecho tenga en cuenta. Sin embargo, no estamos de momento en este punto ni es la función del RIA llevar a cabo esto, que queda para un estado legislativo posterior y en gran parte para la responsabilidad en Derecho interno de los propios Estados Miembros e, incluso, para la propia autoorganización

-
16. Martín Delgado, I., «La aplicación del principio de transparencia a la actividad administrativa algorítmica», en E. Gamero Casado (dir.) y F. L. Pérez Guerrero (coord.), *Inteligencia artificial y sector público: retos, límites y medios*, Tirant Lo Blanch, 2023, p. 138.
 17. Boix Palop, A., «Los algoritmos son reglamentos...», cit.
 18. Boix Palop, A., «Transparencia en la utilización de inteligencia artificial por parte de la Administración», *El Cronista del Estado Social y Democrático de Derecho*, 100, 2022, pp. 99-105.

administrativa de cada poder público. Se trata, pues, de un debate pendiente sobre el que no apuntaremos más.

Una vez encuadrado lo que pretende hacer el RIA y lo que no, y enmarcado en este análisis moderadamente crítico de la ambición del marco normativo vigente, centraremos el presente estudio directamente en el articulado del que disponemos en el RIA y realizaremos un análisis descriptivo de los preceptos que consideramos que, más allá del modelo general de control descrito, afectan más directamente al sector y administración pública. En concreto, vamos a referirnos a:

- Los considerandos 4, 5, 6, 131 y 157.
- Las prohibiciones del artículo 5, que vinculan a las Administraciones.
- Las precauciones del artículo 6, en relación con las obligaciones que el artículo 27 impone a los organismos públicos (en conexión con los artículos 49 y 71).
- Las situaciones que contempla en Anexo III que atañen a los poderes públicos, especialmente los puntos 5 a 8.
- Las referencias a las actuaciones administrativas de los artículos 30, 34, 43, 45, 56, 57, 58, 59, 63, 66, 79, 82, 99 y 100.

II. EVOLUCIÓN, TRAMITACIÓN Y CONTENIDO FINAL DE LOS PRECEPTOS DEL REGLAMENTO QUE AFECTAN EN MAYOR MEDIDA A LOS PODERES PÚBLICOS

Conviene analizar, por su interés, la evolución de las modificaciones y ampliación de la ambición legislativa de los preceptos que afectan a los poderes públicos e el RIA, desde la primera propuesta preparado por la Comisión Europea en abril de 2021, hasta el último texto propuesto por el Consejo, en la versión que incorpora las enmiendas aprobadas por el Parlamento Europeo, sintetizando y concretando la parte del articulado que realmente vincula directamente a las administraciones públicas.

1. CONSIDERACIONES GENERALES QUE PUEDEN PROYECTARSE SINGULARMENTE SOBRE LOS PODERES PÚBLICOS EN SU USO DE INTELIGENCIA ARTIFICIAL

La Propuesta inicial de Reglamento partía de una organización de la regulación categorizando el riesgo del uso de los sistemas de IA, desde un riesgo mínimo o bajo, hasta el alto o el inaceptable, y este modelo es el que finalmente ha acabado aprobándose. Para aquellos usos que se consideren como de riesgo mínimo, el RIA lanza recomendaciones, que se aceptan por voluntariedad, con códigos de conducta y buenas prácticas, y alguna obligación de transparencia¹⁹; por contra, los sistemas de riesgo inaceptable se encuentran prohibidos; en el punto intermedio, copando el grueso de la normativa, aquellos que son de alto riesgo, que nos interesan particularmente pues son los que afectan a los asuntos del sector público, si bien como veremos, con bastantes excepciones. El tratamiento en este esquema regulador de las especificidades regulatorias que afectan a los poderes públicos es parco, y ello lo

19. Boix Palop, A., «Transparencia en la utilización de inteligencia artificial...», cit.

vemos incluso desde la propia exposición de motivos razonada de la necesidad de una regulación, donde las menciones a la posición diferencial de los poderes públicos son menores, parcas y con poca incidencia estructural sobre la normativa.

Así, por ejemplo, el Considerando 4, analizando la IA como conjunto tecnológico al que reconoce su rápida evolución y del que enumera una serie de beneficios, aunque se centra principalmente en la vertiente económica y de competitividad, añade referencias a beneficios sociales o medioambientales y a algunos ámbitos relacionados, mencionando los servicios públicos, la seguridad o la justicia. Como puede verse, el foco no es en ningún caso el sector público, sino que éste aparece como un beneficiado tangencial e indirecto de unas mejoras y avances que, en esencia, se están produciendo en otros ámbitos y que despliegan sus ventajas en el conjunto de la economía. Eso sí, el Considerando 5 ha reconocido, finalmente, ciertos riesgos para los intereses públicos y derechos fundamentales, categorizando los menoscabos como tangibles o intangibles, y los perjuicios como físicos, psíquicos, sociales o económicos. Debemos recordar en este sentido, una vez más, que la afeción a dichos derechos, y por tanto los riesgos inherentes, son mucho mayores cuando la IA se utiliza para la toma de decisiones administrativas, por la propia naturaleza del sector público, y el deber del mismo de protección reforzada frente al que compete a las entidades privadas. Sin embargo, no parece que esta visión sobre los mayores riesgos en estos casos vaya a traducirse después en la normativa, pues no habrá normas propias ni específicas ni cautelas adicionales. Eso sí, el Considerando 6, reconociendo la importancia y la repercusión en varios ámbitos de la IA, y la necesidad de que el RIA asegure el respeto a los valores de la unión (art. 2 i 6 del TUE, derechos y libertades fundamentales de los Tratados y la Carta), permite proyectar sobre las actividades públicas estas cautelas.

Algo más de concreción sobre el poder público, pues del mismo se pueden extraer obligaciones que hay que tener en cuenta muy directamente, se deriva del Considerando 131, en materia de transparencia, cuando se establece el deber de los proveedores de la IA de alto riesgo a los que no se les impondría en principio la legislación de armonización de la Unión, y a aquellos que consideren que no se encuentran encuadrados en alto riesgo por aplicárseles una excepción, que deben registrarse en la base de datos de la UE creada por la Comisión. Y concretamente impone a las autoridades, órganos u organismos públicos el deber de registrarse en la base de datos indicando el sistema cuyo uso tengan previsto. Por su parte, el Considerando 157 se refiere al ámbito competencial, estableciendo que la aplicación del mismo se entiende sin perjuicio de las competencias, funciones, poderes e independencia de las autoridades u organismos públicos nacionales, asegurando el acceso de los mismos a la documentación creada en virtud del RIA. Añade en relación con lo anterior, y sobre todo en cuanto a protección de derechos fundamentales, la necesidad de un procedimiento concreto de salvaguarda, cuando la IA presente un riesgo alto para la salud, seguridad o derechos fundamentales, a aplicar: A los sistemas de alto riesgo, a los sistemas prohibidos introducidos en el mercado, a los puestos en servicio o utilizados contraviniendo las prácticas prohibidas del RIA y a los sistemas comercializados infringiendo los requisitos de transparencia que presenten un riesgo.

2. AFECCIÓN A LAS ADMINISTRACIONES PÚBLICAS Y A LOS PODERES PÚBLICOS DE LOS USOS PROHIBIDOS DE INTELIGENCIA ARTIFICIAL ESTABLECIDOS EN EL ART. 5 REGLAMENTO

Aunque el análisis del art. 5 RIA y de las prohibiciones establecidas en el mismo se hace en otro lugar de esta obra, hay que dejar constancia en este análisis de cómo se proyecta el RIA sobre la actividad de los poderes públicos de que algunas de estas prohibiciones impactan muy directamente sobre ámbitos de actuación pública que, directamente, pasan a estar prohibidos (en realidad, ya o estaban en todos los casos: el RIA simplemente incide en que ciertas actuaciones que nuestros poderes públicos ya no podían hacer por exigencias derivadas de la protección básica de nuestros derechos fundamentales tampoco podrán realizarse empleando IA).

El art. 5 RIA trata de evitar este impacto en los derechos de los ciudadanos al enumerar las prácticas de IA que estarían prohibidas por la Unión Europea, por ser potencialmente peligrosas en cuanto a la vulneración de valores y derechos fundamentales del ordenamiento europeo, o susceptibles de ejercer una manipulación general de la población y, sobre todo, de aquellos sectores más desprotegidos y vulnerables. Establece una serie de supuestos prohibidos que han ido aumentando cuantiosamente desde la Propuesta de Reglamento abril 2021 de la Comisión Europea a la Posición común («enfoque general») sobre la Ley de IA el 6 de diciembre de 2022, del Consejo Europeo, con la Posición negociadora del Parlamento sobre el RIA, junio de 2023 Parlamento Europeo, EPRS, agrupando finalmente un compendio con una serie de elementos comunes.

De dicha enumeración de prohibiciones tiene poco sentido centrarnos en aquellas que suponen, como se ha dicho, una prohibición absoluta que pueda impedir a las Administraciones públicas el uso de la IA para ciertas finalidades o funciones, pero no porque se quiera impedir el empleo de IA para estas actividades, sino porque es la actividad en sí la que está prohibida por incompatible con nuestro sistema de garantías, derechos y Estado de derecho. Por ejemplo, los sistemas de IA no puede ser empleados para generar modelos de crédito social, sencillamente, porque estos modelos se entienden propios de dictaduras o modelos autoritarios, sean aplicados usando IA o no. Del mismo modo, una vigilancia constante de los ciudadanos es incompatible con un modelo democrático, ya sea empleando IA o no, y por ello queda prohibida también cuando se emplea IA.

Más interesante es señalar los casos en que ciertos usos de la IA generalmente prohibidos encuentran cierta relajación cuando son los poderes públicos los que los llevan a cabo para actividades que se consideran socialmente justificadas, excepcionalmente, por su finalidad para la prevención del delito. Así, podemos destacar la diferenciación que se percibe, dentro de estos sistemas de tan alto riesgo que normalmente quedan totalmente prohibidos, a los sistemas de identificación biométrica remota (a diferencia de otros sistemas, donde cabe cierta regulación²⁰), que tratan de identificar a personas a distancia con sus datos biométricos incorporados en bases de referencia, necesitados de concesión previa por autoridad judicial o administración independiente (artículo 5.3) y sometidos a controles especiales

20. Cotino Hueso, L., «Sistemas de inteligencia artificial con reconocimiento facial y datos biométricos. Mejor regular bien que prohibir mal», *El Cronista del Estado Social y Democrático de Derecho*, 100, 2022, pp.68-79.

de transparencia (artículo 52) y unos sistemas a los que considera en todo caso de tal riesgo que, con carácter general, quedan prohibidos (artículos 5.1.d) y 5.2). Así, el apartado 5.1. d) prohíbe al introducción en el mercado, puesta en servicio o uso de la IA para evaluar riesgos y probabilidad delictiva en las personas físicas, con elaboración de perfiles y evaluación de personalidad, la excepción vacía prácticamente de contenido la prohibición para el ámbito público, pues indica que no se aplicará *«a los sistemas de IA utilizados para apoyar la evaluación humana de la implicación de una persona en una actividad delictiva que ya se base en hechos objetivos y verificables directamente relacionados con una actividad delictiva»*. El apartado h) concreta el uso de dicha identificación biométrica «en tiempo real» en espacios públicos, prohibiendo su uso, de nuevo salvo búsqueda de víctimas o personas desaparecidas, prevención de amenazas contra la vida o seguridad (señalando concretamente los atentados terroristas) o la localización de sospechosos por infracciones penales o ejecución de sanciones por los delitos del anexo II con penas o medidas de seguridad de al menos 4 años. Por su parte, dentro del apartado 5.2 se regulan ciertos aspectos para evaluar la aplicación de la letra h), en cuanto a confirmar la identidad de la persona, naturaliza de la situación, gravedad y magnitud del no uso de la IA, y la afectación a los derechos y libertades de la persona implicada. Alude el artículo a la regulación nacional de cada estado miembro, en cuanto a las limitaciones temporales, geográficas y las relativas a las personas, y a la previa autorización de la autoridad encargada en la aplicación de la ley, con una evaluación del impacto en los derechos fundamentales según el artículo 27 que posteriormente analizaremos, y registrado en el sistema de bases de datos del artículo 49. De nuevo juega el RIA con la contra excepción, incluso en este precepto, al matizar que *«(n)o obstante, en casos de urgencia debidamente justificados, se podrá empezar a utilizar tales sistemas sin el registro en la base de datos de la UE, siempre que dicho registro se lleve a cabo sin demora indebida»*.

Añade el punto tercero, en cuanto a la autorización previa, por parte de una autoridad judicial o administrativa independiente, el requisito de solicitud motivada conforme a las normas nacionales, y volvemos de nuevo a rebajar las garantías, pues no será necesaria cuando la situación de urgencia está justificada, siempre que se solicite posteriormente sin demora en un plazo máximo de 24 horas, interrumpiéndose en caso de ser denegada y suprimiéndose la información. Se añade una exigencia posterior de notificación a la autoridad de vigilancia del mercado pertinente y a la nacional de protección de datos, que contenga la información del apartado 6, y *«sin datos operativos sensibles»* con obligación de las mismas de comunicar informes anules a la comisión (que posteriormente serán publicados por ésta). Se realiza por el RIA un llamamiento a los Estados miembros para regular en sus respectivos Derechos nacionales las normas para la solicitud, concesión, ejercicio, supervisión y notificación de las autorizaciones, debiendo notificar las normas a la Comisión (a más tardar en 30 días tras su adopción), y concediendo la opción de regular de modo más estricto el uso de los sistemas de identificación biométrica. Es decir, el RIA trata de establecer un marco intervencionista con unos mínimos a aplicar, que pueden desarrollarse de modo más restrictivo, pero no más laxo por cada parlamento nacional.

Como puede constatarse, en estos casos el RIA levanta la prohibición absoluta de uso de IA cuando entiende que hay una finalidad legítima, pasando en estos casos el uso a ser de alto riesgo. Ha de señalarse, con todo, que la laxa y muy amplia definición de los supuestos que dan paso a la posibilidad de usar estas herramientas

de IA, puede permitir que una apelación muy genérica a estos riesgos habilite para un uso más masivo del querido aparentemente por la norma (por ejemplo, el control de posibles actividades terroristas o la apelación a la persecución de determinados delitos puede llevar, y así ha manifestado su preocupación ya parte de la sociedad civil) a avalar la activación de controles muy genéricos, por ejemplo en frontera, de sistemas de IA que impliquen este tipo de funcionalidades. Para evitar que esta excepción acabe provocando este efecto hará falta normativa de desarrollo que ciña y controle esta habilitación, tanto en Derecho europeo como en su integración nacional.

3. SOBRE LAS PRECAUCIONES DEL ART. 6 REGLAMENTO RESPECTO A LOS USOS DE ALTO RIESGO EN RELACIÓN A LAS OBLIGACIONES QUE SE DERIVAN DE LOS MISMOS PARA LOS PODERES PÚBLICOS

De nuevo, el análisis del art. 6 RIA, en concesión con el Anexo III, a efectos de la delimitación de los sistemas considerados de alto riesgo por el RIA, de lo que se deriva un régimen jurídico de requisitos y exigencias reforzado, corresponde a otra parte de esta obra. Pero vamos a tratar de cartografiar cómo se proyecta específicamente esta regulación respecto de los poderes públicos.

Así, en general, para poder ser permitidos los usos de la IA tenidos como de alto riesgo por la conjunción de estos preceptos, deben cumplirse unos requisitos durante toda su existencia. Las precauciones específicas que afectarán al ámbito público, en tanto que no son diferente a las de cualquier sujeto privado, las encontramos también, por ello, en el propio art. 6 RIA, que deben ser analizadas en este caso siempre junto con las obligaciones adicionales impuestas en el art. 27 RIA para los distribuidores que sean organismos de Derecho público, o entidades privadas que presten servicios públicos, éstas sí directa y específicamente pensadas para estos casos. Y todo ello sin olvidar que serán siempre esenciales también las previsiones generales relativas a evaluaciones, certificaciones y registros, para aportar la necesaria transparencia y seguridad al sistema de IA, dentro de los supuestos de alto riesgo entre los que la Administración se encuentra incardinada.

Empezando por las reglas de clasificación de los sistemas de IA de alto riesgo del art. 6 RIA, tras las dos condiciones genéricas para clasificar el alto riesgo, añade el precepto los sistemas de IA contemplados en el Anexo III. Y he aquí una de las modificaciones más importante a lo largo del proceso legislativo, se amplían los supuestos no considerados de alto riesgo con el cumplimiento de alguna de las condiciones siguientes (no aplicables a la elaboración de perfiles de personas físicas):

- Tareas de procedimiento limitadas.
- Mejoras del resultado una actividad humana previamente realizada.
- La detección de patrones en la toma de decisiones o la desviación de los mismos respecto a decisiones anteriores, sin estar destinado a sustituir la evaluación humana (sin una revisión adecuada) ni a influir ella. (Algo bastante difícil de concebir en la realidad, pues la herramienta de la IA tiende por su propia naturaleza en acabar absorbiendo la competencia real y efectiva de la decisión, una vez la inteligencia humana confía tanto en ella que deja de considerar la necesidad de intervenir).
- Tareas preparatorias de evaluación en los usos del Anexo III.

Como es sabido, si el proveedor considera que su sistema de IA cumple con dichos requisitos y por tanto, pese a encontrarse en el Anexo III no es de alto riesgo, debe documentar su evaluación previamente a la introducción en el mercado, y registrar conforme el artículo 49.2 (debiendo documentar la evaluación a petición de las autoridades, sin ser, por tanto, tan siquiera requisito previo para la introducción en el mercado la presentación de dicha evaluación). Además, la Comisión se reserva el derecho de añadir nuevas condiciones o modificar las existentes (sin reducir el nivel global de protección) cuando considere que existen sistemas contemplados en el anexo III, pero que a su juicio no planteen riesgo importante de causar un perjuicio a la salud, la seguridad o los derechos fundamentales de las personas físicas.

En definitiva, hemos de señalar que en todo caso se consideran las actuaciones administrativas que empleen IA como explícitamente señaladas por el Anexo III, donde se incluye genéricamente la IA en el sector público como de alto riesgo (aunque posteriormente se abre la opción a cierta flexibilización en algunos casos, con una serie de condiciones bastante amplias que quizás permiten extraer algunos actos de esta categoría más garantista al considerarlos de menos afectación a los derechos y garantías, para añadir además posteriormente la opción de no aplicar el apartado tercero, modificando o añadiendo más excepciones, si a criterio de la Comisión no hay afectación de derechos ni garantías del administrado, creemos que hay que entender esta situación como excepcional respecto de usos que incidan sobre la esfera de derechos y deberes de los ciudadanos y que sólo debiera aplicarse a procesos internos). Y ello porque el apoyo constante en las tareas administrativas, tan solo juzgando la corrección de sus decisiones, con una herramienta que facilita la labor con tal magnitud, genera inevitablemente una dependencia de la que resulta muy difícil desvincularse, y si bien teórica y formalmente es la persona responsable la que emite la resolución, de facto es la IA la que la produce, sustituyendo la inteligencia humana, y conllevando finalmente una independencia del algoritmo para llegar a tomar decisiones propias. La no definición concreta de los márgenes de aquello que debe considerarse alto riesgo genera un peligro más que patente para los derechos del administrado en un proceso administrativo concreto y, en general, para los derechos de los ciudadanos.

Por lo que respecta al art. 27 RIA, respecto del impacto en los derechos fundamentales para los IA de alto riesgo, es clave señalar que cuando los responsables del despliegue sean organismos de Derecho público o entidades privadas que prestan servicios públicos el RIA obliga a evaluar el impacto que el uso de la IA pueda tener en los derechos fundamentales siempre con particular rigor. Dicha evaluación (que deberá realizarse en el primer uso del sistema, basándose posteriormente en la anteriores) consiste según el articulado en describir los procesos del responsable del despliegue con su finalidad, el período de tiempo y la frecuencia de uso, las personas o grupos que puedan verse afectados por el sistema, los riesgos específicos y las medidas de supervisión humana, junto con las medidas a adoptarse si los riesgos se materializan, junto con mecanismos para reclamar por ello. Tras la evaluación se deben notificar los resultados a la autoridad de vigilancia del mercado, salvo que esté exceptuado por la propia autoridad conforme al artículo 46.1 RIA.

Por último, respecto al registro del sistema como mecanismo de garantía y transparencia, una de las principales garantías con las que se debe contar, y que regula el RIA en su art. 49 en relación con el art. 71 RIA, es la relativa a la publicidad y control registral. Con el uso de la IA, es necesario un control y una transparencia de modo

que las propias autoridades y el ciudadano sepan de dónde proviene el uso de la IA, en qué condiciones se ha tomado la decisión o resolución, con qué consideraciones y finalmente si ha sido o no revisada por el responsable funcional que la rúbrica. Si la IA predetermina una decisión administrativa, debemos posibilitar siempre que el ciudadano sepa que existe y como se les está aplicando de modo que el ciudadano disconforme con el resultado que quiera recurrir el acto correspondiente podrá tener interés en conocer la configuración del algoritmo, así como su correcta aplicación en el supuesto concreto²¹ y es esencial que el RIA imponga a la Administración el deber de atender dicha reclamación. Establece así el art. 49 RIA que previamente a la introducción en el mercado del sistema de IA de alto riesgo, deben registrarse en la base de datos de la UE. Dicha base de datos, según el art. 71 RIA será elaborada por la Comisión en colaboración con los estados miembros, consultando a los expertos pertinentes, con división entre secciones dependiendo del tipo y persona sujeta a la obligación de inscripción, accesible al público como medio de garantía a excepción de los datos sensibles tan solo visibles por las autoridades de vigilancia del mercado y por la Comisión, que además será la responsable del tratamiento de la misma, y proporcionara apoyo y accesibilidad a obligados y responsables. Ha de ser señalado que, si nos encontramos en el supuesto de aplicación de las excepciones del artículo 6 en su apartado 3, el registro se realizará por el proveedor o representante autorizado, que registrará el sistema por sí mismo, también para el caso de autoridades, órganos u organismos públicos, se registrarán por sus representantes con selección del sistema y de su utilización. Puntualiza la norma que para el caso de sistemas de IA de alto riesgo del anexo III relativos a l uso de la biometría (punto 1), aplicación de la ley (punto 6), y migración, asilo y gestión de control fronterizo (punto 7) el registro se efectuará en una sección segura, ampliando la protección de datos respecto al registro general con información concreta a la que únicamente la Comisión y las autoridades nacionales tendrán acceso.

4. PROYECCIÓN DE LA DELIMITACIÓN DE LOS USOS DE ALTO RIESGO SEGÚN EL ANEXO III SOBRE EL EMPLEO DE INTELIGENCIA ARTIFICIAL POR PARTE DE LOS PODERES PÚBLICOS

Como ya hemos adelantado, la regulación conjunta del control del uso de la IA para entes privados y para la administración pública, tal y como ha optado el RIA, es problemática, pues la competencia de los poderes públicos afecta e incide directamente sobre el estatuto jurídico personal de los afectados de un modo mucho más sensible, por la propia naturaleza de las relaciones del sector público con los administrados. La regulación suscita ciertas dudas en cuanto a su aplicación a las Administraciones Públicas, incluso desde una vertiente sistemática, pues su articulado específico se halla comprendido de modo más concreto en éste Anexo III, el mero hecho de regularse su aplicación concreta en un Anexo denota la distancia entre situaciones jurídicas, en tanto en cuanto la estructura básica de un marco diseñado para los entes privados y grandes empresas puede suponer ignorar algunos riesgos adicionales para ciertos derechos al aplicarse al sistema público. A los agentes privados se les pueden imponer restricciones, intervenciones, auditorías y una serie

21. De la Sierra Morón, S., «Control judicial de los algoritmos: robots, administración y estado de derecho», en Lefebvre (11 de junio de 2021): <https://elderecho.com/control-judicial-de-los-algoritmos-robots-administracion-y-estado-de-derecho>

de condiciones de autoprotección que deben cumplir, y que en ocasiones se trasladan difícilmente en toda su extensión para la administración (véanse a este respecto los problemas en materia sancionadora, semejantes a los que ya se han producido en materia de protección de datos). Pero, además, hay problemas adicionales que afectan al sector público que esta óptica sencillamente no puede abordar. De ello se derivan riesgos evidentes para la protección de los derechos de los ciudadanos cuando se emplea IA por el sector público, por su mayor capacidad de afectación. Ello no obstante, las garantías y cautelas que el RIA introduce para los sistemas de alto riesgo son, al menos, un mínimo sobre el que avanzar y que, a partir de este momento, pasarán a ser exigibles para todos los usos públicos que se entienda que puedan quedar subsumidos en las definiciones del Anexo III.

En este sentido, y a partir de la evolución legislativa sufrida por el Anexo III, de la que se deducen ciertas incoherencias menores, pues en ocasiones parece que todos los usos de IA por parte del poder judicial o de las administraciones públicas pasan a quedar como definidos de alto riesgo si tienen que ver con la toma de decisiones o prestación de servicios públicos, pero por otro lado se añaden algunas previsiones adicionales sobre servicios concretos o actuaciones concretas que podrían hacer albergar la duda respecto de la mayor intensidad del control en unos casos frente a otros. En nuestra opinión, la manera más correcta y garantista de interpretar en este punto el RIA es también la más sencilla: cualquier uso de IA para la ayuda a la toma de decisiones o directamente que sustituya a la misma de forma completa respecto del ejercicio de funciones judiciales o administrativas o respecto de la prestación de servicios públicos es considerada en estos momentos, tras las sucesivas ampliaciones producidas en el proceso legislativo, como de alto riesgo. Y si hay previsiones adicionales para ámbitos más concretos ello sólo refuerza tal consideración respecto de estos ámbitos, obligando a una visión más cuidadosa y a una aplicación más estricta del principio de precaución²² y a entender en esos casos incluidos incluso usos que por ejemplo incidan indirectamente sobre las decisiones y actuaciones.

En concreto, el Anexo III menciona los sistemas de IA de alto riesgo con arreglo al artículo 6.2 en los siguientes ámbitos centrándonos en los que afectan a la administración, que son los relativos a los puntos 5, 6, 7 y 8 a cuyo texto concreto cabe remitir: acceso a servicios privados esenciales y a servicios y prestaciones públicos esenciales y disfrute de estos servicios y prestaciones: acceso a servicios privados esenciales y a servicios y prestaciones públicos esenciales y disfrute de estos servicios y prestaciones (5°); garantía del cumplimiento del Derecho, en la medida en que su uso esté permitido por el Derecho de la Unión o nacional aplicable (6°); migración, asilo y gestión del control fronterizo, en la medida en que su uso esté permitido por el Derecho de la Unión o nacional aplicable (7°) y Administración de justicia y procesos democráticos (8°).

5. PROYECCIÓN DE NORMAS DEL REGLAMENTO SOBRE ACTUACIONES ADMINISTRATIVAS

Adicionalmente a lo ya señalado, el RIA contiene referencias a las actuaciones administrativas de los artículos 30, 34, 43, 45, 56, 57, 58, 59, 63, 66, 79, 82, 99 y 100. Las referimos, siquiera sea brevemente.

22. Cotino Hueso, L., «Riesgos e impactos del Big Data ...», cit.

— Ya hemos indicado que debe seguirse un procedimiento de registro para las IA de alto riesgo (incluso en algunos casos estando las mismas exceptuadas de dicha categorización), y que dicha obligación también compete a las administraciones públicas, artículo 49 RIA en relación con el artículo 71. Junto a este mecanismo de transparencia, existirán en cada estado miembro autoridades encargadas de la supervisión y control:

Las autoridades notificantes, que cada estado nombrará para los procedimientos de evaluación, designación y notificación de la conformidad de las IA con el Reglamento y de su supervisión. El control del cumplimiento de los requisitos impuestos en los sistemas de IA de alto riesgo se ejercerá, por tanto, por la autoridad notificante, un organismo público designado al efecto por cada Estado miembro (artículo 30), que será el encargado de determinar los procedimientos para la evaluación del cumplimiento de la norma. Dichas autoridades requerirán la documentación a los organismos notificados, que a su vez tienen ciertas obligaciones operativas, tal y como indica el artículo 34 (si bien para las herramientas de IA empleadas por actores privados, pues como hemos señalado para algoritmos empleados por el sector público la evaluación de conformidad es interna): Deberán verificar la conformidad de los sistemas de IA de alto riesgo con los procedimientos de evaluación del artículo 43, evitando cargas innecesarias, y considerando el tamaño del proveedor, sector en el que opera, estructura, grado de complejidad de la IA, para minimizar cargas administrativas, respetando el rigor y nivel de protección requeridos. Los organismos notificados pondrán a disposición de la autoridad notificante y presentarán cuando se les pida, toda la documentación para que pueda llevarse a cabo la evaluación, designación, notificación y supervisión por la autoridad. En el artículo 45 y a raíz de lo anterior, se mencionan las obligaciones de información de los organismos notificados.

Junto con este órgano, el RIA señala en su artículo 59 la designación de la autoridad nacional de supervisión (que también puede ejercer el cargo de autoridad de vigilancia del mercado, conforme al artículo 63) con funciones de supervisión de la aplicación y ejecución del RIA, y representar a su Estado en el Comité Europeo de Inteligencia Artificial. Este comité garantizará la uniformidad en la aplicación del RIA en el conjunto de los Estados miembros. La autoridad nacional de supervisión será la encargada de la concesión de la autorización para la introducción o puesta en servicio de la IA de alto riesgo en el mercado, y conforme al artículo 45, todas estas decisiones deberán ser susceptibles de recurso.

— Mención especial merecen los artículos 57 y 58 en cuanto a la regulación de los espacios controlados de pruebas para la IA, es éste caso ya no impone el RIA una actitud supervisora al Estado miembro, sino la creación de espacios para ensayos iniciales, con apoyo y asesoramiento de la Comisión, donde los estados velarán por la asignación de suficientes recursos, y se comprometerán a la cooperación con las autoridades pertinentes, de modo que se consiga un entorno seguro que fomente la innovación, la prueba y la validación de los sistemas innovadores de la IA antes de su introducción al mercado y puesta en funcionamiento, facilitando la autoridad un informe

de salida tras el proceso de evaluación, que las autoridades de vigilancia de mercado y organismos notificados tendrán en cuenta positivamente, sin intervenir en las facultades correctoras o de supervisión de éstos. Se pretende con los espacios controlados mejorar la seguridad, apoyar el intercambio y cooperación entre autoridades, el fomento de la innovación y competitividad y el aprendizaje con pruebas contrastadas, y la Comisión adoptará los actos que especifiquen las disposiciones detalladas para el establecimiento, desarrollo, puesta en práctica, funcionamiento y supervisión de los espacios de prueba controlados, enumerando el RIA los principios comunes que se deben respetar, y debemos comentar el apartado f) del artículo 58.2, al hablar de facilitar la participación de otros agentes del ecosistema de la IA (los organismos notificados y los organismos de normalización, las pymes, las empresas emergentes, las empresas, los agentes innovadores, las instalaciones de ensayo y experimentación, los laboratorios de investigación y experimentación y los centros europeos de innovación digital, los centros de excelencia y los investigadores) para permitir y facilitar la cooperación pública y privada.

— El RIA también otorga al Comité una serie de funciones relacionadas con el asesoramiento y la asistencia a la Comisión y a los estados miembros para la aplicación de la normativa en su artículo 66, dándole la posibilidad, que no la obligación, de contribuir a la coordinación y cooperación de las autoridades de vigilancia del mercado, recopilar información, ofrecer asesoramiento, emitir recomendaciones, dictámenes, códigos de conducta, evaluación del propio RIA y de las normas armonizadas, criterios comunes, integración de instituciones, emisión y recepción de los dictámenes, etc. y en concreto para lo que al sector público interesa, en el apartado d) *contribuir a la armonización de las prácticas administrativas en los Estados miembros, en particular en relación con la exención de los procedimientos de evaluación de la conformidad a que se refiere el artículo 46, el funcionamiento de los espacios controlados de pruebas y las pruebas en condiciones reales a que se refieren los artículos 57, 59 y 60; cerrando el círculo de agentes nacionales y europeos integrados para la aplicación del RIA.*

— Como ya hemos señalado, el art.79 establece un sistema de control de riesgos que puede tener una particular relevancia en los casos de usos de IA por parte de poderes públicos, en relación a las notificaciones de riesgos en estos casos del art. 82.

— Por último, cabe realizar una breve mención a las sanciones y multas administrativas a instituciones, órganos y organismos de la Unión Europea de los artículos 99 y 100 del RIA: para asegurar la aplicación de las disposiciones el texto incluye un sistema sancionador que corresponde a los estados miembros determinar, dentro de los márgenes de ésta articulado. Las sanciones deben ser efectivas, proporcionadas y disuasorias, teniendo en cuenta criterios subjetivos como los intereses de las empresas emergentes, viabilidades económicas, etc. Las cuantías dependen de la gravedad de la infracción, y se contempla la posibilidad de imposición de multas adicionales a medidas no monetarias, como órdenes o advertencias. Respecto a la multa administrativa, también debe decidirse la cuantía concreta dependiendo

de la situación, naturaleza, gravedad, dilación, etc. Además, se prevé también que el supervisor europeo de protección de datos pueda imponer multas administrativas a las instituciones, las agencias y los organismos de la Unión comprendidos en el ámbito de aplicación del RIA, quedando sino equiparados a los entes privados, si en consonancia con el espíritu sancionador que la administración también debe soportar. Igualmente se gradúan tomando en consideración todas las circunstancias pertinentes de la situación de que se trate, y en concreto conforme a la gravedad, duración, consecuencias, número de afectados y nivel de daños, grado de responsabilidad del organismo, acciones emprendidas para mitigar los perjuicios o el grado de cooperación con el Supervisor y su puesta en conocimiento, así como infracciones anteriores similares. Las cuantías de estas multas administrativas pueden llegar hasta unas cantidades, a partir de unos umbrales que dependen del tipo de incumplimiento y que son calculadas con arreglo a la cifra de negocios de la empresa afectada, lo que exigirá de una concreción diferenciada y adaptada para los poderes públicos.

III. ALGUNAS CONCLUSIONES SOBRE LA APLICACIÓN DEL REGLAMENTO AL SECTOR PÚBLICO

El RIA introduce, como hemos ido viendo y señalando desde un primer momento, un sistema de intervención flexible, que analiza el riesgo a los derechos fundamentales y valores de la Unión como premisa para condicionar el uso de la IA de los tres grados de afectación, pero que está esencialmente orientado para regular, ordenar y controlar el uso de IA, y los posibles riesgos que se derivan de la misma respecto de la introducción de productos o prestación de servicios en el mercado a partir de dinámicas esencialmente industriales y comerciales. Por ello, no está específicamente orientado a la regulación, control y minimización de riesgos respecto de usos por poderes públicos de estos productos o servicios, pero ello no es óbice para que no se les aplique esta regulación en sus mismos términos, y con algunas de las particularidades señaladas, cuando sean poderes públicos los que hagan uso de estos sistemas. Con independencia de la probabilidad de que en el futuro haya normas adicionales específicas para la acción administrativa o judicial, por los riesgos adicionales que conlleva respecto de la esfera de derechos y deberes de los ciudadanos el empleo de la IA por los poderes públicos, no puede sino valorarse positivamente este primer paso, que ya introduce importantes controles, hasta ahora inexistentes, sobre el empleo de IA para la adopción, por ejemplo, de decisiones administrativas o judiciales (o para ayudar a las mismas).

En primer lugar, y como ocurre con el sector privado, nos encontramos con usos donde directamente la IA está prohibida para los poderes públicos. El RIA no prohíbe su empleo, como algunos ordenamientos, para la adopción de decisiones discrecionales. Sin embargo, sí se vedan aquellos usos que pueden generar una gravísima afeción global a derechos a partir de dinámicas securitarias que podrían avalar derivas autoritarias de control desproporcionado sobre la población y que se centran, tal y como se desprende del RIA, en las funciones de las Administraciones Públicas relacionadas con la policía de seguridad, con la evaluación o clasificación de personas, con los sistemas de identificación biométrica remota «en tiempo real», así como con los sistemas públicos en los que tan sólo se permite su uso en

determinadas condiciones cuando no generen perjuicios o trato desfavorable o sean indispensables para localización de víctimas, presuntos delincuentes o prevención de amenazas²³.

Un grado por debajo de los anteriores, encontramos los usos de alto riesgo, que comprenden ciertas actividades del sector público en los que el administrado suele encontrarse en una posición desfavorable y más vulnerable frente a la autoridad (a modo de ejemplo la gestión de la migración, asilo y control fronterizo, con evaluación de riesgos y verificación de documentos, o la seguridad pública, con identificación biométrica y categorización de personas o los sistemas de evaluación de acceso a servicios y prestaciones). Los requisitos que este segundo grupo debe cumplir, serán determinados, aplicados y supervisados por varias entidades, con figuras jurídicas vitales para el funcionamiento del sistema como las autoridades notificantes, la autoridad nacional de supervisión (que también puede ejercer el cargo de autoridad de vigilancia del mercado), o los espacios controlados de pruebas para las IA, junto con los elementos de publicidad y transparencia de los registros, y los sistemas de sanción que aseguran el cumplimiento de la norma. Además, y de alguna manera como cierta plasmación del principio de «reserva de humanidad»²⁴, se impone la vigilancia humana de los sistemas de IA que entrañan mayor riesgo para los derechos (artículo 14) y de resultados de la extensa lista del Anexo III, puntos 5 a 8, podemos considerar que la práctica totalidad de actividades administrativas o adopción de decisiones judiciales que se realizan con ayuda o enteramente por IA e impacten sobre el estatuto jurídico de los ciudadanos, afectando a su esfera de derechos y deberes, serán también considerados de alto riesgo. Por último, y junto a reglas específicas para tipos de IA como *chatbots* o equivalentes, que evidentemente la Administración habrá de tener en cuenta cuando los utilice, el RIA también hace en la práctica equiparables en casi todo a los sistemas de alto riesgo a los sistemas de propósito general, de modo que cuando se empleen éstos, en la práctica, habrá de adoptar las mismas cautelas y exigir el cumplimiento a los proveedores del software.

En tercer lugar, y como ocurre también en el sector privado, un tercer grupo de usos de IA son calificados como con riesgo bajo o inexistente (esencialmente, los que ayuden a mejorar procesos y *backoffice*, sin incidencia directa en el estatuto de los ciudadanos, por mencionar el ejemplo más claro en materia de administración pública), y son considerados de desarrollo y uso libre, sin las restricciones del RIA, pero sin perjuicio de que puedan someterse voluntariamente a los previstos para los sistemas de alto riesgo a través de códigos de conducta.

Por último, una breve reflexión de las innovaciones tecnológicas y su regulación, la rápida evolución de la tecnología requiere de un marco jurídico eficaz pero adaptable a las constantes novedades, característica que se pretende asegurar con la posibilidad del Consejo de ampliar o modificar el contenido del RIA conforme

23. Cotino Hueso, L., «Sistemas de inteligencia artificial...», cit.

24. Cotino Hueso, L., «Los usos de la IA en el sector público...», cit., pp. 227-228, sobre los mayores riesgos en ciertos casos y cómo establecer cautelas de otro tipo en las decisiones completamente automatizadas, con todo, contempla acertadamente la situación en que esos menores riesgos nos lleven a este escenario.

a las vicisitudes que puedan surgir, revisiones por vía de reedición de los anexos que deben ir adecuándose a la realidad cambiante de la IA.

En las páginas precedentes hemos tratado de sistematizar la nueva normativa reglamentaria de la Unión Europea, que si bien, como ya hemos adelantado desde el inicio del análisis, creemos que no debería necesariamente regularse en un mismo texto normativo que el aplicable a los entes privados, al menos no en todos sus efectos, sí nos garantiza un mínimo marco jurídico al que atenerse (control humano, registros, organismos de vigilancia, requisitos mínimos, y sobre todo clasificación de riesgos), pues como en tantas otras ocasiones, la sociedad avanza a más velocidad que la normativa por la que debe regirse. Sería muy recomendable la elaboración en un futuro no muy lejano, de una normativa específica, un régimen jurídico concreto y adecuado que comprendiese la aplicación de la IA para el sector público, y contribuya realmente a conservar las específicas garantías que en este sistema se deben salvaguardar, a fin de disponer de un marco común europeo en la materia que luego los poderes públicos de cada Estado Miembro concreten, desarrollen y adapten a sus particularidades y derecho interno. En este sentido, y respecto de la normativa objeto de análisis, lo que se propone regular de manera armonizada estableciendo un mínimo en cuanto a control por los estados miembros, garantías esenciales para los derechos fundamentales y libertades públicas que informan el derecho de la Unión, ha de hacerse sin imponer a los Estados una regulación tan exhaustiva que no puedan establecer reglas propias. Y, por lo demás, ha de destacarse que en ningún caso la actual regulación europea vigente impide a los legisladores nacionales elaborar bases normativas propias internas más detalladas y garantistas que armonicen derechos e intereses públicos dentro del marco que establece el reglamento para los países de la Unión con la protección de los derechos de los ciudadanos. Se trata, con todo, de un segundo paso para el que era necesario, primero, comenzar a caminar. Es lo que ha hecho, también para el control de los usos de la IA por parte de los poderes públicos, tanto administrativos como judiciales, el RIA, que marca ya unos mínimos de protección nada desdeñables y ha de ser por ello bien valorado.

Grandes plataformas y sistemas de inteligencia artificial destinados a la influencia política: la intersección entre la «Ley de Servicios Digitales» y el Reglamento de inteligencia artificial desde la perspectiva del riesgo

ROSA CERNADA BADÍA

Profesora de Derecho Administrativo
Universidad Católica de Valencia San Vicente Mártir

I. INTRODUCCIÓN. FUNDAMENTO DEL TRATAMIENTO ESPECÍFICO DE LAS PLATAFORMAS DIGITALES Y LOS SISTEMAS DE INFLUENCIA POLÍTICA EN EL REGLAMENTO

Corría la segunda década de los años 80 del siglo XX cuando Ulrich Beck redefinía los contornos de nuestra modernidad al declarar superada la lucha de clases y esbozar el paso de una sociedad basada en el reparto de riqueza a una sociedad basada en reparto de riesgos¹. En esta sociedad del riesgo los conflictos surgen de los retos e intereses contrapuestos derivados del desarrollo científico-técnico e implica, a juicio del filósofo alemán, una pérdida de protagonismo de los Estados y el nacimiento en su lugar de «comunidades objetivas de amenaza» que requieren soluciones globales².

La idea, que ya se antojaba jugosa en el último cuarto del siglo XX, constituye una referencia necesaria en la actual sociedad hiperconectada, en la que el desarrollo digital ha reformulado las estructuras de mercado y los resortes sociales provocando un desafío jurídico sin precedentes. En este contexto, el nacimiento de las grandes plataformas digitales, a modo de nuevos cuerpos intermedios, ha provocado un impacto global, toda vez que: i) afecta de forma necesaria al contenido y ejercicio de los derechos fundamentales de los ciudadanos³; ii) ha generado nuevos modelos

-
1. Beck, U., *La sociedad del riesgo. Hacia una nueva modernidad*, Ediciones Paidós Ibérica, Barcelona 1998, p. 25.
 2. *Ibidem*, pp. 53-54.
 3. *Vid.* Resolución del Parlamento Europeo, de 14 de marzo de 2017, sobre las implicaciones de los macrodatos en los derechos fundamentales: privacidad, protección de datos, no discriminación, seguridad y aplicación de la ley (2016/2225(INI)).

de negocio, de dimensión transnacional⁴ bajo la lógica de la intermediación y iii) condiciona la relación entre el poder público y los ciudadanos, pudiendo afectar en último término al funcionamiento de los sistemas democráticos con fenómenos como la desinformación.

Precisamente estos ecosistemas particulares de naturaleza informativa, comercial y social que generan las plataformas digitales sobre estructuras algorítmicas son el caldo de cultivo idóneo para la proliferación de riesgos de muy diversa índole. Estos riesgos requieren de un examen específicamente orientado a la actividad de los servicios digitales y los sistemas de inteligencia artificial (en adelante, IA) que les sirven de base. La idea de las comunidades de amenaza de Beck subyace en esta lógica. La necesaria búsqueda de respuestas globales, también. Y la Unión Europea, fiel a sus principios constitutivos y a su vocación político social, está desarrollando una respuesta europea al reto de la gobernanza digital y el desarrollo de la IA como tecnología disruptiva. Esta respuesta, construida sobre la centralidad de la persona como piedra angular de la transición digital, se enmarca en la llamada estrategia digital europea⁵ que aglutina normas de gran calado. En particular, y junto al Reglamento General de Protección de Datos (en adelante RGPD)⁶, el llamado paquete normativo europeo, conformado por el Reglamento de Mercados Digitales⁷ y el Reglamento de Servicios Digitales⁸ (en lo sucesivo, DSA).

La DSA tiene por objeto contribuir al correcto funcionamiento del mercado interior de servicios intermediarios (...) para crear un entorno en línea seguro, predecible y fiable, que promueva la innovación y el respeto a los derechos fundamentales. Para ello regula la incidencia estatal sobre las grandes plataformas como complemento al control contractual estrictamente privado efectuado por aquéllas. Al efecto, partiendo del *safe harbour*, establece un régimen específico de responsabilidad de los prestadores de servicios intermediarios. Y, en concreto, en su título III consagra una serie de obligaciones de diligencia debida. Estas obligaciones son objeto de supervisión pública a través de un entramado institucional liderado por la Comisión Europea que garantiza el cumplimiento de la normativa por las grandes empresas

4. Otero Martín, D.; Infante González, J. y Ruiz Mérida, M., «Experiencia comparada: regulación y control de mercados digitales de plataforma en EE UU y China», *Plataformas digitales: regulación y competencia*, n.º 925 (marzo-abril 2022), p. 114.
5. Comisión Europea, «Shaping Europe's Digital Future», (febrero de 2020), disponible en: https://eufordigital.eu/wp-content/uploads/2020/04/communication-shaping-europes-digital-future-feb2020_en_4.pdf, último acceso, 15 de febrero de 2024.
6. Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos).
7. Reglamento (UE) 2022/1925 del Parlamento Europeo y del Consejo de 14 de septiembre de 2022 sobre mercados disputables y equitativos en el sector digital y por el que se modifican las Directivas (UE) 2019/1937 y (UE) 2020/1828 (Reglamento de Mercados Digitales) (Texto pertinente a efectos del EEE), DOUE n.º L 265/1, de 12 de octubre de 2022.
8. Reglamento (UE) 2022/2065 del Parlamento Europeo y del Consejo de 19 de octubre de 2022 relativo a un mercado único de servicios digitales y por el que se modifica la Directiva 2000/31/CE (Reglamento de Servicios Digitales) (Texto pertinente a efectos del EEE), DOUE L277/1, de 27 de octubre de 2022.

tecnológicas sometiéndolas a procedimientos de investigación y, en su caso, sanción por incumplimiento de sus responsabilidades.

Esta estructura normativa ha sido objeto de atención por el RIA⁹, en lo sucesivo RIA, en la medida en que las grandes plataformas basan su modelo de negocio en un servicio fundamental: la personalización, que hace uso de sistemas y modelos de IA. Los riesgos derivados de la incorporación de estos sistemas a la actividad de las grandes plataformas digitales resultan de gran calado, atendido el volumen de usuarios de estas plataformas y su potencial influencia en los derechos fundamentales, la seguridad en línea y la conformación de la opinión pública.

Junto a esta circunstancia, e íntimamente relacionada con ella, el RIA también otorga un tratamiento específico a los sistemas destinados a la influencia política. Su especial consideración parte de la preocupación de las instituciones europeas por el desarrollo y utilización de técnicas de manipulación política mediante sistemas de IA en el marco amplio de la lucha contra la desinformación.

La relación entre la desinformación y sus efectos sobre política han sido puestas de manifiesto en diversas circunstancias, si bien fue especialmente sangrante en la campaña de las elecciones presidenciales de 2016 en Estados Unidos (caso *Cambridge Analytica*) o el BREXIT. El origen de la reacción europea parte precisamente de un evento electoral, las elecciones al Parlamento Europeo de mayo de 2019, que incitó esta respuesta particularmente referida a la comunicación estratégica y prácticas de publicidad política. Se puso así de manifiesto¹⁰ la necesidad de intervención pública en este ámbito, dotando de prioridad a la transparencia frente a la postergación de contenidos de manera que los usuarios puedan entender cómo se construyen los resultados de salida de sus búsquedas de información o cómo se personalizan sus *feeds*.

Sin embargo, los sistemas diseñados para la influencia política no siempre hacen uso de las grandes plataformas para lograr sus objetivos. Un ejemplo reciente ha sido la campaña telefónica orquestada en Estados Unidos por la que los votantes recibían una llamada con la voz del presidente Biden en la que instaba a la abstención en New Hampshire¹¹. Por lo tanto, junto a la regulación de las grandes plataformas, resulta necesaria la atención específica a este fenómeno en un contexto creciente del uso de técnicas de IA para incidir en los resultados electorales y, en particular, la segmentación de la publicidad electoral, regulada por el Reglamento Europeo sobre Transparencia y Segmentación de la Publicidad Política (en lo sucesivo, RTSP)¹².

9. Propuesta de Reglamento del Parlamento Europeo y del Consejo, de 21 de abril de 2021, por el que se establecen normas armonizadas en materia de inteligencia artificial (ley de inteligencia artificial) y se modifican determinados actos legislativos de la unión, COM (2021) 206 final.

10. Mardsen, C. y Meyer, T., «Regulating disinformation with artificial intelligence», *Parliamentary Research Service of the European Parliament*, Brussels, European Union (2019), p. 6.

11. Vid. Doménech, E., «El “deepfake” que imita a Biden en plena campaña alerta a los expertos ante el uso de IA para manipular elecciones», NEWTRAL (21 de enero 2024), disponible en: <https://www.newtral.es/ia-imita-biden-deepfake-expertos/20240124/>, último acceso, 17 de enero de 2024.

12. Reglamento (UE) 2024/900 del Parlamento Europeo y del Consejo de 13 de marzo de 2024 sobre transparencia y segmentación en la publicidad política, «DOUE» n.º 900, de 20 de marzo de 2024.

En el presente trabajo, por lo tanto, se va a examinar el tratamiento específico que el RIA y el RTSPP dispensan a las grandes plataformas y a los sistemas destinados a la influencia política, valorando su regulación y examinando la articulación de las diversas regulaciones vigentes y su sistematicidad. Para ello, se examinará en primer término el *iter legis* del RIA, se analizará la diversa aproximación al tratamiento del riesgo en las normativas aplicables y finalmente se detallarán las especialidades previstas en el RIA.

II. UNA BREVE MIRADA AL «*ITER LEGIS*» DEL REGLAMENTO EN LA REGULACIÓN DE LAS GRANDES PLATAFORMAS Y SISTEMAS DE INFLUENCIA POLÍTICA

El RIA forma parte de la estrategia europea de regulación del mercado digital. Así lo indicaba en el texto inicial de la Comisión¹³, en cuya exposición de motivos hacía referencia a la necesaria coherencia entre el futuro RIA y la normativa de servicios de la Unión y la DSA (en aquel momento en fase de propuesta). En particular, por lo que se refiere a las grandes plataformas, el RIA en la posición común del Consejo Europeo detallaba en el considerando 12 y en el artículo 2.5, la aplicación del texto sin perjuicio de las disposiciones relativas a la responsabilidad de los servicios intermediarios¹⁴. Con esta escueta mención, se ventilaba la relación entre las diversas normativas, llamadas a aplicarse de forma simultánea.

Sin embargo, en el *iter legis* de la propuesta y, concretamente, en los trabajos del Parlamento Europeo se hace una especial referencia a las grandes plataformas. En efecto, en línea con la preocupación mostrada en relación con la desinformación y la gobernanza de las grandes plataformas¹⁵, la exposición de motivos del informe del Parlamento Europeo¹⁶ incorpora precisiones de interés a los efectos de este trabajo. Así, en primer término, la valoración como de alto riesgo de los sistemas de IA utilizados por candidatos o partidos para influir en los votos de las elecciones de todo nivel territorial y, junto a ellos, los sistemas de IA utilizados para contabilizar esos votos. Se destaca en sede parlamentaria el gran potencial de estos sistemas, que cuentan con capacidad para influir en un amplio número de ciudadanos de la Unión y, en último término, en el propio funcionamiento de la democracia. Asimismo, la

13. Texto disponible en: https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0008.02/DOC_1&format=PDF, p. 5, último acceso, 12 de noviembre de 2023.

14. Consejo Europeo, «Orientación general sobre la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Reglamento de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión», (6 de diciembre de 2022), disponible en: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=consil%3AST_15698_2022_INIT, último acceso, 14 de noviembre de 2023.

15. Argelich Comelles, C., «Gobernanza de las plataformas en línea ante la DSA y las propuestas de reglamento de mercados digitales e inteligencia artificial (DMA y RIA)», *ADC*, tomo LXXV, fasc. II (abril-junio 2022), pp. 501-530.

16. Parlamento Europeo, «INFORME sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión» (22 de mayo de 2023), *Vid.*: Exposición de motivos, p. 398.

exposición de motivos se refiere a la relación entre la normativa de protección de datos y de servicios digitales.

Del examen de los trabajos parlamentarios en el proceso legislativo del RIA¹⁷ no se manifiesta una correlación directa entre las aportaciones de las comisiones y el texto final en relación con el tratamiento de las grandes plataformas. No obstante, conviene destacar la labor de la Comisión de cultura y educación, en cuya justificación se hace expresa referencia no tanto a las grandes plataformas como a su actividad. Al efecto, sugiere el ponente Marcel Kolaja que se consideren sistemas de alto riesgo los utilizados por los medios de comunicación para crear o difundir artículos de noticias generadas de manera automática y las tecnologías de IA utilizadas para recomendar o clasificar contenidos audiovisuales¹⁸. Si bien su propuesta de enmienda 55, que sugería garantías de transparencia algorítmica sobre parámetros utilizados para la moderación de contenidos y personalización, no fue acogida, su aportación fue fundamental para la redacción del Punto 8 de Anexo III en la versión Parlamento.

En particular, del texto aprobado en sede parlamentaria previa al triduo destacaron dos aportaciones fundamentales en la materia que nos ocupa. Se trata de las enmiendas 739 y 740 que propusieron la modificación del punto 8 del Anexo III considerando de alto riesgo los sistemas destinados a la influencia política y los sistemas de recomendación utilizados por plataformas de gran tamaño. La conexión con la DSA resultaba pues directa y expresa, pero la acogida final ha sido diversa.

1. ATENCIÓN ESPECÍFICA DEL REGLAMENTO A LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL PARA LA INFLUENCIA POLÍTICA

El párrafo 8.1.a) bis del Anexo III del RIA versión Parlamento consideraba como sistemas de IA de alto riesgo los destinados a ser utilizados para «influir en el resultado de una elección o referéndum o en el comportamiento de voto de personas físicas en el ejercicio de su voto en elecciones o referendos». Este texto ha sido incorporado sin modificaciones en el punto 8.b) del Anexo III en la versión final del RIA.

A este respecto resulta fundamental estar a la letra del RIA que se refiere a sistemas «destinados a ser utilizados». Aquí conviene tener en cuenta una matización importante. No es lo mismo un sistema generado con un objetivo que utilizado para ese objetivo. Parece una mera sutileza, pero no lo es. Veamos un ejemplo. En la relación de prácticas prohibidas del artículo 5 del RIA, el texto se refiere específicamente a sistemas de IA que desplieguen técnicas subliminares «con el objetivo o efecto de» distorsionar el comportamiento de una persona. En este caso, el texto distingue estos dos supuestos de forma meridiana. En el caso de los sistemas de influencia política

17. Parlamento Europeo, «Enmiendas aprobadas por el Parlamento Europeo el 14 de junio de 2023 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión (COM(2021)0206 — C9-0146/2021 — 2021/0106(COD) (Procedimiento legislativo ordinario: primera lectura)», disponible en: https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_ES.html, último acceso, 15 de noviembre de 2023.

18. Accesible en: https://www.europarl.europa.eu/doceo/document/A-9-2023-0188_ES.html#_section4, pp. 452 y 453, último acceso, 14 de noviembre de 2023.

la redacción no resulta tan clara al determinar si se trata de sistemas creados con el objetivo de manipular políticamente o si incluyen sistemas que, sin haber sido creados para la manipulación política, puedan ponerse al servicio de este fin. Tal vez convendría una mayor claridad positiva al respecto, toda vez que los sistemas de IA utilizados para la influencia política abarcan el uso de técnicas de naturaleza diversa:

1. las técnicas de microsegmentación¹⁹ o focalización²⁰ políticas que sirven de base a la publicidad política comportamental²¹ y que caerían de lleno en la consideración de sistemas de alto riesgo del Anexo III;

2. la creación de perfiles falsos en redes sociales (*bots*) con un determinado perfil ideológico que genera información sintética mediante IA²². Si estos sistemas son creados para la influencia política constituyen un sistema de alto riesgo conforme al RIA;

3. la creación automatizada de noticias falsas, que puede llegar a la ultra falsificación, como ocurrió en las recientes elecciones al parlamento eslovaco²³;

4. técnicas no estrictamente dirigidas a la influencia política, pero con innegable impacto como la priorización de contenidos o los sistemas de recomendación²⁴.

A la vista del tratamiento final en el RIA de los sistemas algorítmicos de recomendación o de priorización de contenidos, la interpretación idónea del punto 8.b) del RIA parece ser la estricta, esto es la referida a sistemas creados con el objetivo específico de influir en el comportamiento electoral de los ciudadanos. Esta interpretación es además acorde con la definición de publicidad política del artículo 3.2.b) del RTSPP, en la medida en que, para ser calificada como tal, exige un doble

19. Que el Parlamento Europeo considera una forma particularmente perniciosa de publicidad digital. Mardsen, C. y Meyer, T., *op. cit.*, p. 13.

20. Vid. Comité Europeo De Protección De Datos, Directrices 8/2020 de sobre la focalización de usuarios en medios sociales, adoptadas el 13 de abril de 2021, disponible en: https://edpb.europa.eu/system/files/2021-11/edpb_guidelines_082020_on_the_targeting_of_social_media_users_es_0.pdf, último acceso, 20 de febrero de 2024.

21. Comité Europeo De Protección De Datos, «EDPB Urgent Binding Decision on processing of personal data for behavioural advertising by Meta», *Sala de prensa* (1 de noviembre de 2023), disponible en: https://edpb.europa.eu/news/news/2023/edpb-urgent-binding-decision-processing-personal-data-behavioural-advertising-meta_en, último acceso, 20 de febrero de 2024.

22. Panditharatne, M., «Cómo la inteligencia artificial pone en riesgo las elecciones y las medidas que se requieren para protegernos», *Brennan Center-Análisis* (21 junio de 2023, actualizado el 13 julio de 2023), disponible en: <https://www.brennancenter.org/es/our-work/analysis-opinion/inteligencia-artificial-pone-en-riesgo-elecciones-medidas-protoger-democracia>, último acceso, 22 de febrero de 2024.

23. Cuyo impacto en la derrota de Michal Šimečka frente al candidato prorruso Robert Fico todavía está por determinar. Vid. Solon, O., «Trolls in Slovakian Election Tap AI Deepfakes to Spread Disinfo», *Bloomberg News* (29 de septiembre de 2023), disponible en: <https://www.bloomberg.com/news/articles/2023-09-29/trolls-in-slovakian-election-tap-ai-deepfakes-to-spread-disinfo>, último acceso, 22 de febrero de 2024.

24. Por ejemplo, sistemas de personalización de mensajes políticos, que parten de técnicas de segmentación y adaptación de discursos mediante IA y que alcanzan a los usuarios a través de los sistemas de recomendación, generando el riesgo de generar burbujas informativas.

requisito: que pueda influir en el resultado o en el comportamiento electoral «y esté diseñada para ello».

En todo caso, parece innegable que el RIA se preocupa por ligar el concepto de influencia política a la formación de la opinión política y a la protección de los derechos fundamentales de participación política de los destinatarios finales de los sistemas de IA. Esta voluntad deriva de forma necesaria del inciso final del punto 8.b) del Anexo III, que excluye de la consideración de alto riesgo a los sistemas a cuya información de salida no están directamente expuestas las personas físicas, como los sistemas de IA de gestión logística y administrativa de las campañas políticas. Por ejemplo, sistemas utilizados como auxilio en la financiación o diseño de campañas políticas (asesores algorítmicos de campaña)²⁵.

En consecuencia, el tratamiento de los sistemas destinados a la influencia política exige tener en cuenta la aplicación sistemática de la normativa vigente. En particular, la aplicación de las garantías de transparencia algorítmica que propone el RIA para sistemas de alto riesgo en relación con:

- i) el RGPD, toda vez que los datos constituyen «el arma más poderosa» para generar confrontación política a través del perfilado y la personalización informativa²⁶;
- ii) la normativa sectorial; específicamente, las exigencias de transparencia de plataformas en línea de la DSA y el RTSPP, al que se refiere el considerando 62 del RIA al declarar la aplicación conjunta de ambas normativas. Si bien, debe tenerse en cuenta que el ámbito de aplicación del RTSPP no se limita a la publicidad política emitida en plataformas o buscadores, sino que alcanza a todos los sujetos que trabajan en el proceso de preparación, inserción, promoción, publicación y difusión de publicidad política (proveedores o editores de publicidad política y servicios conexos²⁷).

2. LOS SISTEMAS DE RECOMENDACIÓN ALGORÍTMICA: TRATAMIENTO EN LA PROPUESTA DE INTELIGENCIA ARTIFICIAL VERSIÓN PARLAMENTO Y SU FUNDAMENTACIÓN

La propuesta de RIA definía como sistema de alto riesgo los sistemas algorítmicos de recomendación de plataformas de gran tamaño. El fundamento básico de esta decisión descansaba en el impacto de estos sistemas en los derechos fundamentales, circunstancia que la propia exposición de motivos identificaba como criterio «de especial relevancia» para su clasificación como de alto riesgo. Asimismo, y atendiendo al volumen de usuarios de las plataformas digitales de gran tamaño, el considerando 40 ter del RIA destacaba su potencial influencia «en la seguridad en línea, la configuración de la opinión y el discurso públicos, en los procesos electorales y democráticos y en las preocupaciones sociales» como justificación para su calificación como sistema de alto riesgo.

-
- 25. Scheiner, B., «Seis formas en que la IA podría cambiar la política», *MIT Technology Review* (7 de agosto 2023), disponible en: <https://www.technologyreview.es/s/15580/seis-formas-en-que-la-ia-podria-cambiar-la-politica>, último acceso, 24 de febrero de 2024.
 - 26. García Mahamut, R., «Elecciones, protección de datos y transparencia en la publicidad política: la apuesta normativa de la UE y sus efectos en el ordenamiento español», *Revista Española de Transparencia*, n.º 17 extraordinario (2023), p. 78.
 - 27. *Ibidem*, p. 4.

Como se ha advertido *supra*, las plataformas estructuran su actividad sobre la personalización, que se concreta en la capacidad de la plataforma para recomendar a sus usuarios contenido específico creado por los propios usuarios de la red. Sin embargo, como se deduce del considerando 70 de la DSA, el legislador europeo manifiesta un doble concepto de sistema de recomendación: i) el estricto referido a esta actividad de propuesta o sugerencia y ii) un concepto amplio, que también abarcaría otras técnicas como la clasificación y priorización algorítmica de la información, la distinción de texto y otras formas visuales o la organización personalizada de la información. Precisamente en este sentido amplio, el Supervisor Europeo de Protección de Datos recordaba que algunas de estas técnicas, como el perfilado o microsegmentación pueden afectar significativamente los derechos fundamentales²⁸.

A pesar de estas consideraciones, la clasificación como de alto riesgo de los sistemas de recomendación de plataformas digitales no ha sido finalmente acogida en el RIA. En esta decisión, el legislador ha tenido en cuenta la relación entre el RIA y la DSA, que constituye la norma de referencia respecto al tratamiento de los servicios digitales. Corresponde, por lo tanto, analizar la intersección entre ambas normas a los efectos de valorar el tratamiento de las plataformas y sistemas de IA utilizados por ellas en Derecho europeo y la decisión final del legislador respecto a su tratamiento específico en el RIA.

III. LA LÓGICA DEL RIESGO EN LAS PLATAFORMAS DE GRAN TAMAÑO: LA COMPLEMENTARIEDAD ENTRE LA DSA Y EL REGLAMENTO

La actual tendencia en el tratamiento de las plataformas de gran tamaño por el legislador europeo parte de la lógica del riesgo, esto es, un enfoque que consiste en adaptar los derechos y obligaciones a los riesgos que derivan de una cierta actividad²⁹. Adoptada esta perspectiva originariamente por el RGPD, el cumplimiento basado en el riesgo inspira las normas clave del Derecho digital europeo, en particular, el paquete normativo y el RIA, si bien desde una perspectiva distinta.

En efecto, sin desdeñar los innegables beneficios que las grandes plataformas conllevan para los ciudadanos, lo cierto es que su uso indebido puede generar un profundo impacto en los derechos fundamentales y los sistemas democráticos que conviene abordar³⁰, habida cuenta de la mutación digital de los mercados de información citados *supra*. El tratamiento del riesgo en la DSA parte también, si

28. European Data Protection Supervisor, «Opinion 3/2018 EDPS Opinion on online manipulation and personal data» (19 de marzo de 2018), p. 9, disponible en: https://edps.europa.eu/sites/edp/files/publication/18-03-19_online_manipulation_en.pdf, último acceso, 21 de noviembre de 2023.

29. Barrio Andrés, M., «El cumplimiento basado en el riesgo o *risk based compliance*, pieza cardinal del nuevo Derecho digital europeo», *Análisis del Real Instituto Elcano (ARI)*, n.º 34 (2023), p. 3.

30. Se trata de trascender el examen estrictamente tecnológico para evaluar cualitativamente el impacto de la actividad de estas plataformas en los derechos de los particulares (*Risk to rights approach*, tomada de la normativa sobre protección de datos), *vid.* MAHLER, T., «Between risk management and proportionality: The risk-based approach in the EU's Artificial Intelligence Act Proposal», *Nordic Yearbook of Law and Informatics* (September 2021), disponible en: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4001444, p. 259, último acceso, 17 de enero de 2023.

bien de forma indirecta, de una cierta escala de riesgos que se articula sobre la base de una exigencia cumulativa de obligaciones. Al efecto distingue cinco escalas de obligaciones en su capítulo III:

1. las generales referidas a los prestadores de servicios intermediarios, en la sección primera (artículos 11 a 15);
2. las disposiciones adicionales aplicables a las plataformas en línea, entre otros, los servicios de alojamiento (sección 2ª, artículos 16 a 18);
3. las obligaciones de las plataformas en línea (sección 3ª, artículos 19 a 28);
4. como especialidad en materia de protección de consumidores, las disposiciones adicionales de la sección cuarta (artículos 29 a 32) relativas a las plataformas en línea que permitan los negocios B2C (*Business to Consumer*);
5. obligaciones de las plataformas de gran tamaño (sección quinta, artículos 33 a 43).

La perspectiva regulatoria de la DSA es asimétrica, al exigir el cumplimiento de obligaciones adicionales de información, transparencia y *accountability* por las plataformas y buscadores de gran tamaño (en lo sucesivo VLOPs). Éstas se definen en el artículo 33 como aquellas que cuentan con más de 45 millones de usuarios activos al mes en la Unión europea (o un 10% de la población de la Unión). En virtud del mandato del apartado 4 de este precepto, en abril de 2023 la Comisión Europea adoptó la decisión por la que designaba las plataformas de gran tamaño, con no poca polémica³¹.

El considerando 75 de la DSA justifica esta decisión en el número de destinatarios del servicio y la posición central de estos sistemas para facilitar el ejercicio de la libertad de expresión e información y la conformación de la opinión pública. La DSA se preocupa por fundamentar la proporcionalidad de estas medidas precisamente en la valoración de los riesgos derivados de las VLOPs (considerando 76) correlacionando la intensidad de las medidas con el impacto social de este tipo plataformas que, en último término, disponen de los recursos necesarios para realizar una evaluación de los riesgos que originan (*ex artículo 34*) y afrontar sus consecuencias³².

Estos riesgos se han venido a definir como riesgos sistémicos, que derivan del diseño o del funcionamiento del servicio «y de los sistemas relacionados con éste, incluidos los sistemas algorítmicos». El riesgo sistémico remite a una perspectiva holística en la medida en que los riesgos para la salud humana, el medio ambiente o los derechos fundamentales se integran en un contexto más amplio de riesgos y

-
31. Las plataformas designadas han mostrado su disconformidad. El listado está disponible en: https://ec.europa.eu/commission/presscorner/detail/es/ip_23_2413 Incluso se han viralizado rumores sobre las reticencias de algunas redes de cumplir con las exigencias de la DSA, si bien en referencia a X (antiguo Twitter) han sido desmentidas: *Vid.* <https://www.lavanguardia.com/tecnologia/20231023/9320565/elon-musk-desmiente-rumores-eliminar-twitter-paises-ue.html>, último acceso, 12 de noviembre de 2023.
 32. Castelló Pastor, J. J., «Nuevo régimen de responsabilidad de los servicios digitales que actúan como intermediarios a la luz de la propuesta de Reglamento relativo a un mercado único de servicios digitales», en Castelló Pastor, J. J. (dir.), *Desafíos jurídicos ante la integración digital: aspectos europeos e internacionales*, Aranzadi-Thomson Reuters, Cizur Menor (Navarra) 2022, p. 73.

oportunidades sociales, financieras y económicos, combinando fenómenos naturales, evolución socio económica, tecnología y actuaciones políticas multinivel³³. En este sentido define también el artículo 1.44. del RIA el riesgo sistémico en referencia al impacto significativo que los modelos de uso general (en adelante, MUG) pueden tener sobre el mercado interior, en particular, a los «efectos negativos reales o razonablemente previsibles en la salud pública, la seguridad, la seguridad pública, los derechos fundamentales o la sociedad en su conjunto, que puede propagarse a escala a través de la cadena de valor».

Por tanto, un análisis de los riesgos sistémicos requiere un triple proceso de identificación, evaluación y gestión de riesgos desde una perspectiva multidisciplinar que permita analizar las interdependencias y relaciones entre diversos grupos de riesgo. Respecto a las VLOPs el artículo 34 de la DSA exige que evalúen los siguientes riesgos sistémicos:

- i) los derivados de la difusión de contenidos ilícitos; por ejemplo, referido al discurso del odio y la desinformación;
- ii) los efectos reales y previsibles del servicio para el ejercicio de los derechos fundamentales. El texto expresamente se refiere a los riesgos derivados del diseño de sistemas algorítmicos de VLOPs destinados a limitar la libertad de expresión (moderación automatizada de contenidos);
- iii) los efectos reales o previsibles sobre los procesos democráticos y electorales y la seguridad pública;
- iv) la afectación real o previsible de la salud pública, los menores, efectos negativos sobre el bienestar físico y mental de las personas o violencia de género.

Esta enumeración no constituye un *numerus clausus*, en la medida en que concreta la cláusula general del inciso primero que exige la detección, el análisis y la evaluación «de cualquier riesgo sistémico» en la Unión derivado del diseño o funcionamiento de su servicio y los sistemas utilizados, con especial referencia a los sistemas algorítmicos.

Puede observarse que los riesgos sistémicos identificados en la DSA parten de una postura garantista, de un refuerzo de la posición de la persona física destinatario del servicio. Conviene advertir en este punto el diferente enfoque de la DSA y el RIA. Mientras la DSA está orientada a las garantías respecto a los usuarios de las plataformas que, en último término, son los destinatarios finales de los sistemas de IA por ellas utilizados, el RIA se articula sobre el protagonismo de proveedores, responsables del despliegue, importadores y distribuidores de sistemas que no tienen por qué ser sus destinatarios finales³⁴. Estas aproximaciones resultan complementarias en la valoración global de los riesgos subyacentes al uso de sistemas de IA en plataformas digitales.

33. Renn, O. y Klinke, A., «Systemic risks: a new challenge for risk management», *EMBO Reports*, Volumen 5, Special Issue (octubre 2004), p. 41, disponible en: <https://doi.org/10.1038/sj.embor.7400227>, último acceso, 27 de febrero de 2024.

34. Jiménez-Castellanos Ballesteros, I., «Decisiones automatizadas y transparencia administrativa: nuevos retos para los derechos fundamentales», *Revista Española de la Transparencia*, n.º 16 (2023), pp. 202-203.

En efecto, la perspectiva general del riesgo que inspira la DSA difiere del esquema del RIA. La normativa en materia de IA no deja la evaluación del riesgo a los sujetos obligados, sino que impone un análisis legislativo del riesgo acompañado de sistemas de gestión por los obligados. Así pues, este *risk-based approach* en la lógica del RIA se ha venido a considerar más bien como una técnica legislativa para acotar el ámbito de aplicación del reglamento y asegurar la proporcionalidad legislativa³⁵. En todo caso, la solución propuesta en la versión Parlamento suponía incorporar la protección del usuario persona física en la lógica de una norma centrada en la plataforma como implementador/desarrollador. La necesidad de traer a la norma esta perspectiva parece absolutamente adecuada, en la medida en que responde al objetivo primario del artículo 1 del RIA. El usuario de VLOPs por derecho propio, y con fundamento también axiológico, bien merece ser tenido en cuenta por la normativa que regula los sistemas de IA. El matiz es de tintes sistemáticos y no sólo en relación con el propio RIA sino, más allá, en la deseable sistematicidad del Derecho digital europeo. Por tanto, en román paladino, la pregunta no es si los usuarios deben ser tenidos en cuenta, sino dónde y cómo.

El hecho de que en el RIA decaiga la calificación como de alto riesgo de los sistemas de recomendación algorítmica de VLOPs podría considerarse *a priori* una pérdida de centralidad en la persona, una suerte de traición al objetivo basal del reglamento. Sin embargo, un examen sistemático de la norma en relación con las exigencias de la DSA indica que este tratamiento resulta, a la postre, más congruente con el sistema europeo de Derecho digital en la consideración de la intensidad de los riesgos generados por VLOPs.

En efecto, no puede dejar de reconocerse que el legislador europeo desde la versión del RIA de la Comisión, tuvo presente al destinatario final de los sistemas de IA en el original artículo 52 al regular las obligaciones de transparencia para proveedores o implementadores de sistemas de IA «destinados a interactuar directamente con personas físicas». Las obligaciones, centradas en la transparencia algorítmica, buscaban que el destinatario final fuera informado de que estaba interactuando con un sistema de IA.

Junto a estas previsiones, el texto final ha afinado el concepto de MUG, incorporando una suerte de sistema intermedio entre éste y el alto riesgo. Se trata de los modelos de uso general con riesgo sistémico (MUGRS), previstos en el artículo 51, de nueva factura. Estos modelos son declarados como tales por la Comisión europea:

— en base a sus capacidades de alto impacto, que debe evaluarse haciendo uso de herramientas o metodologías apropiadas, conforme a los criterios del Anexo XIII.

— o dependiendo de su capacidad técnica de cálculo o potencia de procesamiento. El artículo 51.2 concreta este criterio al definir como de riesgo sistémico los modelos cuya cantidad acumulada de cálculo utilizada para su entrenamiento medida en operaciones en coma flotante (FLOPs) sea superior a 10^{25} . Aquí se incluiría, por ejemplo, Chat GPT-4.

— Sin embargo, el apartado 3 permite a la Comisión adecuar estos umbrales mediante actos de conformidad con el fin de atender a la evolución tecnológica.

35. Mahler, T., *op. cit.*, p. 247.

En el supuesto de ser calificados como MUGRS, los modelos de IA deben cumplir algunas de las obligaciones de los sistemas de alto riesgo, en particular: i) llevar a cabo una evaluación del modelo de conformidad con protocolos y herramientas normalizadas; ii) evaluar y mitigar los riesgos sistémicos a escala de la Unión Europea; iii) hacer un seguimiento documentado de incidentes graves y medidas correctivas, comunicando estas circunstancias sin demora indebida a la Oficina de IA; y iv) garantizar un nivel adecuado de protección de ciberseguridad para el modelo

Lo cierto es que el precepto resulta suficientemente flexible para adecuar la norma a la evolución tecnológica. Pero más allá de abrazar la contingencia técnica, la clave para analizar su aplicabilidad a las VLOPs pasa por comprender la incidencia de esta nueva categoría en el sistema de riesgos del RIA. Debe tenerse en cuenta que la norma habla de capacidad de alto impacto, no del impacto efectivamente producido. Es decir, está valorando riesgos estrictamente y sometiéndolo a un concreto tipo de riesgo a unas obligaciones específicas. La matización del original sistema de gradación de riesgos, que en el texto final se disfraza de especificación en la clasificación de MUG, afecta de forma directa a las grandes plataformas. Así pues, la aplicación conjunta de ambas regulaciones parece desembocar en la cuestión de si una plataforma VLOPs puede ser incorporada en la categoría de MUGRS por la Comisión Europea. Sin embargo, un examen sosegado de esta cuestión invita a reformular el planteamiento.

En este sentido, conviene traer a colación el considerando 118 del RIA que, atendidas las obligaciones impuestas por la DSA, considera que deben entenderse cumplidas las obligaciones del RIA salvo que se identifiquen riesgos sistémicos no cubiertos por la DSA. Por tanto, en este punto no se trata tanto de seguir la lógica confesa del artículo 51, esto es, la posible clasificación de los modelos de IA que pueden incorporar las VLOPs como MUGRS, sino de evaluar los riesgos verdaderamente cubiertos por las plataformas en cumplimiento del modelo de gestión de riesgos de la DSA en relación con el RIA. Sobre este tema se volverá en el próximo apartado al examinar las especialidades de la aplicación del RIA en materia de gestión de riesgos.

IV. ESPECIALIDADES DE LA APLICACIÓN DEL REGLAMENTO EN LAS GRANDES PLATAFORMAS Y SISTEMAS DE IA PARA LA INFLUENCIA POLÍTICA

Por lo que se refiere a la aplicación del RIA en el sector de las plataformas digitales, debemos partir del considerando 118 del RIA que reconoce la complementariedad con la DSA respecto a las obligaciones impuestas a los proveedores de servicios de intermediación. Por lo tanto, partiendo de la aplicación conjunta de ambas normativas, la cuestión clave radica en diseccionar las garantías que la DSA impone a los sistemas de IA utilizados por las plataformas y examinar su impacto sobre el RIA. Asimismo, se tendrán en cuenta las exigencias de transparencia del RTSPD respecto a los sistemas para la influencia política.

El paquete obligacional que la DSA impone en su relación con el RIA parte de un deber general de cumplimiento teniendo en cuenta los siguientes aspectos: el estado de la técnica generalmente reconocido, la finalidad del sistema, los usos indebidos previsibles y el sistema de riesgos. A partir de este esquema general, su aplicación

conjunta con el RIA se condiciona por el ámbito subjetivo de aplicación de ambas normas. Así, conforme a la terminología del artículo 2.1 del RIA, las plataformas pueden ser proveedoras³⁶ o responsables del despliegue³⁷ de los sistemas de IA que contraten para prestar su servicio. A partir de esta idea, puede examinarse la aplicación conjunta del RIA, la DSA y el RTSPP atendiendo a las garantías que la DSA impone a los sistemas de IA utilizados por VLOPs y teniendo en cuenta las especialidades derivadas del uso de sistemas de IA para la influencia política por las plataformas. Estas garantías pueden clasificarse en cuatro bloques: i) garantías de gestión de riesgos; ii) garantías de transparencia algorítmica; iii) garantías procedimentales y iv) garantías orgánicas.

1. GARANTÍAS EN MATERIA DE GESTIÓN DE RIESGOS

Respecto al sistema de gestión de riesgos, como ya se ha apuntado, el RIA somete los sistemas para la influencia política al procedimiento de gestión de riesgos de su artículo 9 en tanto que constituyen sistemas de alto riesgo. Asimismo, el RTSPP tiene en cuenta el amplio espectro de servicios de publicidad política que pueden generar riesgos. Y así, en la medida en que las VLOPs presten sus servicios en calidad de editores de dicha publicidad, se someten al sistema de gestión de riesgos de la DSA por remisión expresa del considerando 46 del RTSPP. El texto no deslinda claramente los supuestos, pero una lectura sistemática de las tres normas permite entender esta remisión a la DSA referida a los servicios de publicidad política no considerados de alto riesgo por el RIA (por ejemplo, si no utilizan sistemas de IA).

Por su parte, y respecto a sistemas o modelos de IA integrados en VLOPs, el RIA remite al marco de gestión de riesgos de la DSA y, a tal efecto, presume que se cumplen las obligaciones del RIA salvo que surjan o se identifiquen riesgos sistémicos significativos no cubiertos por la DSA. Esta remisión constituye una presunción *iuris tantum* de cumplimiento del estándar impuesto por el RIA, sometiéndose en caso de cobertura insuficiente a las obligaciones intensificadas del artículo 55. Por tanto, lo significativo no es la calificación de los modelos de IA incorporados en las VLOPs como MUGRS, sino la adecuada cobertura de riesgos en aplicación del sistema de gestión de la DSA y el RIA.

En este sentido, el considerando 118 del RIA exige a las VLOPs que evalúen los posibles riesgos sistémicos derivados del diseño, funcionamiento y uso de sus servicios y, en particular, extiende esta evaluación a: i) los sistemas algorítmicos utilizados en el servicio que puedan contribuir a estos riesgos y ii) los riesgos sistémicos que deriven de posibles usos indebidos. La conexión con la sección 5ª de la DSA es clara y, en especial, con el artículo 34 cuyo contenido reproduce. Este deber de evaluación de riesgos surge en un momento muy concreto: cuando se designe

36. Es el caso de Meta y sus plataformas, *vid.* <https://ai.meta.com/blog/powered-by-ai-instagram-explore-recommender-system/>, último acceso, 28 de febrero de 2024.

37. El RIA versión Consejo amplía el ámbito obligacional de los proveedores a los usuarios de los sistemas de IA, una figura que en la versión Parlamento se redefine como implementador. El texto definitivo, matiza el ámbito subjetivo refiriéndose a proveedores y responsables del despliegue, sin perjuicio de su aplicación a otros sujetos como importadores o distribuidores.

como VLOP y en todo caso una vez al año o antes de desplegar funcionalidades que puedan tener impacto crítico sobre los riesgos.

En la evaluación de los riesgos debe tenerse en cuenta la finalidad del sistema. Y al respecto, matiza la DSA los factores a evaluar, que incluyen: i) el diseño de sistemas de recomendación; ii) sistemas de moderación de contenidos; iii) condiciones generales aplicables y su ejecución; iv) sistemas de selección y presentación de anuncios; v) efectos de los servicios de publicidad política, por remisión del RTSP, ya citada y vi) prácticas del prestador relacionadas con los datos.

Junto a estas cuestiones, y en línea con la versión parlamento, la DSA exige el análisis de ciertos usos indebidos, en particular, los efectos derivados de la manipulación del servicio, por ejemplo, el uso no auténtico (*bots*) o explotación automatizada. Asimismo, también deben valorarse los efectos perniciosos de usos adecuados de la IA cuando la ilicitud radica en otros aspectos del servicio, como la amplificación y difusión potencialmente rápida y amplia de contenido ilícito (viralización) o información incompatible con las condiciones generales. Esta misma lógica se aplica a prácticas de publicidad segmentada y otras técnicas de limitadas por el artículo 18 del RTSP.

Como viene sosteniéndose, la lógica del RIA es preventiva, por lo tanto, la evaluación de riesgos exige la correlativa adopción de medidas «adecuadas y específicas» para minimizar el riesgo y facilitar el cumplimiento adecuado y proporcionado de los requisitos del capítulo II. Esta misma lógica se adopta por el artículo 35 de la DSA, que exige a las VLOPs la adopción de medidas de reducción de riesgos razonables, proporcionadas y efectivas, adaptadas a los riesgos sistemáticos y teniendo en cuenta las consecuencias de dichas medidas sobre los derechos fundamentales. Estas medidas son objeto de supervisión por la Comisión europea con el inestimable apoyo técnico del Centro Europeo para la Transparencia Algorítmica (ECAT).

Por lo tanto, se observa en ambos textos una visión sistemática de las medidas de corrección, que operan en dos momentos: i) desde el diseño y desarrollo del sistema y ii) una vez diseñado, como mecanismos de control y mitigación de riesgos no eliminables (piénsese en la difusión de publicidad electoral en jornadas de reflexión), todo ello acompañado de la necesaria transparencia algorítmica y alfabetización de los responsables del despliegue del sistema.

Esta visión sistemática de los mecanismos de corrección invita, en el contexto de la DSA, a tener en cuenta la aplicación de medidas específicas. Por ejemplo, la realización de pruebas de sistemas algorítmicos del artículo 35 de la DSA³⁸, cuyo escrutinio concreta artículo 40 al exigir a las plataformas el deber de explicar a la Comisión o al coordinador de servicios digitales el diseño, lógica, funcionamiento y realización de las pruebas de sus sistemas algorítmicos. Por su parte, una vez implantado el sistema, la adaptación de los sistemas algorítmicos incluidos los sistemas de recomendación.

El artículo 34 de la DSA exige a las grandes plataformas el deber de conservación de los documentos justificativos de las evaluaciones de riesgos durante al menos tres

38. Exigencia que el RIA descansa en los sistemas de alto riesgo ex artículo 9, párrafos 5 a 7.

años. El plazo es sustancialmente inferior al de 10 años previsto para los sistemas de alto riesgo del artículo 18 del RIA. Habida cuenta de que el RIA no establece ningún plazo mínimo de conservación para los MUG este plazo intermedio, que responde a la visión sistémica del riesgo, parece razonable en términos de sistemática regulatoria.

2. GARANTÍAS EN MATERIA DE TRANSPARENCIA ALGORÍTMICA: EXPLICABILIDAD VS. OPACIDAD

Respecto a las obligaciones de transparencia algorítmica debe tenerse en cuenta el diferente foco en términos subjetivos que preside el RIA y la normativa sectorial que analizamos. Por tanto, su examen debe abordarse desde una perspectiva amplia, en términos de cadena de valor, centrada el RIA en los proveedores/responsables del despliegue y la DSA y RTSPP en los destinatarios finales. En esta circunstancia radica la adecuada articulación de la transparencia como garantía última de la explicabilidad de los sistemas de IA utilizados por las VLOPs en el Derecho digital europeo.

En efecto, la DSA y el RTSPP se centran en el usuario final de la plataforma o receptor de la publicidad, sin embargo, esta perspectiva también se aborda por el artículo 50 del RIA que establece obligaciones de transparencia para proveedores y usuarios de sistemas de IA destinados a interactuar con personas físicas. Las garantías de transparencia algorítmica de la DSA parten del artículo 14 que prevé una suerte de transparencia algorítmica contractual al exigir que las plataformas incluyan en sus condiciones generales o de prestación del servicio información sobre toma de decisiones automatizadas. Esta exigencia puede verse reflejada en el artículo 50.2 del RIA al exigir que, a más tardar, la información relativa al uso de algoritmos en redes se facilite a las personas físicas con ocasión de la primera interacción o exposición. La explicabilidad del sistema se garantiza por el artículo 50.5 en último término con: i) la accesibilidad de la información en virtud del último inciso del precepto, completado con la referencia a la explicabilidad adaptada a los menores del artículo 14.3 de la DSA; ii) una adecuada redacción que el RIA define en términos de expresión «clara y distinguible»³⁹; iii) las medidas de alfabetización del artículo 4 del RIA.

Una concreción de esta garantía de transparencia algorítmica se recoge en el artículo 27 de la DSA que exige a las VLOPs que incluyan en las condiciones generales información relativa a los parámetros principales utilizados en sus sistemas de recomendación y las opciones que la plataforma pone a disposición de los destinatarios del servicio para modificar o influir en estos parámetros. Esta información se relaciona con la explicabilidad del sistema ya que implica la motivación de la decisión, es decir, la explicación de por qué se sugiere un determinado contenido. El artículo 27.2 guía a las plataformas en el cumplimiento de esta obligación identificando dos parámetros mínimos: i) los criterios más significativos a la hora de determinar la información sugerida al destinatario del servicio, ii) las razones de la importancia relativa de dichos parámetros.

Ahondando en la protección del destinatario final, el artículo 52.3 del RIA regula el supuesto específico de las ultrafalsificaciones, conocidas como *deep fakes*, que pueden

39. El artículo 14.6 de la DSA exige la publicación de las condiciones generales en todas las lenguas oficiales de todos los Estados miembros en los que la VLOPs preste sus servicios.

ser utilizadas, en su caso, como elementos para la influencia política. Estas técnicas en la medida en que no nacen estrictamente con la finalidad prevista en el punto 8.b) del anexo III, también se consideran *a priori* de riesgo limitado si bien el RIA los somete a obligaciones de transparencia adicionales. En concreto, para las VLOPs la técnica del etiquetado (*flagging*) de forma que los usuarios puedan conocer que se trata de una falsificación. Asimismo, el inciso segundo se refiere a las noticias falsas (texto manipulado con el fin de informar al público sobre asuntos de interés público). En este caso, el etiquetado constituye la garantía de transparencia clave en la protección de los usuarios finales de los servicios de VLOPs.

En el contexto específico de la publicidad, la exigencia de etiquetado del RIA se suma a la estrictamente identificativa de la naturaleza política del anuncio que recoge el artículo 11 del RTSPP y el artículo 26.1.a) de la DSA. Por su parte, el artículo 26.1.d) de la DSA exige que las VLOPs ofrezcan de forma sencilla, información significativa, accesible y directa desde los propios anuncios acerca de los principales parámetros utilizados para determinar el destinatario del anuncio y, en su caso, cómo cambiar esos parámetros. Sin embargo, en lo que aquí nos interesa, el artículo 26.3 prohíbe a las plataformas en línea la presentación de publicidad política segmentada, es decir, basada en la elaboración de perfiles⁴⁰ a partir de datos personales que revelen opiniones políticas o las demás categorías especiales de datos del artículo 9 del RGPD. Aquí deben tenerse en cuenta las garantías de transparencia previas en los artículos 6 y siguientes del RTSPP, en especial, de los artículos 7, 11, 12 y 19 y los requisitos para la segmentación de la publicidad política del artículo 18 RTSPP así como el deber de conservación de los anuncios políticos y avisos de transparencia y sus modificaciones, que el RTSPP alarga hasta los 7 años desde la entrega o difusión del anuncio (artículo 9.3) o desde la última publicación del aviso (artículo 12.4).

Por su parte, el artículo 15.1 de la DSA consagra las obligaciones de transparencia informativa de los proveedores de servicios intermediarios en términos de *accountability*, y hace una triple referencia a la transparencia de los sistemas de IA, al exigir una rendición de cuentas mediante la emisión de informes en los que se incluya, entre otras cuestiones, información relativa a:

- i) el número de notificaciones tratadas únicamente por medios automatizados y el tiempo medio necesario para adoptar medidas (apartado b);
- ii) la moderación automatizada de contenidos y el tipo de medidas adoptadas que afecten a la disponibilidad, visibilidad y accesibilidad de la información proporcionada por los destinatarios del servicio y otras restricciones conexas (apartado c). Aquí se incluyen los sistemas de recomendación que utilizan sistemas de IA, que deben quedar identificados por disposición del precepto;
- iii) el uso de medios automatizados con fines de moderación de contenidos (apartado e). El propio precepto concreta la información mínima que debe publicarse: descripción cualitativa, especificación de los fines precisos, indicadores de la precisión y posible tasa de error de los medios automatizados empleados para cumplir dichos fines, y las salvaguardias aplicadas. Respecto a las VLOPs, el artículo 42.2 exige que

40. El artículo 4.4. del RGPD define la elaboración de perfiles como «toda forma de tratamiento automatizado de datos personales consistente en utilizar datos personales para evaluar determinados aspectos personales de una persona física», en particular, para este caso, para analizar o predecir el sentido de su voto.

estos indicadores de precisión e información conexa se desglosen por cada lengua oficial de los Estados miembros. Junto con el artículo 42, las previsiones del artículo 15 se completan para las VLOPs con las del artículo 24.1 que, en lo que aquí nos interesa, se refiere entre otras cuestiones a la publicación del número de suspensiones de cuentas, por ejemplo, *bots* utilizados como mecanismo de influencia política. En este ámbito, la transparencia *ex post* se completa con el artículo 14 del RTSPP en lo referido a la información relativa al uso de técnicas de segmentación.

Respecto a la temporalidad de los informes, la cadencia anual que, con carácter general reconoce el artículo 15.1 de la DSA se matiza en el artículo 42.1 de la DSA al exigir su publicación a los seis meses de la notificación a la plataforma de su designación como VLOP y, una vez cumplida ésta, al menos con cadencia semestral. El deber de rendición de cuentas para VLOPs se concreta en el deber de remisión a la Comisión y al coordinador de servicios digitales de los informes del artículo 42.4, en particular, los resultados de la evaluación de riesgos y las medidas de reducción específicas, el informe de auditoría y el de aplicación de la auditoría y, en su caso, informe de las consultas que ha realizado el prestador en apoyo a las evaluaciones de riesgo y el diseño de medidas de reducción. Estos informes estarán a disposición del público salvo solicitud motivada del perjuicio que pudiera provocar a la plataforma su plena disponibilidad de acceso. Esta obligación viene por lo tanto a reforzar las garantías del usuario del sistema y del destinatario final de VLOPs sin necesidad de considerarlas como sistemas de alto riesgo.

Una manifestación específica de la transparencia algorítmica *ex post* en el ámbito de la publicidad descansa en el deber de publicación y actualización de un repositorio en el que se incluya la información básica y fácilmente accesible sobre los anuncios publicitados en las plataformas, de acuerdo con el contenido del artículo 39.2 de la DSA y 13 del RTSPP. En particular, deben hacerse públicos los parámetros de personalización de los anuncios (que pueden hacer uso de sistemas de IA), esto es, los criterios utilizados para la presentación o exclusión del anuncio a determinados usuarios. En todo caso, este deber de información nace con la presentación del anuncio en plataforma y se mantiene hasta un año después de la última vez en que se presentó.

Estas medidas de transparencia algorítmica llevan consigo el derecho de los destinatarios del servicio, es decir, de los usuarios de las plataformas, a conocer que están interactuando con un sistema de IA. Este derecho se reconoce de forma más rotunda en el artículo 50.1 del RIA que exige esta garantía de transparencia desde el diseño o desarrollo del sistema. Es notable la matización que incorporó la versión parlamento al descansar este deber de información no solo en el sistema sino también en el proveedor o usuario. De esta forma, las plataformas (ya se considerasen proveedoras, usuarias o implementadoras del sistema) estaban llamadas a este deber de comunicación⁴¹ a las personas físicas usuarios de sus servicios. Este matiz decae en el texto final del RIA, que se refiere tan solo a los proveedores del sistema y exceptiona el deber de información cuando resulte evidente atendidas

41. Comisión Europea, Dirección General de Redes de Comunicación, Contenido y Tecnologías, «Directrices éticas para una IA fiable», *Oficina de Publicaciones* (2019), p. 22, disponible en: <https://data.europa.eu/doi/10.2759/14078>, último acceso, 3 de marzo de 2024.

las circunstancias «desde el punto de vista de una persona física razonablemente informada, atenta y perspicaz». Esta redacción no supone una reducción efectiva de las garantías de transparencia de los usuarios de VLOPs en la medida en que el deber de transparencia contractual del artículo 14 de la DSA y las obligaciones de transparencia algorítmica del artículo 27 de la DSA, someten a la plataforma que no actúe como proveedora a este deber de información.

3. GARANTÍAS PROCEDIMENTALES

En la línea impuesta por el Derecho internacional, la DSA parte de una posición de exención de responsabilidad de las plataformas por los contenidos ilícitos subidos por sus usuarios salvo conocimiento efectivo de la infracción (*safe harbour*) si bien reconoce la cláusula de buen samaritano. En consecuencia, la propia norma admite la posibilidad de que las plataformas puedan adoptar diversas medidas contra usos indebidos, ya se trate de: i) bloqueo, relegación de información, ii) suspensión o cese del servicio para ciertos usuarios, iii) suspensión o cese de cuentas o iv) suspensión, cesación o restricción de la monetización de cuentas, conforme a los artículos 3.t) y 17). Estas decisiones pueden adoptarse de forma automatizada mediante el uso de sistemas de IA que deberán cumplir las exigencias de los MUG del RIA. Asimismo, su actividad se someterá a auditorías externas independientes con cadencia, al menos, anual conforme al artículo 37 de la DSA.

Como se viene advirtiendo, la DSA refuerza la posición del usuario de las plataformas y le dota de vías concretas de intervención y gestión que pueden tener trascendencia a efectos del RIA. En primer término, la DSA articula en su artículo 16 un sistema de notificaciones cuyo establecimiento exige a los prestadores de servicios de alojamiento. A través de este sistema cualquier usuario (persona física o jurídica) o singularmente un alertador fiable del artículo 22 puede notificar la detección de contenido ilícito. Más concretamente, a la hora de gestionar este sistema de notificaciones, el artículo 20 prevé, respecto de las plataformas en línea, el establecimiento de un sistema interno de gestión de reclamaciones que puede ser automatizado. A través de este sistema, las plataformas arbitran un procedimiento para resolver reclamaciones contra las decisiones de moderación de contenido contrario a las condiciones generales del servicio o para detectar anuncios políticos que incumplan las previsiones del RTSP.

El tratamiento de estas notificaciones debe realizarse «de manera no discriminatoria, diligente y no arbitraria» de acuerdo con el artículo 20.4, cumpliendo con los requisitos de los MUG del RIA. Por lo tanto, a la hora de evaluar el resultado de salida de un sistema de IA para la gestión automatizada de estas reclamaciones, una adecuada explicabilidad permite al usuario comprender los parámetros seguidos por la plataforma en su decisión y, en su caso, proceder como mejor convenga a su interés. A este respecto, la referencia a la garantía de transparencia algorítmica resulta clave.

Este sistema interno de gestión de notificaciones implica situar a las plataformas en una posición cuasi judicial, matizando la vuelta al Derecho público del modelo de gobernanza. No obstante, este sistema se entiende sin perjuicio de la posibilidad del usuario de la plataforma de acudir a un sistema de resolución extrajudicial de conflictos o los tribunales ordinarios correspondientes. Un sistema de notificación

análogo al de la DSA, con posibilidad de tramitación automatizada, se prevé en el artículo 16 del RTSP para la identificación y, en su caso, retirada de anuncios que incumplan los requisitos del RTSP.

En todo caso, no puede dejarse de destacar la labor de homogeneización en este ámbito que realizó el Parlamento Europeo en el *iter legis* de la propuesta de RIA al hacer extensivo el derecho de información a este sistema de gestión de reclamaciones en la redacción que propuso al artículo 52, párrafos 1 y 3 ter, en relación con los artículos 27 y 38⁴² de la DSA. Estas previsiones decaen en el texto final, si bien el considerando 170 recuerda la existencia de vías de recurso efectivas para las personas físicas o jurídicas que prevé el Derecho europeo cuando sus derechos o intereses queden afectados por un sistema de IA y recuerda la posibilidad de presentar una denuncia ante la autoridad de vigilancia del mercado ante la infracción de las previsiones del RIA.

Sobre el sistema de gestión de reclamaciones conviene hacer un último apunte de interés. En la tramitación de esta garantía se puede poner de manifiesto una falta de conformidad del sistema de IA utilizado por la plataforma, en cuyo caso, en la gestión de estas reclamaciones puede articularse el mecanismo de colaboración para la adopción de acciones correctoras, particularmente interesante si la reclamación se realiza por los alertadores del artículo 22 de la DSA.

Finalmente, la DSA parte de la necesidad de combinar el uso de sistemas algorítmicos en la prestación de servicios por las plataformas con la revisión humana. Esta cuestión queda sin regulación específica en el texto final del RIA, que nada prevé respecto a los MUG⁴³. No obstante, la garantía que consagra la DSA resulta adecuada en la medida en que refiere la necesidad de supervisión humana de dos tipos de decisiones:

— la moderación automatizada de contenidos, conforme al artículo 14 de la DSA. Este precepto debe ponerse en relación con el artículo 42.2, que completa las previsiones de los artículos 15 y 24.1 exigiendo la transparencia informativa *ex post* (*accountability*) de los datos relativos a los recursos humanos destinados a esta labor de revisión cuando la moderación es automatizada.

— la supervisión por personal humano cualificado de las decisiones adoptadas en el marco del sistema de reclamaciones, conforme al artículo 20.6.

La DSA se refiere a esta participación de personas físicas en la supervisión del funcionamiento de sistemas algorítmicos sin ahondar en su carácter potestativo u obligatorio, si bien del considerando 58 parece derivarse una cierta obligatoriedad en la medida en que obliga a las plataformas a establecer sistemas internos de gestión de reclamaciones «que estén sujetos a revisión humana cuando se usen medios automáticos».

4. GARANTÍAS ORGÁNICAS Y GOBERNANZA DIGITAL

El Derecho digital europeo ha transitado desde un modelo de autorregulación hacia un modelo de corregulación, en el que se combinan elementos de *soft law*

42. Se refiere a la obligación de las grandes plataformas de habilitar al menos una opción para que los sistemas de recomendación no se basen en la elaboración de perfiles.

43. Respecto a los sistemas para la influencia política considerados de alto riesgo, se prevé en el artículo 14.

(códigos de conducta o de buenas prácticas⁴⁴) con una estructura institucional que materializa la vuelta al Derecho público, para supervisar el cumplimiento de la normativa y consolidar el modelo de gobernanza digital en la Unión. Al efecto, articula una serie de garantías orgánicas que se concretan en el nombramiento de autoridades y otros sujetos con competencias de supervisión. Este esquema se sigue en las tres normas que analizamos. La relación entre la DSA y el RIA respecto a estas garantías orgánicas se ponía de manifiesto en el considerando 40 ter del RIA versión Parlamento al afirmar, desde la necesaria perspectiva de impacto, que las autoridades designadas en virtud de la DSA debían actuar como autoridades encargadas de la aplicación de la ley a los efectos de cumplimiento del RIA. Esta previsión, no obstante, puede entenderse subsumida en la cláusula de complementariedad ya citada que recoge el RIA.

De acuerdo con las previsiones del Capítulo IV de la DSA (artículos 49 y siguientes), las autoridades competentes para la supervisión de los prestadores de servicios de supervisión y de la ejecución de la DSA son el coordinador de servicios digitales y la Comisión europea, que ejerce importantes labores de supervisión y puede adoptar actos de ejecución. El engarce de estas figuras se produce de la siguiente manera:

— La Comisión europea está llamada a participar en la estructura de Gobernanza de la IA, a través de la Oficina europea de IA, que colaborará con el Consejo europeo de IA y que, de acuerdo con el artículo 68, podrá asumir la competencia exclusiva de la Comisión para supervisar el cumplimiento de las obligaciones de los sistemas y MUG.

— Por su parte, la Comisión Nacional de los Mercados y la Competencia ha sido designada coordinador de servicios digitales en España,⁴⁵ que deberá en su caso coordinarse con la Agencia Española de Supervisión de la Inteligencia Artificial como autoridad nacional que controla el cumplimiento y ejecución del RIA.

— En todo caso, y habida cuenta de la centralidad de la gestión de riesgos y la transparencia algorítmica en ambas regulaciones, debe destacarse la trascendencia del ECAT (ya citado), un organismo sito en Sevilla y especializado en el análisis multidisciplinar (técnico, científico y jurídico) del uso de los algoritmos, sus riesgos e impacto. El ECAT constituye un apoyo fundamental para la Comisión a la hora de examinar los informes de transparencia y autoevaluación de riesgos de las VLOPs así como en la práctica de actos de ejecución, especialmente las medidas de investigación⁴⁶. Asimismo colaborará, entre otros, con panel científico de expertos independientes previsto en el RIA.

44. Artículos 45 a 47 de la DSA y artículo 56 del RIA, por remisión en el ámbito de este estudio de los artículos 53.5. y 55.2.

45. Comisión Nacional De Los Mercados Y La Competencia, «El Ministerio para la Transformación Digital y de la Función Pública designa a la CNMC como Coordinador de Servicios Digitales de España», Nota de prensa (24 de enero de 2024), disponible en: https://www.cnmc.es/sites/default/files/editor_contenidos/Notas%20de%20prensa/2024/NdP-CNMC-DSA.pdf, último acceso, 6 de marzo de 2024.

46. Comisión Europea, «Aplicación de la Ley de Servicios Digitales: la Comisión pone en marcha el Centro Europeo para la Transparencia Algorítmica», *Comunicado de prensa* (17 de abril de 2023), *vid.*: <https://ec.europa.eu/commission/presscorner/api/files/>

Esta estructura institucional debe entenderse sin menoscabo de otras figuras, previstas en la DSA a efectos de coordinación (Junta europea de derechos digitales); o de bisagra entre usuarios, autoridades y plataformas (nombramiento de puntos de contacto entre plataformas y autoridades del artículo 11) y entre plataformas y destinatarios finales del servicio (artículo 12); de representantes legales de las plataformas (artículo 13 de la DSA o 14 del RTSPP)⁴⁷ o de jefes de comprobación del cumplimiento (artículo 41) que aseguran el cumplimiento de la DSA. Esta estructura se relaciona con la establecida en el artículo 15 del RTSPP que, sin perjuicio del nombramiento de autoridades competentes para ámbitos no regulados por la DSA, somete la supervisión del cumplimiento del RTSPP al entramado institucional de la DSA respecto a los servicios de intermediación.

V. CONCLUSIONES

1. El tratamiento que realiza el RIA respecto a VLOPs y sistemas destinados a la influencia política resulta adecuado en términos sistemáticos y responde de forma más precisa a la lógica de la normativa aplicable, centrada en el proveedor/responsable del despliegue en el caso del RIA y en el usuario final respecto a la DSA y el RTSPP. Estas distintas perspectivas resultan complementarias en la valoración global de los riesgos subyacentes al uso de sistemas de IA en plataformas digitales.

2. El tratamiento final de las VLOPs, al decaer su consideración como sistema de alto riesgo en el RIA, resulta más congruente con el sistema europeo de Derecho digital en la consideración de la intensidad de los riesgos generados, en la medida en que la DSA y el RTSPP consagran una regulación fuertemente garantista basada en la transparencia.

3. Respecto a los sistemas de influencia política, la consideración como de alto riesgo debe interpretarse circunscrita a aquellos sistemas desarrollados específicamente para este objetivo. Por tanto, los sistemas que sirven a este fin de manera accesoria (como los sistemas de recomendación o de creación de ultrafalsificaciones), se someten al régimen de los capítulos IV y V del RIA.

4. El artículo 50 del RIA acoge la orientación al destinatario final de los sistemas destinados a interactuar directamente con personas físicas (como las VLOPs), matizando el artículo 51 y siguientes obligaciones adicionales si concurre riesgo sistémico (MUGRS). El reconocimiento de los MUGRS incorpora en la práctica una nueva gradación de riesgos que articula, merced a su régimen específico de obligaciones, un régimen de intervención intermedio entre los sistemas y modelos de uso general y los sistemas de alto riesgo. Este régimen intermedio se alinea con el régimen obligacional de la DSA y el régimen de transparencia del RTSPP.

5. El reconocimiento de la categoría de MUGRS incide tangencialmente en el régimen de las VLOPs. El RIA no exige la categorización como MUGRS de los modelos de IA que incorporan las VLOPs, sino evaluar el cumplimiento del estándar de gestión de riesgos del RIA, que se presume *iuris tantum* para las VLOPs sometidas

document/print/es/ip_23_2186/IP_23_2186_ES.pdf, último acceso, 7 de marzo de 2024.

47. Que se corresponden con los representantes autorizados de los proveedores del artículo 54 del RIA.

al sistema de gestión de riesgos de la DSA. La presunción opera salvo en aquellos riesgos sistémicos que se demuestren no cubiertos, en cuyo caso serían exigibles las obligaciones cumulativas del RIA. En este mismo sentido, las obligaciones de evaluación y mitigación de riesgos, así como de seguimiento y adopción de medidas correctivas de la DSA son equiparables a las de los MUGRS del RIA en términos de garantía.

6. Las garantías clave descansan en la transparencia algorítmica y la explicabilidad del sistema previstas en la normativa sectorial, que se complementan en cuestiones técnicas con las especificaciones del RIA, en particular: i) la información de los anexos XI a XIII; ii) el deber de información de estar interactuando con un sistema de IA (artículo 50.1); iii) el etiquetado de contenido generado mediante sistemas de IA y iv) el régimen de rendición de cuentas.

7. Más allá de las garantías procedimentales de la DSA para fortalecer la posición del usuario final, en el ámbito objeto de este estudio el sistema de gobernanza del RIA mantiene el protagonismo de la Comisión europea (apoyada en el ECAT) y responde adecuadamente al régimen de corregulación con la incorporación de autoridades nacionales y europeas, plataformas y alertadores fiables con un protagonismo renovado de las técnicas de intervención del Derecho público en perspectiva tecnológica y humanista.

RÉGIMEN GENERAL APLICABLE A LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL DE ALTO RIESGO

La aplicación de las normas armonizadas y de las especificaciones comunes en el ámbito de la inteligencia artificial (artículos 40 y 41 Reglamento)

VICENTE ÁLVAREZ GARCÍA

Catedrático de Derecho Administrativo Universidad de Extremadura

I. INTRODUCCIÓN

1. LA REGULACIÓN DE LA INTELIGENCIA ARTIFICIAL A TRAVÉS DE LA TÉCNICA ARMONIZADORA DEL NUEVO ENFOQUE

La regulación de la inteligencia artificial en la Unión Europea ha seguido la técnica del nuevo enfoque armonizador, de tal manera que las Instituciones europeas han renunciado a la ordenación de este campo por ellas solas, para apelar a una colaboración con ellas de los sujetos privados¹.

Este régimen de colaboración público-privada consiste en que los poderes públicos comunitarios van a establecer el marco normativo general al que estará sometido este producto, incluyendo los requisitos esenciales que necesariamente habrá de respetar para poder ser introducido y comercializado válidamente en el mercado europeo, pero las especificaciones técnicas que sirven para cumplimentar tales requisitos serán fijadas, como regla general, por los organismos europeos de normalización. También serán sujetos privados quienes realicen las evaluaciones de la conformidad (si se quiere, los controles) que permitirán comprobar si el producto ha sido elaborado siguiendo las especificaciones técnicas pertinentes y, en definitiva, si respeta los requisitos esenciales obligatorios en materia de salud, de seguridad, de derechos fundamentales y de valores europeos impuestos por el acto legislativo regulador de la inteligencia artificial.

1. Sobre la aplicación de la técnica del nuevo enfoque armonizador en el campo de la inteligencia artificial, véanse, por todos, los dos siguientes estudios: Álvarez García V. y Tahiri Moreno, J., «La regulación de la inteligencia artificial en Europa a través de la técnica armonizadora del nuevo enfoque», *Revista General de Derecho Administrativo*, n.º 63, 2023; y Álvarez García, V., «Los instrumentos normativos reguladores de las especificaciones técnicas en la Unión Europea: un breve ensayo de identificación de nuevas fuentes del Derecho», *Revista General de Derecho Administrativo*, n.º 64, 2023.

La decisión de seguir el modelo regulatorio del nuevo enfoque provoca que para atender a la ordenación de la inteligencia artificial deba considerarse, en primer término, su acto legislativo regulador, que adopta la forma jurídica de Reglamento, con la base jurídica ofrecida por el artículo 114 TFUE, dado que con dicha norma se pretende «la adopción de medidas para garantizar el establecimiento y el funcionamiento del mercado interior». Es verdad que históricamente la forma seguida por los actos normativos nuevo enfoque reguladores de los productos era la Directiva, pero en la actualidad se ha impuesto de manera rotunda el Reglamento, por la necesidad de conseguir una uniformidad normativa del régimen de las mercancías en el seno de la Unión.

El RIA no es, no obstante, la única norma por la que se regula este producto, sino que, en segundo término, resultan de aplicación diferentes actos legislativos transversales de primer orden, que constituyen lo que en la jerga comunitaria se denomina el «nuevo marco legislativo para la comercialización de los productos». De toda esta normativa, la más importante para este capítulo es el Reglamento sobre la normalización europea de 2012², si bien en el ámbito de los controles técnicos son indispensables el Reglamento por el que se regula la acreditación³, la Decisión ordenadora de los mecanismos de evaluación de la conformidad⁴ y el Reglamento sobre la vigilancia del mercado⁵.

Enmarcado el RIA dentro del modelo armonizador del nuevo enfoque, me parece que es esencial subrayar en este momento que esta norma europea reviste frente al resto de los actos legislativos que siguen esta técnica regulatoria una extraordinaria particularidad, y es que sirve de base jurídica para establecer normas armonizadas para el conjunto de la Unión no de productos físicos, como ha sido tradicional desde la generalización de la política del nuevo enfoque desde mediados de los años ochenta del pasado siglo⁶, sino de las distintas categorías de programas informáticos que se incardinan dentro de la gran familia de la inteligencia artificial.

2. Reglamento (UE) n.º 1025/2012 del Parlamento Europeo y del Consejo, de 25 de octubre de 2012, sobre la normalización europea, por el que se modifican las Directivas 89/686/CEE y 93/15/CEE del Consejo y las Directivas 94/9/CE, 94/25/CE, 95/16/CE, 97/23/CE, 98/34/CE, 2004/22/CE, 2007/23/CE, 2009/23/CE y 2009/105/CE del Parlamento Europeo y del Consejo y por el que se deroga la Decisión 87/95/CEE del Consejo y la Decisión n.º 1673/2006/CE del Parlamento Europeo y del Consejo.
3. Reglamento (CE) n.º 765/2008 del Parlamento Europeo y del Consejo, de 9 de julio de 2008, por el que se establecen los requisitos de acreditación.
4. Decisión n.º 768/2008/CE del Parlamento Europeo y del Consejo, de 9 de julio de 2008, sobre un marco común para la comercialización de los productos y por la que se deroga la Decisión 93/465/CEE del Consejo.
5. Reglamento (UE) 2019/1020 del Parlamento Europeo y del Consejo, de 20 de junio de 2019, relativo a la vigilancia del mercado y la conformidad de los productos y por el que se modifican la Directiva 2004/42/CE y los Reglamentos (CE) n.º 765/2008 y (UE) n.º 305/2011.
6. En relación con la política del nuevo enfoque armonizador véanse, por todos, los libros de Álvarez García, V., *Industria*, Iustel, 2010, pp. 47 y ss., y *Las normas técnicas armonizadas (Una peculiar fuente del Derecho europeo)*, Iustel, 2020, pp. 21 y ss.; además de las pioneras y esenciales obras de M. López Escudero, *Los obstáculos técnicos al comercio en la Comunidad Económica Europea*, Universidad de Granada, 1991; de Valencia Martín, G., *La defensa frente al neoproteccionismo en la Comunidad Europea*, Cámara Oficial de Comercio, Industria y Navegación, 1993; y de Mattera, A., *Le Marché Unique*

2. UNA BREVE INTRODUCCIÓN A LOS ELEMENTOS BÁSICOS DE LA TÉCNICA ARMONIZADORA DEL NUEVO ENFOQUE APLICADOS A LA INTELIGENCIA ARTIFICIAL

Un análisis de la técnica armonizadora⁷ del nuevo enfoque revela que cuenta con tres grandes pilares: en primer lugar, la existencia de unos requisitos esenciales obligatorios que debe respetar un producto, y que se fijan directamente por el acto legislativo aprobado por las Instituciones europeas (en el caso de la inteligencia artificial, por un Reglamento); en segundo lugar, la regulación de las especificaciones técnicas que facilitan la justificación de la conformidad del producto con dichos requisitos imperativos; y, en tercer lugar, la existencia de controles que permiten acreditar en última instancia que el producto cumple con dichas exigencias esenciales obligatorias. Dedicemos unas breves líneas para explicar estos elementos básicos del nuevo enfoque aplicados a la inteligencia artificial.

A) Los requisitos esenciales imperativos que deben respetar los sistemas de inteligencia artificial de alto riesgo tienen como finalidad la protección de la salud, de la seguridad y de los valores y derechos fundamentales en la Unión Europea.

Estas exigencias imperativas establecidas por el RIA se ordenan en torno a los siguientes ámbitos: los datos y su gobernanza, la documentación técnica, los registros, la transparencia y la comunicación de información a los usuarios, la vigilancia humana, la precisión, la solidez y la ciberseguridad.

B) Las especificaciones técnicas que posibilitan el cumplimiento de los requisitos esenciales impuestos a los sistemas de inteligencia artificial por su Reglamento regulador son de tres órdenes: en primer término, las normas técnicas armonizadas europeas (o, más abreviadamente, las normas armonizadas); en segundo término, las especificaciones comunes; y, en tercer término, otras soluciones técnicas «como mínimo equivalentes» a las normas armonizadas o a las especificaciones comunes referidas, y que pueden ser ofrecidas, por ejemplo, mediante normas técnicas elaboradas por organismos de normalización (internacionales, europeos y nacionales) o por los propios operadores económicos que están detrás de la elaboración de los programas informáticos de inteligencia artificial.

C) Los controles técnicos destinados a acreditar el cumplimiento de los requisitos obligatorios por los sistemas de inteligencia artificial de alto riesgo responden a la idea de que de nada vale establecer normas si no se acompañan de mecanismos destinados a asegurar su cumplimiento.

En el marco del nuevo enfoque, los controles de los productos pueden ser previos a su introducción en el mercado o a su comercialización, pero también pueden serlo

Européen, Ses règles, son fonctionnement, Jupiter, 1988. En el primero de estos trabajos se recoge, asimismo, una amplísima muestra bibliográfica sobre la construcción histórica y sobre el funcionamiento de la libertad europea de circulación de mercancías, para quien desee profundizar más sobre esta cuestión.

Me parece imprescindible sobre esta materia, además, la consulta de la Comunicación de la Comisión titulada *Guía azul sobre la aplicación de la normativa europea relativa a los productos de 2022* (DOUE C 247, pp. 1 y ss.).

7. Véase Álvarez García, V., *Las normas técnicas ...cit.* pp. 21 y ss.

a posteriori. Este esquema se reproduce en relación con los sistemas de inteligencia artificial de alto riesgo⁸.

Los controles *ex ante* pueden ser realizados bien por el propio operador económico directamente (este es el caso de los autocontroles y de las autocertificaciones) o bien por una tercera parte independiente (que es un organismo de evaluación de la conformidad previamente notificado a las instancias europeas —o, simplemente, organismo notificado—) previamente acreditada por un organismo destinado a esta tarea (organismo notificante).

Los controles *ex post* pueden revestir, asimismo, diferentes modalidades en función de los sujetos que los realicen. En efecto, estos controles pueden ser realizados por operadores de naturaleza privada (y es que, en primer término, la supervisión se realizará por los propios proveedores directamente o por una tercera parte, esto es, por los organismos notificados), pero también por entidades públicas, dado que juegan un papel fundamental las autoridades nacionales de vigilancia del mercado en el control último de los sistemas de inteligencia artificial de alto riesgo comercializados dentro de la Unión, sin olvidar que deben participar en este tipo de controles las autoridades u organismos estatales encargados de supervisar o de hacer respetar los derechos fundamentales.

Pues bien, centrados en estos términos los tres tipos de elementos básicos de la técnica armonizadora del nuevo enfoque aplicados a la inteligencia artificial, me voy a dedicar a la explicación a lo largo de las próximas páginas exclusivamente de los segundos, esto es, de las dos grandes categorías de especificaciones técnicas que pueden dictarse para desarrollar el nuevo Reglamento regulador de esta trascendental familia de programas informáticos que conforman la inteligencia artificial, empezando por las normas armonizadas, para continuar con las especificaciones comunes. En todo caso, debe subrayarse antes de comenzar su estudio individualizado que, a diferencia de lo que sucede con los requisitos esenciales establecidos directamente por el acto legislativo nuevo enfoque que analizamos (que son, recordemos, obligatorios para los sistemas de inteligencia artificial de alto riesgo), tanto las normas armonizadas como las especificaciones comunes son voluntarias jurídicamente, aunque, eso sí, están dotadas de un efecto jurídico-público de primer orden: los software generados siguiendo estos dos tipos de documentos técnicos se presume que son conformes con los referidos requisitos esenciales que deben ser respetados imperativamente para poder introducirlos y comercializarlos válidamente en el mercado comunitario⁹. Esta presunción de conformidad abre, en otras palabras, todo este mercado al producto. En todo caso, debe entenderse que esta voluntariedad jurídica significa que los operadores económicos pueden seguir otras regulaciones técnicas alternativas a las fijadas en las normas armonizadas o en las especificaciones comunes para elaborar sus productos. Esta configuración jurídica voluntaria ha chocado *de facto*, no obstante, con los elevados costes burocráticos y económicos que

8. Álvarez García V. y Tahiri Moreno, J. «La regulación de la inteligencia ...» cit.
 9. Sobre este efecto jurídico-público de la presunción de conformidad, puede consultarse Álvarez García, V., *Las normas técnicas ...cit.* pp. 160 y ss. En relación con esta cuestión resultan muy interesantes las Conclusiones de la Abogada General Laila Medina presentadas el 22 de junio de 2023, en el asunto «*Public.Resource.Org, Inc., Right to Know CLG contra Comisión Europea*», C-588/21 P (puntos 33 y ss.).

se derivan de fabricar los productos siguiendo otras soluciones técnicas, que ven multiplicados los controles técnicos para acceder al mercado europeo.

II. LAS NORMAS ARMONIZADAS

1. UNA CUESTIÓN PREVIA BÁSICA: LAS NORMAS ARMONIZADAS TIENEN LA NATURALEZA JURÍDICA DE DERECHO COMUNITARIO

Desde que en los años ochenta del pasado siglo empezó a generalizarse la técnica del nuevo enfoque las normas armonizadas se han convertido en un instrumento transcendental de desarrollo de los actos legislativos europeos de esta naturaleza.

A pesar de la abundancia de este tipo de documentos técnicos y de su gran importancia, tan sólo en el año 2016 el Tribunal de Justicia de la Unión Europea abordó por primera vez su naturaleza jurídica, declarando en su crucial Sentencia *James Elliott* que estas normas constituían Derecho de la Unión¹⁰.

La justificación de esta caracterización ha sido realizada por esta Alta Institución comunitaria a partir de un doble elemento: por un lado, su proceso de elaboración; y, por otro, el efecto jurídico-público de la presunción de conformidad al que me he referido ya brevemente.

En relación con su generación, estas normas son elaboradas por unos sujetos privados, que son los organismos europeos de normalización, siguiendo un procedimiento interno acordado en su seno. Ahora bien, dada su función de complemento de los actos legislativos nuevo enfoque, el proceso de intervención de la Comisión en su producción es, realmente, muy relevante: es cierto que se elaboran por entes normalizadores, pero lo hacen previa petición (o mandato) de esta Alta Institución; y una vez finalizadas por estos sujetos, deben ser aceptadas por la propia Comisión, que, además, debe publicar sus referencias (esto es, su código numérico y su título) en el Diario Oficial de la Unión Europea, si se quiere que empiecen a producir efectos jurídicos.

Con respecto a estos efectos jurídicos, recuerda el Tribunal de Justicia comunitario que, a pesar de su carácter voluntario, gozan de presunción de conformidad, lo que hace que a menudo se transformen de hecho en obligatorias, porque, gracias a su respeto en el proceso de fabricación de los productos, los operadores económicos ven enormemente reducidas sus cargas a la hora de demostrar el cumplimiento de los requisitos esenciales impuestos por el acto legislativo (Directiva o Reglamento),

10. STJUE de 27 de octubre de 2016, asunto «*James Elliott Construction Limited contra Irish Asphalt Limited*», C-613/14. Un amplio estudio sobre esta Sentencia puede verse en Álvarez García, V., «La confirmación por parte de la jurisprudencia del Tribunal de Justicia de la Unión Europea de la capacidad normativa de los sujetos privados y sus lagunas jurídicas (el asunto “*James Elliott Construction Limited contra Irish Asphalt Limited*”)», *Revista General de Derecho Administrativo*, n.º 46, 2017. Pueden consultarse, asimismo, los trabajos de B. Lundqvist, «European Harmonised Standards as ‘Part of EU Law’: The implications of the James Elliott Case for Copyright Protection and, possibly, for EU Competition Law», *Legal Issues of Economic Integration*, n.º 44, 2017; y A. Volpato, «The Harmonised Standards before the ECJ: James Elliott Construction», *Common Market Law Review*, n.º 54(2), 2017.

y que obligatoriamente habrán de ser respetados para introducir y comercializar las mercancías en el mercado europeo.

2. LA DISTINCIÓN ENTRE LAS NORMAS EUROPEAS Y LAS NORMAS ARMONIZADAS EUROPEAS

Las normas armonizadas¹¹ están reguladas de manera general en el Reglamento sobre la normalización europea de 2012¹². A esta norma deben añadirse para cada producto sometido al ámbito armonizador del nuevo enfoque las previsiones específicas que pueda contener su concreto acto legislativo regulador. En el caso de los sistemas de inteligencia artificial, las particularidades son, realmente, mínimas, sin realmente añadir nada significativo a la referida regulación general de 2012.

A pesar de su denominación, el Reglamento sobre la normalización europea no ordena en su plenitud todo este campo, dado que, si bien es verdad que contiene una regulación de las normas armonizadas (y ahora también de las normas que desarrollan el acto legislativo sobre Seguridad General de los Productos¹³), no realiza esta función con respecto a las simples normas técnicas europeas (o, abreviadamente, normas europeas).

Estas categorías normativas tienen ciertamente un sustrato común: en ambos supuestos se trata de especificaciones técnicas aplicables a los productos (o a los servicios), que son elaboradas por los organismos europeos de normalización y que están destinadas a una aplicación repetitiva o continuada en el tiempo. Es decir, son realmente normas de naturaleza técnica de ámbito continental, que tienen su origen en un sujeto privado, siguiendo un procedimiento privado y que poseen una naturaleza jurídica voluntaria.

Las diferencias entre ambas figuras (esto es, entre las normas técnicas europeas y las normas técnicas armonizadas europeas) radican en los tres elementos siguientes:

11. Sobre esta distinción, véanse, por todos, los trabajos de Álvarez García, V. «El lugar de las normas técnicas y de las normas técnicas armonizadas en el ordenamiento jurídico europeo», en Jiménez de Cisneros Cid, F.J. (Dir.), *Homenaje al Profesor Ángel Menéndez Rexach*, Aranzadi, 2018, pp. 99 y ss.; y *Las normas técnicas... cit.* pp. 71 y ss.
12. Reglamento (UE) n.º 1025/2012 del Parlamento Europeo y del Consejo, de 25 de octubre de 2012, sobre la normalización europea, por el que se modifican las Directivas 89/686/CEE y 93/15/CEE del Consejo y las Directivas 94/9/CE, 94/25/CE, 95/16/CE, 97/23/CE, 98/34/CE, 2004/22/CE, 2007/23/CE, 2009/23/CE y 2009/105/CE del Parlamento Europeo y del Consejo y por el que se deroga la Decisión 87/95/CEE del Consejo y la Decisión n.º 1673/2006/CE del Parlamento Europeo y del Consejo.
13. Reglamento (UE) 2023/988 del Parlamento Europeo y del Consejo, de 10 de mayo de 2023, relativo a la Seguridad General de los Productos, por el que se modifican el Reglamento (UE) n.º 1025/2012 del Parlamento Europeo y del Consejo y la Directiva (UE) 2020/1828 del Parlamento Europeo y del Consejo, y se derogan la Directiva 2001/95/CE del Parlamento Europeo y del Consejo y la Directiva 87/357/CEE del Consejo.

Sobre esta cuestión, véase, desde un punto de vista doctrinal, Álvarez García, V. «Los documentos técnicos normativos que sirven para garantizar la seguridad de los productos en la Unión Europea», *Revista General de Derecho Administrativo* n.º 64 Iustel, octubre 2023.

A) En primer término, las normas europeas no tienen por objeto complementar ningún acto legislativo armonizador de la Unión (sino que tienen una vida independiente), mientras que las normas armonizadas constituyen un desarrollo indispensable de los actos legislativos nuevo enfoque. Recordemos que estos actos legislativos establecen los requisitos obligatorios esenciales que deben respetar los productos para su válida introducción y comercialización dentro del mercado europeo, en tanto que las normas armonizadas fijan las especificaciones técnicas cuyo cumplimiento permite justificar el respeto por dichos productos de tales requisitos obligatorios. De esta forma, la técnica armonizadora del nuevo enfoque se basa, como ya indiqué con anterioridad, en una suerte de colaboración pública y privada en Europa, en la medida en que, por un lado, las Instituciones europeas (sujetos de Derecho público) elaboran los actos legislativos nuevo enfoque que establecen los requisitos esenciales que obligatoriamente deben respetar los productos para poder ser introducidos válidamente en el mercado interior comunitario y, por otro lado, estos actos legislativos son desarrollados técnicamente por los organismos europeos de normalización (sujetos de Derecho privado) mediante la elaboración de normas armonizadas.

B) En segundo término, las normas armonizadas se elaboran por los organismos europeos de normalización a través de procedimientos fijados por ellos mismos, pero en su proceso de elaboración existe una fuerte intervención por parte de las Instituciones comunitarias. Así, las normas armonizadas sirven para desarrollar actos legislativos nuevo enfoque adoptados por el Parlamento Europeo y por el Consejo; son generadas previo mandato de la Comisión (es verdad, en todo caso, que, en el desarrollo del mandato —esto es, para la elaboración de la norma—, se sigue el procedimiento fijado por los organismos de normalización); una vez elaboradas son examinadas por la Comisión y, en su caso, aceptadas por esta Institución; y finalmente sus referencias son publicadas por la Comisión en el Diario Oficial europeo. Toda esta intervención pública es inexistente en el caso de las meras normas técnicas europeas, que se elaboran libremente por los organismos europeos de normalización siguiendo sus procedimientos internos.

C) En tercer término, las normas armonizadas y las normas europeas tienen, como decíamos hace unos instantes, una naturaleza jurídica voluntaria, pero las normas armonizadas están dotadas de unos efectos jurídico-públicos de primer orden, que hacen de ellas, como ya sabemos, verdaderos actos de Derecho comunitario fiscalizables por el Tribunal de Justicia de la Unión. El principal de estos efectos consiste en la presunción de conformidad del producto fabricado siguiendo sus prescripciones técnicas con los requisitos obligatorios establecidos por el acto legislativo nuevo enfoque que lo regula para el conjunto del mercado interior europeo. Téngase en cuenta que las normas armonizadas producen estos efectos jurídico-públicos, esenciales para asegurar la libertad comunitaria de circulación de mercancías, a pesar de ser elaboradas por sujetos privados y de que su contenido completo no está publicado oficialmente, sino tan sólo sus referencias (esto es, su código numérico y su título). Además, este contenido de las normas, que es traducido a los idiomas oficiales de los distintos Estados miembros de la Unión por sus correspondientes organismos nacionales de normalización (transformando así las normas europeas en normas nacionales), está protegido por los derechos

de propiedad intelectual pertenecientes a los organismos normalizadores, que las venden para financiarse¹⁴.

Una última consideración con respecto a la voluntariedad jurídica de las normas europeas y de las normas armonizadoras: esta voluntariedad significa que, si los operadores económicos no las siguen, no serán castigados administrativamente con ninguna sanción, pero la realidad práctica demuestra que el mercado a menudo las impone fácticamente, porque los operadores económicos tienden a adquirir únicamente productos fabricados conforme a estas normas y así se acredita mediante la correspondiente certificación¹⁵. A esta imposición del mercado, común en el caso de ambas categorías normativas, se añade en el supuesto de las normas armonizadas su efecto jurídico de la presunción de conformidad que facilita el comercio transnacional dentro de la Unión y reduce las cargas (administrativas y económicas) que rodean a la demostración de que un producto respeta los requisitos esenciales del acto legislativo nuevo enfoque que lo regula.

3. LA INTERVENCIÓN DE LOS ORGANISMOS EUROPEOS DE NORMALIZACIÓN Y DE LA COMISIÓN EN EL PROCEDIMIENTO DE ELABORACIÓN DE LAS NORMAS ARMONIZADAS

El texto de las normas armonizadas se redacta por alguno de los tres organismos europeos de normalización en función de la materia de que se trate. Estas tres entidades, que tienen en común que son sujetos privados de naturaleza asociativa constituidos conforme al Derecho belga o al francés, son: 1) El Comité Europeo de Normalización (CEN), cuyas funciones normalizadoras se extienden a todos los ámbitos de la industria y de los servicios (excluyendo la electrotecnología y las telecomunicaciones); 2) El Comité Europeo de Normalización Electrotécnica (CENELEC), cuya actividad se centra en la electrotecnia; y 3) El Instituto Europeo de Normas de Telecomunicaciones (ETSI, por sus siglas su inglés), que está encargado de la elaboración de las normas en el mundo de las telecomunicaciones¹⁶.

Las normas armonizadas sólo se elaboran por una o varias de estas entidades asociativas europeas. Ahora bien, es verdad que en ocasiones los organismos continentales se limitan a convertir en europeas las normas adoptadas por sus homólogos internacionales, que son igualmente tres: la Organización Internacional de Normalización (ISO), la Comisión Electrotécnica Internacional (CEI) y la Unión Internacional de Telecomunicaciones (UIT). Es cierto que las dos primeras entidades, que tienen una naturaleza de organizaciones internacionales no gubernamentales (compuestas por los organismos nacionales de normalización de buena parte de los Estados del planeta), realizan funciones exclusivamente normalizadoras, mientras

14. Sobre esta cuestión, puede profundizarse en Álvarez García, V., *Las normas técnicas...* cit. pp. 146 y ss.

15. Con respecto a la cuestión de la obligatoriedad fáctica tanto de las normas técnicas como de las certificaciones voluntarias, véanse Álvarez García, V., «El proceso de privatización de la calidad y de la seguridad industrial y sus implicaciones desde el punto de vista de la competencia empresarial», *Revista de Administración Pública*, n.º 159, 2002, pp. 344 y ss.; y Muñoz Machado, S. *Tratado de Derecho Administrativo y Derecho Público General*, T. XIV: *La actividad regulatoria de la Administración*, BOE, 4ª ed., 2015, pp. 80 y ss.

16. Álvarez García, V., *La normalización industrial*, Tirant lo Blanch, 1999, pp. 367 y ss.

que la UIT tiene una naturaleza jurídico-pública, puesto que es el organismo especializado de las Naciones Unidas para las tecnologías de la información y la comunicación (TIC), y que, entre la pluralidad de funciones que desarrolla en este ámbito, se encuentra la realización de tareas normalizadoras¹⁷.

En la medida en que los organismos europeos de normalización son sujetos privados, que actúan siguiendo procedimientos privados, sus normas han sido tradicionalmente privadas. Y así sigue siendo con las normas técnicas europeas. No obstante, con el desarrollo de la técnica armonizadora del nuevo enfoque, las Instituciones europeas vienen encargando a las referidas entidades normalizadoras continentales desde hace décadas la elaboración de soluciones técnicas que completen los actos legislativos nuevo enfoque, que reciben la denominación, como bien sabemos, de normas armonizadas.

Esta función jurídico-pública que, en definitiva, cumplen las normas armonizadas ha justificado tradicionalmente, tal y como ya hemos avanzado anteriormente, una intervención de la Comisión en su proceso de adopción, prevista singularmente en cada uno de los actos legislativos nuevo enfoque, pero desde el año 2012 existe una regulación general en el Reglamento sobre la normalización europea ordenando dicha intervención¹⁸. Los hitos esenciales de la misma son:

A) La emisión del concreto mandato de normalización por parte de la Comisión que va dirigido a uno o varios organismos europeos de normalización y que, por lo tanto, es previo al inicio del trabajo normativo de estas entidades (que siempre pueden aceptar o rechazar el mismo). En todo caso, sin un previo mandato no existirá una norma armonizada. Las previsiones regulatorias esenciales sobre los mandatos están establecidas por los apartados 1 y 2 del artículo 10 del Reglamento de 2012, aunque, ciertamente, la Comisión ha desarrollado profusamente su contenido en su *Vademécum de la normalización europea en apoyo de la legislación y las políticas de la Unión Europea*¹⁹.

17. *Ibidem*, pp. 423 y ss.

18. Sobre la intervención de la Comisión en el proceso de elaboración de las normas armonizadas, véase Álvarez García, V., *Las normas técnicas... cit.* pp. 103 y ss.

19. Estos apartados 1 y 2 del Reglamento sobre la normalización europea disponen: «1. Dentro de los límites de las competencias que establece el TFUE, la Comisión podrá pedir a una o varias organizaciones europeas de normalización que elaboren una norma europea o un documento europeo de normalización en un plazo determinado. Las normas europeas y los documentos europeos de normalización deberán basarse en el mercado, tomar en consideración el interés público, así como los objetivos de política claramente expuestos en la petición de la Comisión, y ser fruto del consenso. La Comisión fijará los requisitos respecto del contenido que deberá cumplir el documento solicitado y un plazo para su adopción.

2. Las decisiones a que se refiere el apartado 1 se adoptarán con arreglo al procedimiento establecido en el artículo 22, apartado 3, tras consultar a las organizaciones europeas de normalización y las organizaciones de partes interesadas europeas que reciban financiación de la Unión de conformidad con el presente Reglamento, así como al comité establecido por la correspondiente legislación de la Unión, si existe tal comité, o a través de otros medios de consulta de expertos sectoriales».

El procedimiento al que se refiere el art. 10.2 del Reglamento sobre la normalización europea de 2012 es el procedimiento de examen, que está regulado en el Reglamento (UE) n.º 182/2011 del Parlamento Europeo y del Consejo, de 16 de febrero de 2011, por el que se establecen las normas y los principios generales relativos a las

B) El contenido de la norma armonizada lo elaboran los organismos europeos de normalización siguiendo su normativa interna. Es cierto que, durante su proceso de elaboración, estas entidades se coordinan con la Comisión, pero la competencia para elaborar el texto de la norma y para aprobarla es exclusiva de estos organismos. Cuando la norma está aprobada, el correspondiente organismo europeo de normalización debe mandar a la Comisión su texto completo en los idiomas oficiales de trabajo del mismo (inglés, francés y alemán), además de las referencias de dicha norma (que incluyen su código numérico y su título en todas las lenguas oficiales de la Unión).

C) La Comisión debe analizar el contenido de la norma armonizada para verificar si su texto se adecua a las estipulaciones del acto legislativo nuevo enfoque que desarrolla y al concreto mandato que esta Institución dirigió a los entes normalizadores y que le sirve de base jurídica específica. Esta tarea de «recepción» de la norma armonizada por parte de la Comisión está regulada en el apartado 5 del artículo 10 del Reglamento sobre la normalización europea²⁰.

Aunque no lo dice el precepto indicado, en el proceso de verificación del cumplimiento del mandato por la correspondiente norma armonizada participan primero consultores externos a la Comisión (o *Harmonised Standards Consultants*—HAS—) y, finalmente, los propios funcionarios de esta Alta Institución²¹.

D) En caso de ser aceptada la norma armonizada por la Comisión, las referencias (y sólo las referencias, que no su texto) se publicarán en el Diario Oficial de la Unión Europea. Únicamente con esta publicación oficial limitada la norma armonizada gozará del efecto jurídico-público de la presunción de conformidad. La obligación de proceder a esta publicación oficial, una vez que ha sido recepcionada la norma armonizada por la Comisión, está prevista en el apartado 6 del artículo 10 del Reglamento sobre la normalización europea²².

modalidades de control por parte de los Estados miembros del ejercicio de las competencias de ejecución por la Comisión.

20. Este precepto dispone que: «Las organizaciones europeas de normalización informarán a la Comisión sobre las actividades emprendidas para la elaboración de los documentos contemplados en el apartado 1. La Comisión, de forma conjunta con las organizaciones europeas de normalización, evaluará la conformidad de los documentos elaborados por las organizaciones europeas de normalización con su petición inicial».
21. Sobre el proceso de recepción por la Comisión de las normas técnicas, Álvarez García, V., *Las normas técnicas ...cit.* pp. 133 y ss.
22. Este precepto prevé que: «Cuando una norma armonizada cumpla los requisitos que está previsto que regule, establecidos en la correspondiente legislación de armonización de la Unión, la Comisión publicará sin demora una referencia a dicha norma armonizada en el Diario Oficial de la Unión Europea o por otros medios, con arreglo a las condiciones establecidas en el correspondiente acto de la legislación de armonización de la Unión».

En relación con la publicidad de las normas armonizadas y los problemas jurídicos que plantea, véanse Álvarez García, V., *Las normas técnicas ...cit.* pp. 136 y ss., y 182 y ss.; Bellis, M. De, «Private standards, EU law and access — The General Court's ruling in *Public.Resource.Org*», *Eulawlive*, 10-9-2021; Volpato, A. «Rules Behind Paywall: the Problem with References to International Standards in EU law», *Eulawlive*, 19-7-2021; Volpato, A., «Transparency and Legal Certainty of the References to International Standards in EU Law: Smoke Signals from Luxembourg?: Stichting

4. LAS PREVISIONES SOBRE LAS NORMAS ARMONIZADAS DURANTE EL PROCESO DE TRAMITACIÓN DE LA PROPUESTA DE REGLAMENTO EFECTUADAS POR LA COMISIÓN, POR EL CONSEJO Y POR EL PARLAMENTO EUROPEO

Las regulaciones de las normas armonizadas efectuadas por la propuesta de RIA elaborada por la Comisión, por el texto transaccional del Consejo y por las enmiendas del Parlamento contienen algunas variantes.

A) Las previsiones del texto de la Comisión contenidas en su escueto artículo 40 se refieren únicamente al efecto de la presunción de conformidad con los requisitos esenciales obligatorios establecidos en la propuesta de RIA para los sistemas de inteligencia artificial de alto riesgo que respeten las normas armonizadas cuyas referencias hayan sido publicadas en el Diario Oficial europeo. Recordemos que este efecto jurídico-público fundamental provoca en los diferentes productos sometidos a la legislación armonizadora del nuevo enfoque dentro de la Unión Europea que, aunque las normas armonizadas sean voluntarias jurídicamente, resulten seguidas amplísimamente por los fabricantes, dado que su cumplimiento reduce las cargas burocráticas y económicas y, en muchas ocasiones, permite tras meros controles de conformidad internos (esto es, realizados por los propios fabricantes) el empleo del mercado CE y, por tanto, el acceso al mercado interior europeo.

B) El texto del Consejo también prevé este trascendental efecto de la presunción de conformidad, pero añade algunas precisiones en relación tanto, en primer término, con el contenido de los mandatos que la Comisión debe dirigir a los organismos europeos de normalización encargándoles la elaboración de normas armonizadas de desarrollo de las previsiones del RIA en relación con los sistemas de inteligencia artificial de alto riesgo (a los que añade los sistemas de inteligencia artificial de uso general²³), como, en segundo término, sobre las obligaciones que pesan sobre estas entidades normalizadoras una vez que consideran cumplimentado dicho mandato con la aprobación del contenido de las correspondientes normas armonizadas.

a) Dispone, en efecto, el texto transaccional del Consejo, en relación con la primera de estas cuestiones, que los mandatos de normalización deben especificar que las normas armonizadas deben ser «coherentes» y «claras», y perseguir, «en particular», estos cuatro objetivos: en primer término, la garantía de que los sistemas de inteligencia artificial de alto riesgo sean seguros, respeten los valores de la Unión y garanticen «su autonomía estratégica abierta»; en segundo término, la promoción de

Rookpreventie Jeugd and Others (C-160/20)», Eulawlive, 1-3-2022; y Volpato A. y Eliantonio, M., «The Butterfly Effect of Publishing References to Harmonised Standards in the L series», European law blog, 7-3-2019.

23. El concepto de «sistema de inteligencia artificial de uso general» se incorpora en el texto transaccional del Consejo con el siguiente tenor: es, dice este documento, «un sistema de inteligencia artificial que, con independencia de la manera en la que se introduzca en el mercado o se ponga en servicio, incluido el software de código abierto, ha sido concebido por el proveedor para desempeñar funciones de aplicación general, como el reconocimiento de imágenes y de voz, la generación de audio y vídeo, la detección de patrones, la respuesta a preguntas y la traducción, entre otras. Un sistema de inteligencia artificial de uso general puede utilizarse en una pluralidad de contextos e integrarse en una pluralidad de otros sistemas» [art. 3.1 ter)].

«la inversión y la innovación en inteligencia artificial, incluso mediante el incremento de la certidumbre jurídica, así como la competitividad y el crecimiento del mercado de la Unión»; en tercer término, el fomento de la participación («gobernanza») de todas las partes interesadas en la normalización (desde, por ejemplo, la industria hasta la sociedad civil, pasando por las pymes y los investigadores); y, en cuarto término, el reforzamiento de la cooperación «mundial» en la normalización de la inteligencia artificial, de manera «que sea coherente con los valores e intereses de la Unión».

b) En relación, en segundo lugar, con las obligaciones de los organismos de normalización europeos que redacten las normas armonizadas, el Consejo establece en su texto transaccional la carga de que aporten «pruebas de los esfuerzos que dediquen a cumplir los objetivos referidos».

C) El Parlamento Europeo formula, por su parte, cuatro enmiendas al breve texto originario formulado por la Comisión sobre la regulación de las normas armonizadas.

a) La primera enmienda (esto es, la 437) amplía el efecto de la presunción de conformidad, más allá de los sistemas de inteligencia artificial de alto riesgo, a los modelos fundacionales²⁴.

b) Las enmiendas segunda y tercera (es decir, las 438 y 439) se refieren tanto a la forma de elaborar las peticiones (o mandatos) por la Comisión, como a su contenido, con el siguiente tenor: 1) Habilita a la Comisión para formularlas en relación con todos los requisitos esenciales establecidos por el RIA, respetando las previsiones sobre esta cuestión establecidas por el Reglamento sobre la normalización europea; 2) Fija un plazo máximo de dos meses desde la entrada en vigor del Reglamento para presentar las peticiones; 3) En la preparación de las peticiones, se impone a la Comisión la consulta al Comité Europeo de Inteligencia Artificial previsto en el artículo 56 de la propuesta de la Comisión y finalmente en el artículo 64) y al foro consultivo; 4) A la hora de formular el contenido de estas peticiones, se impone a la Comisión que especifique que las normas serán «coherentes» con lo establecido en el Reglamento y en toda la legislación armonizadora dictada hasta ahora en el seno de la Unión, reiterando, además, la obligación de que dichas normas estén «destinadas a garantizar que los sistemas de inteligencia artificial o modelos fundacionales introducidos en el mercado o puestos en servicio en la Unión cumplan los requisitos pertinentes establecidos en el presente Reglamento» (tanto los establecidos para los sistemas de inteligencia artificial de alto riesgo como para los modelos fundacionales).

c) La cuarta enmienda (la 440) se refiere a las obligaciones que incumben a los agentes que participen en el proceso de normalización. Estas obligaciones son las cuatro siguientes: 1) Deberán tener en cuenta los principios generales para una inteligencia artificial fiable establecidos expresamente por el propio Reglamento regulador de esta materia; 2) Tratarán de promover la inversión, la innovación, la

24. La enmienda 168 del Parlamento Europeo a la propuesta de Reglamento de inteligencia artificial propone la introducción de un nuevo punto 1 quater a su art. 3, con la siguiente definición de «modelo fundacional»: es «un modelo de sistema de inteligencia artificial entrenado con un gran volumen de datos, diseñado para producir información de salida de carácter general y capaz de adaptarse a una amplia variedad de tareas diferentes». En la versión final se maneja el concepto de 3. 63^b «modelo de IA de uso general».

competitividad y el crecimiento del mercado, todo ello en el ámbito de la inteligencia artificial; 3) Contribuirán al reforzamiento de la cooperación internacional en materia de normalización de la inteligencia artificial, teniendo en cuenta, además, las normas internacionales que existan en este ámbito, siempre y cuando sean coherentes «con los valores, los derechos fundamentales y los intereses de la Unión»; y 4) Garantizarán la participación efectiva y equilibrada de todas las partes interesadas en la normalización contempladas en el Reglamento sobre la normalización europea.

D) En fin, aunque las previsiones regulatorias de la Comisión sobre las normas armonizadas en el proyecto de RIA son realmente breves, no creo que los añadidos propuestos ni por el Consejo ni por el Parlamento Europeo tengan ni gran valor declarativo ni ninguna eficacia práctica. Debe tenerse en cuenta que, como ya hemos visto hace unos instantes, existe un régimen general en el Reglamento sobre la normalización europea de 2012, que, si bien es cierto que no es muy detallado, sí que establece una regulación jurídica reducida del sistema de elaboración y adopción de las normas armonizadas, que es «completada» por otros documentos de la Comisión como el *Vademécum* al que antes aludí.

Con esta perspectiva normativa, creo que se debería proceder a establecer un marco normativo para el conjunto del sistema europeo de normalización, más allá de la mínima regulación que ahora se establece para las normas armonizadas (y desde hace unos pocos meses para las normas que desarrollan el nuevo Reglamento General de Seguridad de los Productos²⁵), que, como decía, es francamente reducida, por no decir que resulta mínima. En todo caso, no me voy a centrar ahora en la primera de estas cuestiones (esto es, en el establecimiento de un verdadero marco regulador general del sistema de normalización europea), pero sí quiero hacer una precisión sobre la intervención de la Comisión en el proceso de gestación de las normas armonizadas en la que he insistido en otras ocasiones. Y esta precisión es la siguiente: este tipo de normas técnicas de desarrollo de los actos legislativos nuevo enfoque tienen unos importantísimos efectos jurídico-públicos, y, sin embargo, no son objeto de publicación en el Diario Oficial comunitario²⁶. Tan sólo se publican, en efecto, sus códigos numéricos y sus títulos, pero no su contenido. La justificación para su falta de publicación consiste en que estos textos, que son propiedad de los organismos de normalización, sirven para la financiación de los mismos. No me parece que el coste de dichos textos sea tan elevado para que no se puedan adquirir por la Comisión²⁷, dado que estas normas son fundamentales para el buen

25. Reglamento (UE) 2023/988 del Parlamento Europeo y del Consejo, de 10 de mayo de 2023, relativo a la Seguridad General de los Productos.

26. Sobre la importantísima problemática de la falta real de publicación oficial de las normas técnicas, véase Álvarez García, V., *Las normas técnicas ...cit.* pp. 179 y ss.; y «La problemática de la publicidad oficial de las normas técnicas de origen privado que despliegan efectos jurídico-públicos», *Revista de Derecho Comunitario Europeo*, n.º 72, 2022. Con respecto a esta cuestión son muy relevantes las Conclusiones de la Abogada General Laila Medina presentadas el 22 de junio de 2023, en el asunto «Public.Resource.Org, Inc., Right to Know CLG contra Comisión Europea», C-588/21 P.

27. Téngase en cuenta, a este respecto, el siguiente dato tomado de las Conclusiones de la Abogada General Laila Medina presentadas el 22 de junio de 2023, en el asunto «Public.Resource.Org, Inc., Right to Know CLG contra Comisión Europea», C-588/21 P: «Según alegó el CÉN en la vista, el 4,6 % del presupuesto de normalización procede

funcionamiento de la política armonizadora del nuevo enfoque (esto es, del mercado interior y de la industria europea). Nosotros conocemos en nuestro país técnicas de colaboración mixta público-privada que podrían servir para profundizar en las relaciones entre la Comisión y los organismos de normalización europeos, que, por lo demás, ya existen, articulándose a través de las Directrices Generales para la Cooperación entre CEN, CENELEC y ETSI con la Comisión Europea y la Asociación Europea de Libre Comercio de 28 de marzo de 2003²⁸.

La generación de normas armonizadas es realmente una tarea pública encargada a sujetos privados²⁹. Pues bien, la Comisión tendría que compensar financieramente el ejercicio de esas tareas públicas. Esta solución podría contribuir a la mejora del sistema normalizador, dotándolo de los recursos necesarios, que serían esenciales para toda la normalización y, en particular, para la referida a las tecnologías de inteligencia artificial. Esta última tarea normalizadora requiere disponer de recursos económicos y personales suficientes para elaborar normas en cortos períodos temporales. Cosa que ahora, con la duración de los actuales procesos de normalización, parece prácticamente imposible. Téngase en cuenta que, «desde la primera propuesta hasta la publicación final, el desarrollo de una norma técnica lleva usualmente tres años»³⁰.

E) Con independencia de las previsiones sobre los efectos jurídicos de las normas armonizadas (comunes a los textos de la Comisión, del Consejo y del Parlamento Europeo), sobre los mandatos de normalización y la justificación de su cumplimiento (específicas de los textos del Consejo y del Parlamento Europeo) o sobre las obligaciones de los agentes que participen en el proceso de normalización (propias del Parlamento), los tres documentos prevén que el Comité Europeo de Inteligencia Artificial (o, en la formulación del Parlamento Europeo, la Oficina de Inteligencia Artificial), en su función de asesoramiento y asistencia a la Comisión Europea sobre esta materia, pueda emitir «dictámenes, recomendaciones o contribuciones por escrito» sobre las normas armonizadas o las especificaciones comunes ya existentes [artículo 58 c) en la propuesta de la Comisión y en el texto del Consejo; y artículo 56 ter, letra h), punto i), de la enmienda 529 del Parlamento Europeo]. En lo que finalmente es el artículo 67. 8º RIA.

5. LAS NORMAS ARMONIZADAS EN EL TEXTO DEFINITIVO DEL REGLAMENTO

La versión final del texto del artículo 40 del RIA sigue regulando, como no podía ser de otra forma, la presunción de conformidad de conformidad de los sistemas de

de la venta de las normas técnicas armonizadas, lo que equivale a aproximadamente 2 millones de euros al año, mientras que, en palabras del propio CEN, la financiación de la Comisión equivale a “alrededor del 20% del presupuesto total del CEN”» (punto 99).

28. En relación con el contenido y con la significación de estas Directrices Generales, véase Álvarez García, V., *Las normas técnicas ...cit.*, pp. 103 y ss.
29. Sobre la posibilidad de que los sujetos privados elaboren normas jurídicas, véase Álvarez García, V., «La capacidad normativa de los sujetos privados», *Revista Española de Derecho Administrativo*, n.º 99, 1998, pp. 343 y ss.; y *Las normas técnicas ...cit.* pp. 197 y ss.
30. McFadden, M. y otros (Oxford Commission on AI&Good Governance), *Harmonising Artificial Intelligence: The role of standards in the EU AI Regulation*, December 2021, p. 17.

Inteligencia artificial de alto riesgo que sean conformes con las normas armonizadas o partes de las mismas cuyas referencias se hayan publicado en el Diario Oficial de la Unión Europea (conforme a las previsiones del Reglamento sobre la normalización europea de 2012) con los requisitos esenciales aplicables a los mismos para su válida comercialización dentro del territorio comunitario.

Las normas armonizadas en el ámbito de la inteligencia artificial, como en el resto de los sectores a los que se aplican las técnicas del nuevo enfoque, serán elaboradas por los organismos europeos de normalización previa la pertinente solicitud o mandato formulado por la Comisión. En el mandato de normalización, que esta Alta Institución europea formulará «sin demoras indebidas», se exigirán, en primer término, resultados sobre los procesos de información y documentación para mejorar el rendimiento de los recursos de los sistemas de inteligencia artificial y, en segundo término, se especificará que las normas han de ser coherentes, claras y destinadas a garantizar que los sistemas de inteligencia artificial comercializados o puestos en servicio dentro de la Unión cumplen los requisitos establecidos en el propio RIA. En la elaboración de los mandatos de normalización la Comisión habrá de consultar al Consejo y a las diferentes partes interesadas (incluido el Foro Consultivo).

Los organismos europeos de normalización desarrollarán los mandatos elaborando las normas armonizadas pertinentes, según sus reglas de funcionamiento interno. Ahora bien, habrán de facilitar a petición de la Comisión pruebas de sus mejores esfuerzos para cumplir los requisitos y los objetivos previstos en el RIA.

Este último acto legislativo prevé, por último, que los agentes implicados en las tareas normalizadoras deberán promover la inversión y la innovación en la esfera de la inteligencia artificial. Para ello, procurarán aumentar la seguridad jurídica, la competitividad y el crecimiento del mercado de la Unión, además de contribuir a reforzar la cooperación mundial en materia de normalización, teniendo en cuenta, por un lado, las normas internacionales existentes en el ámbito de la inteligencia artificial, siempre y cuando sean coherentes con los valores, los derechos fundamentales y los intereses de la Unión, y mejorando, por otro lado, la gobernanza multilateral con una representación equilibrada de los intereses implicados y la participación efectiva de las partes interesadas.

6. LOS PROBLEMAS DE LA APLICACIÓN DE LAS TÉCNICAS NORMALIZADORAS A LA REGULACIÓN DE LA INTELIGENCIA ARTIFICIAL EN LA UNIÓN EUROPEA

La normalización es un proceso destinado a la elaboración de especificaciones técnicas, que, reflejando en un momento dado el desarrollo de la ciencia y de la tecnología, permiten a los operadores económicos fabricar sus productos³¹. Estas

31. Sobre el fenómeno de la normalización pueden verse mis trabajos siguientes: Álvarez García, V., *La normalización industrial... cit.*; *Industria*, Iustel, 2010; *Las normas técnicas ...cit.*; así como Aubry, H., Brunet A. y Peraldi Leneuf, F., *La normalisation en France et dans l'Union Européenne*, Presses Universitaires d'Aix-Marseille, 2012; Bismuth, R., *La standardisation internationale privée (Aspects juridiques)*, Larcier, 2014; Cantero M. y Micklitz, M.W. (Eds.), *The Role of the EU in Transnational Legal Ordering: Standards, Contracts and Codes*, Edward Elgar Publishing, 2020; Carrillo Donaire, J.A., *El derecho de la seguridad y de la calidad industrial*, Marcial Pons, 2000;

especificaciones, que en sus versiones más acabadas reciben el nombre de normas técnicas, son elaboradas por los organismos de normalización, que pueden operar a nivel internacional general, supranacional regional o estatal. En el mundo occidental, estas entidades suelen tener una forma jurídica privada, dado que están integradas preferentemente por representantes de los operadores económicos y de los consumidores, además de por académicos y, cada día más, por organizaciones sociales (de defensa, por ejemplo, de los intereses de los trabajadores o del medio ambiente), sin olvidar también la participación más o menos intensa de las diferentes Administraciones públicas. Las decisiones sobre las normas técnicas se toman, en todo caso, por consenso de los diferentes sujetos.

La normalización se ha utilizado tradicionalmente sobre todo en el mundo de los productos físicos para asegurar su interoperatividad, su seguridad y su calidad. De los productos físicos se ha ido extendiendo progresivamente al mundo de los servicios (aunque aquí su desarrollo resulta todavía limitado) y ahora se persigue su implantación en el ámbito de los software de inteligencia artificial. De forma paralela a esta ampliación del ámbito material de las técnicas normalizadoras, se está haciendo lo propio con sus fines (más allá de la interoperatividad, seguridad industrial y calidad) hasta perseguir objetivos económicos, sociales y políticos de primer orden. Así, por ejemplo, en el ámbito de los sistemas de inteligencia artificial, los procesos de normalización en la Unión, que están dando sus primeros balbucesos, persiguen que la industria europea tenga un papel significativo a nivel mundial o que se respeten los valores y los derechos fundamentales en el viejo continente.

Contreras, J.L., *The Cambridge Handbook of Technical Standardization Law. Volume 2: Further Intersections of Public and Private Law*, Cambridge University Press, 2019; Delimatsis, P., *The Law, Economics and Politics of International Standardisation*, Cambridge University Press, 2015; J. Esteve Pardo, *Técnica, Riesgo y Derecho*, Ariel Derecho, 1999; Falke, J., *Rechtliche Aspekte der Normung in den EG-Mitgliedstaaten und der EFTA*, Band 3: Deutschland, European Communities, 2000; G. Fernández Farreres, «Industria», en Martín-Retortillo Baquer S. (Dir.), *Derecho Administrativo Económico*, T. II, La Ley, 1991; F. Gambelli, *Aspects juridiques de la normalisation et de la réglementation technique européenne*, Eyrolles/Fédération des industries mécaniques, 1994; M. Izquierdo Carrasco, *La seguridad de los productos industriales*, Marcial Pons, 2000; Malaret García, E., «Una aproximación jurídica al sistema español de normalización de productos industriales», *Revista de Administración Pública*, n.º 116, 1988; R. Rodrigo Vallejo, «The Private Administrative Law of Technical Standardization», *Yearbook of European Law*, n.º 40, 2021; H. Schepel, y J. Falke, *Legal aspects of standardisation in the Member States of the EC and EFTA*, Vol. 1: Comparative report, European Communities, 2000; Schepel, H. y Falke J. (Ed.), *Legal aspects of standardisation in the Member States of the EC and EFTA*, Vol. 2: Country reports, European Communities, 2000; Schepel, H. «Private Standards as a Replacement for Public Lawmaking?», en Cantero M. y Micklitz H.W. (eds.), *The Role of the EU in Transnational Legal Ordering*, Edward Elgar Publishing, 2020; y Van Waeyenberge, A. «La normalisation technique en Europe-L'empire (du droit) contreattaque», *Revue Internationale de Droit Économique*, n.º 32 (3), 2018.

Los organismos de normalización han intentado reflejar en las normas técnicas los desarrollos científicos y tecnológicos en un tiempo razonable, pero esto, que ha sido relativamente fácil para los productos físicos e incluso para los servicios, plantea importantes problemas en el mundo de las tecnologías de la información y de la comunicación (TIC) en general³², y de los sistemas de la inteligencia artificial en particular, dados los rápidos avances de la ciencia y de la tecnología en estos ámbitos. En diciembre de 2021 vio la luz un importante estudio sobre las fortalezas y, sobre todo, los desafíos que presenta el sistema europeo de normalización ante la aprobación del marco comunitario de inteligencia artificial. Este trabajo se denomina *Harmonising Artificial Intelligence: The role of standards in the UE AI Regulation*³³. Me parece oportuno destacar ahora algunos de los retos a los que se enfrenta la normalización ante este tipo de tecnologías, que, se subrayaba en dicho estudio, «es tan reciente que los organismos de normalización están ahora comenzando a diseñar sus planes para la actividad normalizadora»³⁴. De entre estos desafíos, vale la pena destacar en este momento los cinco siguientes:

A) Se señala en este estudio, en primer término, que debe procederse a una mejora de la capacidad (esto es, de los recursos) de los organismos europeos de normalización para enfrentarse a la elaboración de las normas técnicas armonizadas en materia de inteligencia artificial. Y es que existe una discordanza significativa entre las normas armonizadas que se consideran necesarias en el ámbito de la inteligencia artificial y los desarrollos de especificaciones técnicas que se han producido de manera efectiva hasta ahora en el viejo continente³⁵. Esta disfuncionalidad se centra en una amplísima medida en la ausencia de los recursos necesarios para la normalización tanto en lo que afecta a la sempiterna cuestión de su financiación como en lo que se refiere al número de expertos procedentes de la inteligencia artificial que se necesitan emplear en las tareas normalizadoras.

32. Es verdad, no obstante, que ya desde hace años las Instituciones europeas han elaborado documentos de *softlaw* propugnando la necesidad de desarrollar la normalización en el ámbito de las TIC. Sirvan de ejemplo el Libro Blanco de la Comisión titulado *Modernizar la normalización de las TIC en la UE-El camino a seguir* de 2009 [COM(2009) 324final] y la Comunicación, también de la Comisión, rubricada *Prioridades de normalización en el sector de las TIC para el mercado único digital* de 2016 [COM(2016) 176final].

33. McFadden, M. y otros, *Harmonising... cit.* pp. 4 y 5.

Son numerosos, además, los trabajos y documentos de naturaleza tanto pública como privada que en los últimos años se han ido elaborando para generar ideas sobre cómo adaptar el sistema europeo de normalización a los nuevos retos que plantean las tecnologías de la información y de la comunicación (TIC) en general, y los sistemas de inteligencia artificial en particular. Por mencionar algunos, pueden señalarse: en primer término, el *Plan de Estandarización de las TIC* desde el año 2019 —su última versión es de 2023— (Su título en inglés es *Rolling Plan for ITC Standardisation*); en segundo término, el *Bildt report on EU Standardisation* (2019); en tercer término, la *Note from 17 member States to Council on Competitiveness* (2021); o en cuarto término, la consulta pública realizada por la Comisión sobre una nueva estrategia de normalización europea (El título de esta nueva estrategia en inglés es *Roadmap for a new European Standardisation Strategy* —junio de 2021—).

34. M. McFadden y otros, *Harmonising ...cit.* p. 4.

35. *Ibidem*, p. 18.

B) El segundo de los grandes desafíos consiste en la necesidad de asegurar una significativa participación en la elaboración de las normas técnicas armonizadas europeas por parte de los entes encargados de proteger los derechos fundamentales y los intereses públicos. Debe tenerse en cuenta a este respecto que uno de los grandes objetivos del RIA consiste en establecer una serie de requisitos obligatorios para los sistemas de inteligencia artificial de alto riesgo con el objetivo de minimizar los potenciales efectos adversos frente a la seguridad, a la salud y a los valores y los derechos fundamentales en la Unión, y, por tanto, debe ser también uno de los grandes objetivos de las normas técnicas armonizadas. Pues bien, por este motivo, resulta evidente que las normas armonizadas europeas que sirvan para el desarrollo de los requisitos esenciales obligatorios establecidos en el RIA «serán más efectivas si están elaboradas con la colaboración de expertos en salud, en seguridad y en derechos fundamentales»³⁶. El problema es si realmente los organismos europeos de normalización están preparados, en el momento presente, para proteger la salud, la seguridad y los valores y los derechos fundamentales en la Unión. Parece, ciertamente, que se impone la necesaria participación de sectores (y, por tanto, de expertos) dentro de los organismos de normalización que tradicionalmente no han estado implicados en las tareas normalizadoras.

C) El tercer gran reto se centra en la necesidad de que las normas armonizadas sean «lo suficientemente flexibles para reflejar la rápida evolución de la tecnología y de los productos de inteligencia artificial». Y es que, en estos momentos, existe «una discordancia entre la velocidad de despliegue de los productos y servicios basados en inteligencia artificial y el desarrollo de las normas técnicas»³⁷. En este orden de cosas, debería simplificarse y agilizarse el proceso de aceptación y de publicación de las normas armonizadas por parte de la Comisión.

D) El cuarto desafío se basa en la necesidad de fortalecer las relaciones de cooperación entre los organismos europeos e internacionales de normalización. En este ámbito, debe partirse de dos ideas que a veces no son fáciles de compatibilizar: por un lado, las normas armonizadas europeas deben respetar los valores europeos y los derechos fundamentales, según el RIA, algo que no tiene por qué suceder con las normas técnicas internacionales; y, por otro, los europeos estamos interesados en que existan estándares técnicos mundiales, abiertos e interoperables que faciliten el comercio entre la Unión Europea y el resto del planeta. Con este trasfondo, debería evitarse en la medida de lo posible la duplicación de los esfuerzos entre los organismos europeos de normalización y los internacionales generales, de tal forma que se aprovechen dentro de la Unión Europea los estándares internacionales en la medida de lo posible, reduciendo la carga de trabajo de las organizaciones europeas de normalización para concentrar su actividad en aquellos campos donde no existan normas internacionales, además de contribuir a eliminar (o, al menos, reducir) las trabas al comercio internacional. La necesidad de evitar duplicidades entre la normalización europea e internacional no es, ciertamente, nueva. Para ello, se han desarrollado los acuerdos de Viena y de Frankfurt entre, por un lado,

36. *Ibidem*, p. 19.

37. *Ibidem*, p. 18.

los organismos europeos de normalización (CEN y CENELEC) y, por otro, los internacionales (ISO y CEI). Recuerda el estudio *Harmonising Artificial Intelligence: The role of standards in the EU AI Regulation* (p. 21) que estos acuerdos dan prioridad a los trabajos normalizadores del conglomerado ISO/CEI frente al conformado por CEN/CENELEC, pero el proyecto de RIA «podría cambiar esto en la práctica, poniendo a CEN/CENELEC en el asiento del conductor»³⁸, de tal forma que este conglomerado europeo dirija la normalización internacional defendiendo los valores y los principios europeos. A este respecto, el estudio señala que, entre los varios factores que podrían contribuir a este resultado, se encuentran, por ejemplo, los fuertes incentivos para los operadores económicos mundiales de producir respetando las normas técnicas armonizadas europeas, puesto que les permitiría reducir los costes de evaluación de la conformidad en la Unión Europea, pudiendo acceder de esta manera más fácilmente al amplio mercado comunitario³⁹.

E) El quinto reto consiste en la necesidad de proceder al desarrollo de mejores herramientas de control del cumplimiento de los estándares técnicos (en particular, de las normas armonizadas) en cooperación con los expertos en normalización, la industria y los proveedores de productos.

7. LOS RESULTADOS ACTUALES DEL TRABAJO DE NORMALIZACIÓN EN INTELIGENCIA ARTIFICIAL

A pesar de que la normalización nacional, europea e internacional se encuentra muy desarrollada en la mayor parte de la industria destinada a la fabricación de productos físicos, no ha sucedido lo mismo en relación con la inteligencia artificial a pesar de que esta familia de tecnologías ha alcanzado ya un uso considerable. La normalización técnica de la inteligencia artificial todavía está en sus primeros momentos de vida.

Aunque hay variados sujetos capaces de elaborar normas técnicas en el mundo de la inteligencia artificial, los únicos con competencia, desde un punto de vista jurídico, para generar normas armonizadas son los organismos europeos de normalización. En esto no se diferencia en nada la inteligencia artificial de cualquiera de los productos físicos afectados por la técnica armonizadora europea del nuevo enfoque.

En la medida en que se acaba de aprobar el RIA, no se han podido, hasta ahora, dictar propiamente normas armonizadas específicas para su desarrollo, si bien es cierto que tanto el ETSI como el conglomerado CEN/CENELEC ya han empezado a trabajar sobre esta cuestión desde hace meses, como se demuestra con la consulta de sus agendas normalizadoras. El primero de estos entes se está centrando preferentemente en el ámbito de la seguridad, mientras que los segundos están trabajando más en los aspectos de la confianza y de la ética⁴⁰. Para el desarrollo de su trabajo normalizador en la materia que nos ocupa, el ETSI creó en 2018 el *Industry Specification Group on Securing Artificial Intelligence* (ETSI

38. *Ibidem*, p. 21.

39. *Ídem*.

40. *Ibidem*, p. 10.

ISG SAI), entre cuyos miembros fundadores estaba Telefónica⁴¹. CEN/CENELEC instauró en 2019, por su parte, el *Joint Technical Committee 21 on Artificial Intelligence (JTC 21)*⁴².

No obstante, el nivel territorial en el que más desarrollada se encuentra en estos momentos la normalización en materia de inteligencia artificial es el internacional general, gracias al trabajo del conglomerado ISO/IEC⁴³, que cuenta con el *Joint Technical Committee 1 (ISO/IEC JTC 1)*, destacando dentro de él, a nuestros efectos, el Subcomité *SC 42 on Artificial Intelligence*⁴⁴.

Téngase en cuenta que la normalización internacional es esencial para la normalización europea, pues muchas de las normas técnicas europeas del CEN/CENELEC tienen su base en normas internacionales elaboradas por ISO/IEC⁴⁵. La incorporación de las normas del ISO y del IEC por sus homólogos organismos europeos de normalización se ha visto facilitada por los acuerdos de Viena y de Frankfurt, que sirven para ordenar las relaciones entre todos estos organismos normalizadores⁴⁶. Esta estrecha relación entre la normalización europea y la internacional hace suponer más que fundadamente que pueda ocurrir algo similar en el ámbito de la inteligencia artificial. El problema es que, mientras que en la Unión Europea el RIA exige que los organismos europeos de normalización elaboren las normas armonizadas respetando los valores europeos y los derechos fundamentales, no tienen por qué hacerlo así las entidades internacionales generales.

41. Téngase presente que este ente normalizador está constituido por más de 900 miembros de más de sesenta países. A diferencia de lo que sucede con el CEN y el CENELEC, que únicamente están conformados por organismos nacionales de normalización, pueden ser miembros del ETSI todos aquellos sujetos u organismos que tengan interés en la normalización en el mundo de las telecomunicaciones. Véase, a este respecto, Álvarez García, V., *La normalización industrial... cit.* pp. 367 y ss.

42. Entre los diversos textos normativos en materia de inteligencia artificial sobre los que este JTC 21 se encuentra trabajando pueden citarse, a título de ejemplo, estos dos: el *prCEN/CLC/TR 17894 (WI=JT021001) Evaluación de conformidad de inteligencia artificial*; o el *prCEN/CLC/TR XXXX (WI=JT021002) Inteligencia artificial: descripción general de las tareas y funcionalidades de inteligencia artificial relacionadas con el procesamiento del lenguaje natural*.

43. IEC son las siglas en inglés de la Comisión Electrotécnica Internacional.

44. Son varias las normas técnicas ya proyectadas por ISO/IEC en este ámbito. Por citar tan sólo algunos ejemplos, indicaremos las cinco siguientes: la *ISO/IEC 22989 Inteligencia Artificial-Conceptos y terminología*; la *ISO/IEC 23894 Tecnología de la información—Inteligencia Artificial—Gestión del riesgo*; la *ISO/IEC 24668 Tecnología de la información—Inteligencia Artificial—Marco de gestión de procesos para analítica mediante Big Data*; la *ISO/IEC 38507 Tecnología de la información—Gobernanza de TI—Implicaciones de gobernanza del uso de la inteligencia artificial por las organizaciones*; o la *ISO/IEC 23053 Tecnología de la información-Inteligencia Artificial-Evaluación del rendimiento de clasificación de los modelos machine learning*.

45. McFadden, M. y otros, *Harmonising... cit.* p. 31: «De las aproximadamente 3500 normas técnicas del CEN/CENELEC citadas en el DOUE, el 44% están basadas en normas internacionales».

46. Un breve apunte sobre las bases en las que se asientan las relaciones entre la normalización internacional y la europea puede verse en Álvarez García, V., *La normalización industrial... cit.* pp. 439 y ss.

A nivel internacional general, además de los organismos internacionales de normalización (ISO y IEC), cuyo trabajo esencial es la elaboración de normas técnicas, existen otras organizaciones bien de naturaleza pública o bien de carácter privado que también pueden generar estándares técnicos como parte accesoria a su función principal. Entre las entidades de este tipo que elaboran especificaciones técnicas en materia de inteligencia artificial pueden indicarse las tres siguientes: en primer término, la ITU-T (*International Telecommunications Union-Telecommunication Standardisation Sector* o, en español, Unión Internacional de Telecomunicaciones —Sector de Normalización de las Telecomunicaciones—); en segundo lugar, el IEEE (*Institute of Electrical and Electronics Engineers*); y en tercer lugar, el consorcio W3C o *World Wide Web Consortium*.

En fin, desde 2021 la Comisión Europea hace un seguimiento de la labor normalizadora en materia de inteligencia artificial realizada por todos los organismos citados a través de su servicio *AI Watch*, que se refleja en su informe *Artificial Intelligence Standardisation Landscape*⁴⁷ de ese año, constando este texto de una última versión en 2023⁴⁸.

III. LAS ESPECIFICACIONES COMUNES

1. UNAS IDEAS INICIALES SOBRE SU CONCEPTO

A diferencia de las normas armonizadas que ya cuentan con una larga tradición en la Unión, las especificaciones comunes⁴⁹ tienen una historia realmente mucho más reciente (aunque, ciertamente, con algún precedente aislado⁵⁰), pues es un instrumento jurídico que sólo ha sido empleado con visos de una cierta generalización desde el año 2017 a través de dos actos legislativos: el Reglamento (UE) 2017/745 del Parlamento Europeo y del Consejo, de 5 de abril de 2017, sobre los productos sanitarios; y el Reglamento (UE) 2017/746 del Parlamento Europeo y del Consejo, de 5 de abril de 2017, sobre los productos sanitarios para diagnóstico *in vitro*⁵¹.

47. Su título completo es *AI Watch: Artificial Intelligence Standardisation Landscape: state of play and link to the EC proposal for an AI regulatory framework*. Sus autores son S. Nativi y S. De Nigris.

48. Esta última edición se ha publicado con el título *AI Watch: Artificial Intelligence Standardisation Landscape Update*.

49. Sobre este tipo de instrumentos normativos, pueden verse Álvarez García V. y Tahiri Moreno, J. «La regulación de la inteligencia ...» cit. y «Los instrumentos normativos ...» cit.

50. Es verdad que ya el art. 5.3 de la Directiva 98/79/CE del Parlamento Europeo y del Consejo, de 27 de octubre de 1998, sobre productos sanitarios para diagnóstico *in vitro*, contemplaba la adopción de este instrumento jurídico. En base a dicho precepto, la Comisión aprobó, incluso, diversas especificaciones técnicas comunes mediante su Decisión 2002/364/CE, de 7 de mayo de 2002.

51. Ya en el año 2023 se han aprobado tanto el Reglamento (UE) 2023/1230 del Parlamento Europeo y del Consejo, de 14 de junio de 2023, relativo a las máquinas, como el Reglamento (UE) 2023/1542 del Parlamento Europeo y del Consejo, de 12 de julio de 2023, relativo a las pilas y baterías y sus residuos, donde se regula este instrumento normativo de las especificaciones comunes.

Esta diferencia temporal hace que hayamos podido estudiar aceptablemente bien el perfil de las normas armonizadas, explicando sus ventajas, y también sus serios problemas jurídicos⁵², pero que conozcamos poco, ciertamente, sobre cuáles son los contornos y cómo están llamadas a operar las especificaciones comunes⁵³.

El RIA recoge la siguiente definición de especificación común: es, dice el artículo 3.28 de este texto, «un documento, distinto de una norma, con soluciones técnicas que proponen una forma de cumplir determinados requisitos y obligaciones establecidos en el presente Reglamento»⁵⁴.

Esta definición nos permite entender únicamente que las especificaciones comunes son unas disposiciones generales que, alejadas de las normas (armonizadas), establecen estándares técnicos, que ofrecen una forma de cumplir con los requisitos esenciales establecidos de manera imperativa por el acto legislativo nuevo enfoque al que estos instrumentos están destinados a desarrollar.

Este concepto es, evidentemente, bien poco concreto. No responde, entre otras muchas cuestiones básicas, a quién debe elaborar este tipo de documento técnico, ni cómo debe hacerse, ni cuáles son sus efectos.

Pues bien, a falta de una regulación general (como la que existe con respecto a las normas armonizadas en el Reglamento sobre la normalización europea de 2012), las respuestas a todos estos interrogantes deben buscarse en la regulación que de este instrumento efectúa el artículo 41 RIA.

2. LA EVOLUCIÓN DE LA REGULACIÓN DE LA FIGURA DE LAS ESPECIFICACIONES COMUNES DURANTE LA TRAMITACIÓN DE LA PROPUESTA DE REGLAMENTO: DESDE EL PROYECTO DE LA COMISIÓN HASTA LAS ENMIENDAS DEL PARLAMENTO EUROPEO, PASANDO POR EL TEXTO TRANSACCIONAL DEL CONSEJO

A) La propuesta de la Comisión sobre el RIA dedicó su artículo 41 a las especificaciones comunes.

52. Álvarez García, V., *Las normas técnicas ...cit.*

53. Ejemplos del empleo de este tipo de actos de ejecución del Derecho de la Unión los encontramos en el Reglamento de ejecución (UE) 2020/1207 de la Comisión, de 19 de agosto de 2020, por el que se establecen disposiciones de aplicación del Reglamento (UE) 2017/745 del Parlamento Europeo y del Consejo, en lo referente a las especificaciones comunes para el reprocesamiento de productos de un solo uso; el Reglamento de ejecución (UE) 2022/2346 de la Comisión, de 1 de diciembre de 2022, por el que se establecen especificaciones comunes para los grupos de productos sin finalidad médica prevista enumerados en el anexo XVI del Reglamento (UE) 2017/745 del Parlamento Europeo y del Consejo, sobre los productos sanitarios; y el Reglamento de ejecución (UE) 2022/1107 de la Comisión, de 4 de julio de 2022, por el que se establecen especificaciones comunes para determinados productos sanitarios para diagnóstico *in vitro* de la clase D de conformidad con el Reglamento (UE) 2017/746 del Parlamento Europeo y del Consejo.

54. La definición es similar a la recogida en los Reglamentos 2017/745 y 2017/746. Así, en el artículo 2 de ambos textos (apartados 71 y 74, respectivamente), se definen las especificaciones comunes como «un conjunto de requisitos técnicos o clínicos, distintos de una norma, que proporciona un medio para cumplir las obligaciones jurídicas aplicables a un producto, proceso o sistema».

a) Este precepto otorga, en primer término, una grandísima discrecionalidad a la Comisión para la adopción de estos instrumentos normativos al prever que esta Alta Institución podría generarlos cuando no existan normas armonizadas «o cuando la Comisión considere que las normas armonizadas pertinentes son insuficientes o que es necesario abordar cuestiones específicas relacionadas con la seguridad o con los derechos fundamentales».

b) En segundo término, dispone que las especificaciones comunes tendrán la naturaleza de actos ejecución adoptados mediante el procedimiento de examen, en cuya tramitación estará obligada a recabar los puntos de vista de organismos o de grupos de expertos.

c) En tercer término, se dota a estos instrumentos normativos del efecto jurídico-público de la presunción de conformidad.

d) En cuarto término, se prevé que los proveedores puedan recurrir a otras soluciones técnicas alternativas a las especificaciones comunes, siempre que como mínimo aquéllas sean equivalentes a éstas.

B) Las variantes introducidas por el texto transaccional del Consejo a la regulación de las especificaciones comunes establecidas por la propuesta de RIA.

Las modificaciones propugnadas por el Consejo a la proyectada regulación de las especificaciones comunes son bien significativas, teniendo como objetivo «limitar la discrecionalidad de la Comisión»⁵⁵, al especificar cuándo la Comisión podría elaborar este tipo de documentos técnicos, cómo se desarrollará el procedimiento para su adopción o cuáles serán las relaciones entre las normas armonizadas y las especificaciones comunes. Debe tenerse en cuenta, en todo caso, que la regulación de las especificaciones comunes prevista en el texto transaccional del Consejo no sólo afecta a los sistemas de inteligencia artificial de alto riesgo, sino también a los de uso general.

a) En relación con los supuestos en los que puede dictar la Comisión especificaciones comunes, el Consejo propugna que únicamente podría hacerse cuando, habiéndose solicitado previamente a los organismos europeos de normalización la elaboración de normas armonizadas, el mandato de la Comisión no haya sido aceptado por estas entidades, cuando las normas no se hayan presentado por éstas en el plazo establecido o, finalmente, cuando, habiéndose elaborado efectivamente una norma, ésta no se ajuste a dicho mandato.

55. Introducción al texto transaccional del Consejo, p. 7. Esta idea se complementa en la nueva redacción del considerando 65 de dicho texto transaccional, que subraya el carácter excepcional que debe tener la adopción de las especificaciones comunes con los términos siguientes: «a falta de referencias pertinentes a las normas armonizadas, la Comisión debe poder establecer, mediante actos de ejecución, especificaciones comunes para determinados requisitos previstos en el presente Reglamento como solución alternativa excepcional para facilitar la obligación del proveedor de cumplir los requisitos del presente Reglamento, cuando el proceso de normalización esté bloqueado o cuando haya retrasos en el establecimiento de una norma armonizada adecuada. Si dichos retrasos se deben a la complejidad técnica de la norma en cuestión, la Comisión debe tenerlo en cuenta antes de considerar la posibilidad de establecer especificaciones comunes».

b) Con respecto al procedimiento para su adopción, el Consejo prevé que sea el de examen, tal y como propone la Comisión, pero con los siguientes condicionantes: en primer término, esta última Institución, al elaborar las especificaciones debe cumplir con los mismos objetivos que, según el Consejo, deben exigirse a los organismos europeos de normalización en los mandatos formulados por la Comisión para la adopción de las normas armonizadas⁵⁶; en segundo término, la Comisión debería recabar los puntos de vista de los organismos o grupos de expertos pertinentes; en tercer término, esta Institución debería efectuar una previa consulta al Comité Europeo de Inteligencia Artificial; y, en cuarto término, debería informar al comité previsto en el artículo 22 del Reglamento sobre la normalización de 2012 (compuesto por representantes de los Estados miembros y presidido por un representante de la Comisión) de que se cumplen los requisitos para adoptar una especificación común.

c) En lo que se refiere a las relaciones entre las especificaciones comunes y las normas armonizadas, el Consejo prevé una suerte de primacía de estas últimas sobre aquéllas, al establecer que: «Cuando las referencias de una norma armonizada se publiquen en el DOUE, se derogarán, según proceda», las especificaciones comunes⁵⁷.

d) En fin, el efecto jurídico-público atribuido a las especificaciones comunes por el texto del Consejo es el de la presunción de conformidad. Ahora bien, a diferencia de lo que sucedía con la propuesta de Reglamento de la Comisión (en la que se preveía que, en caso de que los operadores económicos no utilizaran las especificaciones comunes, podrían recurrir válidamente a otras soluciones técnicas «como mínimo equivalentes» a las establecidas en ellas), el texto del Consejo guarda silencio sobre esta cuestión. Creo, en todo caso, que no parece que haya dudas de que el hecho de que no se fabrique conforme a normas armonizadas o a especificaciones comunes no impide, desde el punto de vista jurídico, que puedan utilizarse, efectivamente, otras soluciones técnicas, aunque ello suponga *de facto* mayores cargas burocráticas y económicas para los proveedores a la hora de introducir sus sistemas de inteligencia artificial en el mercado interior europeo.

C) El Parlamento Europeo formula varias enmiendas al texto de la Comisión (desde, en concreto, la 442 a la 448).

a) Entre estas enmiendas, destacan las referidas, en primer término, a la limitación de la discrecionalidad de Comisión a la hora de regular los supuestos en los que procede la adopción de especificaciones comunes, puesto que se exige que no haya una previa norma armonizada y que se haya solicitado previamente a los

56. Recordemos que estos objetivos, recogidos en el art. 40.2 del texto transaccional del Consejo, y que ya hemos referido con anterioridad, son: en primer término, la garantía de que los sistemas de inteligencia artificial de alto riesgo sean seguros, respeten los valores de la Unión y garanticen «su autonomía estratégica abierta»; en segundo término, la promoción de «la inversión y la innovación en inteligencia artificial, incluso mediante el incremento de la certidumbre jurídica, así como la competitividad y el crecimiento del mercado de la Unión»; en tercer término, el fomento de la participación («gobernanza») de todas las partes interesadas en la normalización (desde, por ejemplo, la industria hasta la sociedad civil, pasando por las pymes y los investigadores); y, en cuarto término, el reforzamiento de la cooperación «mundial» en una normalización de la inteligencia artificial, de manera «que sea coherente con los valores e intereses de la Unión».

57. Art. 41.4 del texto transaccional del Consejo.

organismos europeos de normalización su adopción (y estos no cumplan con esta tarea adecuadamente). La Comisión deberá justificar, además, las razones por las que haya decidido recurrir a las especificaciones comunes.

b) En segundo término, la Comisión deberá consultar con carácter también previo a su adopción a la Oficina de Inteligencia Artificial y al foro consultivo, a los organismos europeos de normalización, a los grupos de expertos establecidos conforme a la legislación sectorial de la Unión y a otras partes pertinentes. Cuando la Comisión decida no seguir el dictamen emitido por la referida Oficina de Inteligencia Artificial deberá proporcionarle una explicación motivada.

c) En tercer término, la Comisión deberá justificar el cumplimiento de los mismos objetivos que el Parlamento Europeo quiere que se exijan para la elaboración de las normas armonizadas: 1) Deberá tener en cuenta los principios generales establecidos por el Reglamento para una inteligencia artificial fiable; 2) Tratará de promover la inversión, la innovación, la competitividad y el crecimiento del mercado de la inteligencia artificial en el ámbito de la Unión; 3) Contribuirá a reforzar la cooperación mundial en materia de normalización, teniendo en cuenta las normas internacionales sobre inteligencia artificial, cuando «sean coherentes con los valores, los derechos fundamentales y los intereses de la Unión»; y 4) Garantizará una participación equilibrada y efectiva de todas las partes interesadas en el ámbito de la normalización de la inteligencia artificial.

d) En cuarto término, el Parlamento Europeo propone la siguiente regla de solución de conflictos entre las normas armonizadas y las especificaciones comunes: «Cuando se publique la referencia de una norma armonizada en el Diario Oficial de la Unión Europea, la Comisión derogará los actos de ejecución» que aprueben las especificaciones comunes o partes de ellas, en la medida en que regulen la misma materia.

e) En quinto término, las previsiones sobre las especificaciones comunes se aplican tanto para las que desarrollan los requisitos esenciales de los sistemas de inteligencia artificial de alto riesgo como los de los modelos fundacionales.

3. LOS ELEMENTOS ESENCIALES QUE CONFIGURAN LAS ESPECIFICACIONES COMUNES EN EL REGLAMENTO

A) La competencia para la elaboración y para la adopción de estos instrumentos corresponde a la Comisión. En efecto, a diferencia de lo que sucede con las normas armonizadas, las especificaciones comunes son documentos técnicos de naturaleza pública en su totalidad, puesto que son generados íntegramente por la Comisión, que es quien tiene la iniciativa, elabora el texto y lo aprueba. Recordemos que las normas armonizadas se generaban por iniciativa (mandato) de la Comisión, pero se elaboraba su contenido por los organismos europeos de normalización, aunque dependía de la aceptación de la Comisión y de la publicación oficial de sus referencias el que estas normas pudiesen gozar del efecto jurídico-público de la presunción de conformidad.

B) ¿En qué supuestos podrían adoptarse especificaciones comunes? La tramitación de la propuesta de RIA reveló importantes divergencias de apreciación a la hora de decidir cuándo procede dictar una especificación común entre, por un lado, la Comisión y, por otro, el Consejo y el Parlamento Europeo.

Frente a la tesis de la Comisión en la que propugnaba su gran discrecionalidad para dictar este tipo de documentos técnicos cuando considerase oportuno (sin importar, por ejemplo, que ya existiesen normas armonizadas sobre la concreta cuestión), el texto definitivo del artículo 41 RIA restringe esta posibilidad, exigiendo la concurrencia de los siguientes tres requisitos: 1) Que no haya elaborada ya una norma armonizada; 2) Que la Comisión haya formulado ya el oportuno mandato de normalización a uno o varios organismos europeos de normalización; y 3) Que, alternativamente, dichos organismos no hayan aceptado dicha petición, o se produzcan retrasos indebidos en la aprobación de la norma armonizada, o ésta no se ajuste al mandato de la Comisión.

C) El procedimiento de elaboración es el de examen. Este procedimiento de examen está regulado en el artículo 5 del Reglamento (UE) n.º 182/2011 del Parlamento Europeo y del Consejo, de 16 de febrero de 2011, por el que se establecen las normas y los principios generales relativos a las modalidades de control por parte de los Estados miembros del ejercicio de las competencias de ejecución por la Comisión. Este acto legislativo europeo impone que la Comisión, antes de la adopción del correspondiente acto de ejecución, obtenga un dictamen favorable sobre su proyecto del pertinente comité compuesto por representantes de los Estados miembros y de la propia Comisión (que preside el comité, aunque carece de voto). La mayoría requerida para sacar adelante este dictamen es la cualificada prevista en el apartado 3 del artículo 238 TFUE.

Durante la tramitación del procedimiento de elaboración de las especificaciones comunes, la Comisión estará obligada a consultar a diferentes expertos en los ámbitos de la inteligencia artificial y de la normalización. En este sentido, será necesaria la consulta a la Oficina de Inteligencia Artificial y al foro consultivo, a los organismos europeos de normalización, a los grupos de expertos establecidos en virtud del Derecho sectorial pertinente de la Unión y a otras partes interesadas.

D) La forma jurídica que habrán de adoptar las especificaciones comunes

Esta forma jurídica será la de acto de ejecución de la Comisión. Debe recordarse, a este respecto, que los actos de ejecución del Derecho de la Unión Europea tienen su base jurídica en el artículo 291 TFUE. Este precepto establece, como regla general, que las medidas de ejecución de los actos jurídicamente vinculantes de la Unión corresponden a los Estados miembros, que adoptarán a tal efecto todas las medidas necesarias de Derecho interno (apartado 1). No obstante, en los supuestos que «requieran condiciones uniformes de ejecución de los actos jurídicamente vinculantes de la Unión, éstos conferirán competencias de ejecución a la Comisión» (o, en algunos casos limitados, al Consejo) (apartado 2), que serán ejercidas, por un lado, por las Instituciones europeas y estarán sometidas, por otro, a las modalidades de control por los Estados miembros, conforme a las reglas establecidas en reglamentos europeos aprobados por el procedimiento legislativo ordinario (apartado 3). Pues bien, la norma de Derecho derivado reguladora de esta cuestión en la actualidad es el ya referido hace unos instantes Reglamento (UE) n.º 182/2011.

Estos actos de ejecución de la Comisión (y, por lo tanto, las especificaciones comunes) se deben publicar íntegramente en el Diario Oficial de la Unión Europea (artículo 297.2 TFUE). Recuérdese que esta situación resulta bien diferente de lo que sucede con las normas armonizadas, que, a pesar de que tienen como efecto

jurídico-público la presunción de conformidad (como sucede con las especificaciones comunes), tienen una publicidad oficial restringida a sus referencias (esto es, a sus códigos numéricos y a sus títulos).

E) El valor jurídico y los efectos de las especificaciones técnicas.

La regla general que parece deducirse del RIA es que estos instrumentos normativos son voluntarios desde una perspectiva jurídica. Me explico: los operadores económicos que quieran introducir en el mercado o comercializar un software de inteligencia artificial deben respetar los requisitos esenciales establecidos imperativamente por su acto legislativo regulador nuevo enfoque, y estas exigencias obligatorias pueden realizarse siguiendo bien las normas armonizadas que puedan existir (elaboradas por los organismos europeos de normalización), bien las especificaciones comunes (generadas por la Comisión) o bien mediante otras soluciones técnicas (establecidas, por ejemplo, por un operador privado o por un organismo de normalización de ámbito nacional —como UNE en España— o internacional general —como ISO—) que puedan presentar un nivel de calidad y de seguridad equivalente cuanto menos al establecido en las especificaciones comunes existentes en este ámbito. En otros términos, las especificaciones comunes ofrecen soluciones técnicas que tienen alternativas que pueden ser libremente escogidas desde un punto de vista jurídico por cada operador económico.

La decisión de elaborar un sistema de inteligencia artificial siguiendo las especificaciones comunes (al igual que sucede al hacerlo con las normas armonizadas) tiene como particularidad su importante efecto jurídico. Y es que, en efecto, la consecuencia jurídico-pública de que los operadores económicos utilicen las especificaciones comunes es nada más y nada menos que la presunción de conformidad de que dichos sistemas generados siguiendo sus prescripciones técnicas respetan los requisitos esenciales establecidos por el acto legislativo nuevo enfoque aplicable, y cuyo cumplimiento, recordemos, es indispensable para poder comercializarlos dentro del mercado europeo.

En definitiva, a pesar del carácter jurídico-público ligado a la autoridad competente para la producción de las especificaciones comunes y del procedimiento utilizado para elaborarlas, este instrumento jurídico no es obligatorio desde el punto de vista jurídico, sino que se le confiere «únicamente» el mismo efecto jurídico-público que a las normas armonizadas, al que nos acabamos de referir, es decir, su presunción de conformidad. Esto significa que, aunque existan estas especificaciones comunes y/o aunque existan normas armonizadas, los operadores de inteligencia artificial podrán adoptar otras soluciones técnicas alternativas para la elaboración de sus sistemas, debiendo justificar, eso sí, que en su elaboración «han adoptado soluciones técnicas como mínimo equivalentes» a las establecidas en las normas armonizadas y en las especificaciones comunes.

G) Las relaciones jurídicas entre las especificaciones comunes y las normas armonizadas.

Hemos visto que la Comisión adopta especificaciones comunes cuando no existen normas técnicas o cuando, existiendo éstas, las mismas son insuficientes para regular una materia. En otras palabras, las especificaciones comunes sirven para salvar las lagunas que han dejado los organismos europeos de normalización.

No parece sensato que cuando se adopten especificaciones comunes sobre una concreta cuestión existan normas armonizadas reguladoras de la misma. La Comisión tiene en sus manos bien la facultad de no publicar en el Diario Oficial comunitario las referencias de las normas armonizadas elaboradas por los entes normalizadores europeos si no está de acuerdo con ellas o bien la potestad de retirar tales referencias de dicho Diario Oficial si ya están publicadas⁵⁸. La Comisión, en otros términos, dispone de la llave para evitar este tipo de eventuales conflictos.

En todo caso, parece que debería establecerse una regla de resolución de potenciales conflictos entre ambas categorías normativas. Esto es lo que parece hacer el RIA. Este acto legislativo prevé, en efecto, «una suerte de primacía» (o, incluso, de «jerarquía») de las normas técnicas sobre las especificaciones comunes, al establecer que: «Cuando se publique la referencia de una norma armonizada en el Diario Oficial de la Unión Europea, la Comisión derogará» las especificaciones comunes. En otros términos, la derogación de las especificaciones comunes (normas jurídico-públicas) por las normas armonizadas (normas de origen privado) no es automática, sino que únicamente se impone a la Comisión la obligación de derogar las especificaciones comunes contrarias a las normas armonizadas.

Pero, a la luz de este texto, las dudas siguen persistiendo: ¿qué sucedería en la situación inversa (esto es, si se aprueban especificaciones comunes posteriores a normas armonizadas)? ¿Hay una relación de jerarquía entre normas armonizadas y especificaciones comunes? Las respuestas a estas cuestiones son realmente inciertas, y sólo resulta posible deducirlas a través del sentido común jurídico, dado que no tenemos elementos ni normativos ni jurisprudenciales que nos permitan acometer dicha tarea.

H) La necesidad de proceder a una regulación general de las especificaciones comunes.

Los actos legislativos nuevo enfoque han sido desarrollados tradicionalmente por normas armonizadas, pero, en la actualidad, parece que éstas han encontrado unas compañeras inseparables en las especificaciones comunes. Desde la regulación de los productos sanitarios en 2017 y, sobre todo, con los recientes Reglamentos de máquinas y de baterías (ambos de 2023), los requisitos esenciales que deben cumplir

58. El art. 11.1 del Reglamento sobre la normalización europea de 2012 (modificado por el art. 48 del Reglamento relativo a la Seguridad General de los Productos de 2023) prevé esta posibilidad. Dice este precepto, en efecto, que: «1. Cuando un Estado miembro o el Parlamento Europeo consideren que una norma armonizada o una norma europea elaborada en apoyo del Reglamento (UE) 2023/988 no cumple del todo los requisitos que está previsto que regule, tal como estén establecidos en la legislación de armonización aplicable de la Unión o en dicho Reglamento, informarán de ello a la Comisión con una explicación pormenorizada. Tras consultar al comité establecido por la correspondiente legislación de armonización de la Unión, en caso de que exista tal comité, o al comité creado por dicho Reglamento, o tras otras formas de consulta de expertos sectoriales, la Comisión decidirá: a) publicar, no publicar o publicar con restricciones las referencias de la norma armonizada o la norma europea en cuestión elaborada en apoyo de dicho Reglamento en el Diario Oficial de la Unión Europea, y b) mantener o mantener con restricciones las referencias de la norma armonizada o la norma europea en cuestión elaborada en apoyo de dicho Reglamento en el Diario Oficial de la Unión Europea o suprimirlas de este».

los productos para ser válidamente introducidos y comercializados en el mercado europeo pueden completarse mediante uno u otro instrumento.

Parece cierto, por tanto, que las especificaciones comunes son un tipo de disposiciones técnicas que han venido para quedarse, puesto que, sin duda, ofrecen una importante ventaja desde una perspectiva práctica. Y es que, al menos sobre el papel, estas especificaciones permitirían rellenar los vacíos que no completan los organismos de normalización europeos⁵⁹, cuando éstos no se encuentren en disposición de adoptar una norma armonizada acorde a las necesidades de la Unión (por ejemplo, cuando no sean capaces de satisfacer las exigencias derivadas de los valores de la Unión o de la protección de los derechos humanos).

Ahora bien, si con respecto a las normas armonizadas el Legislador europeo decidió aprobar un acto transversal que estableciese una ordenación general para todas ellas (me refiero, naturalmente, al Reglamento sobre la normalización europea de 2012), que ha permitido resolver importantes cuestiones jurídicas sobre su elaboración y sobre su aplicación, me parece que debería hacerse lo mismo con las especificaciones comunes. Creo, en efecto, que debería afrontarse la elaboración de una regulación general de las especificaciones comunes, puesto que, analizando las escasas previsiones que sobre ellas existen en los recientes actos legislativos nuevo enfoque que las contemplan, sus perfiles a veces son en extremo difusos, planteándose numerosas dudas jurídicas tanto en su fase de elaboración como en la de aplicación.

4. OTRAS SOLUCIONES TÉCNICAS EQUIVALENTES A LAS OFRECIDAS POR LAS NORMAS ARMONIZADAS Y POR LAS ESPECIFICACIONES COMUNES

En la órbita del nuevo enfoque, las normas armonizadas generadas por los entes normalizadores europeos no han sido tradicionalmente las únicas soluciones técnicas puestas en manos de los operadores económicos para fabricar sus productos desde un punto de vista jurídico, ya que siempre se les ha ofrecido la posibilidad de utilizar otros medios técnicos alternativos para ello. Entre estos medios, se encuentra, por ejemplo, el recurso a normas técnicas internacionales, a normas europeas no armonizadas, a normas nacionales o, simplemente, a especificaciones técnicas establecidas por los propios operadores económicos.

Esta configuración jurídica de voluntariedad ha chocado, no obstante, *de facto* con los costes burocráticos y económicos que los controles de los productos fabricados mediante el uso de otros mecanismos técnicos han tenido. En otras palabras, aunque las normas armonizadas son de uso jurídicamente voluntario, la práctica demuestra que es el sistema seguido preferentemente por los operadores económicos para elaborar sus productos dadas las ventajas que conlleva, ligadas a la presunción de conformidad⁶⁰.

59. En este sentido, se ha afirmado por insignes expertos del mundo de la normalización que: «Las especificaciones comunes actúan como una red o una barrera de seguridad, habilitando a la Comisión a actuar cuando hay una laguna en el espacio de las normas técnicas» [McFadden, M. y otros, *Harmonising... cit.* p. 9].

60. Téngase presente, a este respecto, el dato siguiente recogido en las Conclusiones de la Abogada General Laila Medina presentadas el 22 de junio de 2023, en el asunto «Public.

Esta situación no ha cambiado con la introducción de las especificaciones comunes elaboradas por la Comisión. Tanto éstas como las normas armonizadas son voluntarias jurídicamente, permitiéndose el uso de otras soluciones técnicas «como mínimo equivalentes» a aquéllas. Así lo prevé el RIA. La propuesta de este acto legislativo presentada por la Comisión en abril de 2021 precisaba, a este respecto, que estas soluciones técnicas alternativas pueden «desarrollarse con arreglo a los conocimientos científicos o de ingeniería generales, a discreción del proveedor del sistema de inteligencia artificial de que se trate. Esta flexibilidad reviste una importancia especial, ya que permite a los proveedores de sistemas de inteligencia artificial decidir cómo quieren cumplir los requisitos, teniendo en cuenta el estado de la técnica y los avances tecnológicos y científicos en este campo»⁶¹. Y es que todavía está por ver si en el futuro próximo la normalización será un mecanismo lo suficientemente ágil para establecer normas armonizadas reguladoras de los sistemas de inteligencia artificial, y si, en su defecto, la Comisión podrá suplir las lagunas que puedan dejar pendientes los organismos europeos de normalización; o si, por el contrario, serán más eficaces los operadores económicos en el establecimiento de las especificaciones técnicas necesarias para el desarrollo de un campo que se despliega de una manera tan veloz como lo hace, evidentemente, el de la inteligencia artificial.

Recuérdese, en todo caso, y para finalizar, que cuando se empleen estas soluciones técnicas alternativas a las normas armonizadas o a las especificaciones comunes no existirá presunción de conformidad de los sistemas de inteligencia artificial generados conforme a ellas con los requisitos obligatorios establecidos por el RIA, multiplicándose significativamente las dificultades y los costes de esta demostración a cargo de los proveedores de dichos sistemas.

Resource.Org, Inc., Right to Know CLG contra Comisión Europea, C-588/21 P: «El hecho de que las normas técnicas armonizadas son obligatorias *de facto*, ya que suelen ser el único método aceptado en el mercado para garantizar el cumplimiento del Derecho derivado de la Unión correspondiente, queda confirmado por un estudio encargado por la Comisión: “en la práctica, [las normas técnicas armonizadas] son casi obligatorias para la mayoría de los agentes económicos”. Además, el mismo estudio señala que el precio de las normas técnicas armonizadas es uno de los principales obstáculos en cuanto a su uso efectivo» (punto 45). Este estudio de la Comisión al que alude la Abogada General se explicita en la nota 24 de sus Conclusiones: *EIM Business & Policy Research, Access to Standardisation — Study for the European Commission, [DG] Enterprise and Industry*, 2010.

61. Exposición de Motivos de la propuesta de Reglamento de inteligencia artificial formulada por la Comisión, p. 16.

La evaluación de la conformidad en el diseño y producción de sistemas basados en inteligencia artificial en el contexto del «Nuevo Marco Legislativo»

ADRIÁN PALMA ORTIGOSA

Profesor Ayudante Doctor del Departamento de Derecho Administrativo de la Universitat de València¹

I. INTRODUCCIÓN

El presente trabajo estudia el contenido del RIA que regula el proceso de evaluación de la conformidad de los sistemas de IA². En primer lugar se realiza una aproximación al encuadre jurídico del proceso de evaluación de la conformidad dentro de la legislación europea, el cual, contempla toda una serie de instrumentos que tratan de garantizar que determinados productos puedan comercializarse en la

-
1. Este trabajo se ha llevado a cabo en el marco de los siguientes proyectos de investigación:
«Algorithmical Law» (PROMETEO/2021/009. Financiado por la Generalitat Valenciana. «La regulación de la economía digital: tutela publica de la igualdad y herramientas algorítmicas» (PID2019-108745GB-I00). Ministerio de Ciencia e Innovación. «Derechos y garantías públicas frente a las decisiones automatizadas y el sesgo y discriminación algorítmicas» 2023-2025 (PID2022-136439OB-I00) financiado por el Ministerio de Ciencia e Innovación. «Herramientas algorítmicas para ciudadanos y Administraciones Públicas» (Proyectos de Generación de Conocimiento, Ministerio de ciencia e Innovación, convocatoria 2021, PID2021-126881OB-I00). «Algorithmic Decisions and the Law: Opening the Black Box» (TED2021-131472A-I00) del Plan de Recuperación, Transformación y Resiliencia. Convenio de Derechos Digitales-SEDIA Ámbito 5 y 6 (2023/C046/00228673).
 2. Este trabajo se ha llevado a cabo en el marco de los siguientes proyectos de investigación: «Algorithmical Law» (PROMETEO/2021/009. Financiado por la Generalitat Valenciana. «La regulación de la economía digital: tutela publica de la igualdad y herramientas algorítmicas» (PID2019-108745GB-I00). Ministerio de Ciencia e Innovación. «Derechos y garantías públicas frente a las decisiones automatizadas y el sesgo y discriminación algorítmicas» 2023-2025 (PID2022-136439OB-I00) financiado por el Ministerio de Ciencia e Innovación. «Herramientas algorítmicas para ciudadanos y Administraciones Públicas» (Proyectos de Generación de Conocimiento, Ministerio de ciencia e Innovación, convocatoria 2021, PID2021-126881OB-I00).

Unión Europea con ciertas garantías y requisitos homogéneos. En segundo lugar se analizan los mecanismos de la evaluación de conformidad que ha diseñado el RIA para los diferentes sistemas de IA. En función del tipo de sistema de IA, se aplicará uno u otro procedimiento de verificación del cumplimiento de los requisitos exigidos a los sistemas de IA.

II. EVOLUCIÓN, TRAMITACIÓN Y CONTENIDO FINAL DE LOS ARTÍCULOS DEL REGLAMENTO DE LA INTELIGENCIA ARTIFICIAL IMPLICADOS

Los artículos que regulan el proceso de evaluación de la conformidad contemplados en el RIA son los siguientes:

Considerando	Precepto	Materia
	Artículo 16.f)	Obligación de los proveedores de asegurarse que su sistema se somete a la evaluación de la conformidad antes de que éste se ponga en el mercado.
Considerando 122	Artículos 41.3 y 42	Presunciones de conformidad.
Considerando 123-125 y 128	Artículo 43	Evaluación de la conformidad prevista para cada tipo de sistema de IA.
Considerando 130	Artículo 46	Exenciones de conformidad.
	Anexo VI	Evaluación de la conformidad realizada por el propio proveedor.
	Anexo VII	La evaluación de la conformidad realizada por terceros.

Estos artículos no han sufrido grandes modificaciones desde que se aprobó la propuesta inicial por parte de la Comisión Europea el 21 de abril de 2021³.

Cabe destacar que inicialmente la evaluación de la conformidad no sólo se regulaba en los artículos indicados en la tabla anterior, ya que también aparecía mencionada como obligación de los proveedores en el artículo 19⁴. El contenido de ese artículo fue suprimido durante la tramitación del RIA. Este cambio no ha tenido ninguna trascendencia práctica, ya que el artículo 19 inicial sobre evaluación de la conformidad, ahora suprimido, remitía al proceso de evaluación de la

3. Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de inteligencia artificial) y se modifican determinados actos legislativos de la Unión. De 21 de abril de 2021.

4. El antiguo artículo 19 de la propuesta inicial del RIA establecía la obligación por la cual el proveedor debía asegurarse de llevar a cabo la evaluación de la conformidad indicada en el artículo 43.

conformidad regulado en el artículo 43 y Anexos VI y VII, los cuales no han sido prácticamente modificados desde la versión inicial. Entendemos que esa supresión obedece esencialmente a dos razones: por un lado al hecho de que era una remisión perfectamente prescindible y por otro, porque también se mencionaba esa misma obligación en el apartado e) del artículo 16 del texto inicial del RIA⁵.

III. LA EVALUACIÓN DE LA CONFORMIDAD EN LA LEGISLACIÓN DE LA UNIÓN EUROPEA

El proceso de evaluación de la conformidad contemplado en el RIA y que posteriormente detallaremos sigue el esquema establecido por el llamado «nuevo marco legislativo», en adelante NML. El NML está integrado por varios textos legales europeos que establecen unas bases comunes sobre la comercialización, evaluación y vigilancia de productos en la Unión Europea⁶. Por consiguiente, el legislador europeo, a la hora de legislar sobre un producto, puede tomar como referencia el NML⁷, el cual, contempla una estructura que busca asegurar una evaluación y puesta en el mercado fiable de tales productos y bienes.

La estrategia que ha adoptado la Unión Europea para regular el diseño, la puesta en el mercado y la supervisión de los sistemas de IA es mantener la estructura del NML que lleva ya años implantada para otra serie de productos⁸.

De forma resumida, y siguiendo la estructura marcada por el NML, para que un producto pueda ser comercializado en la UE con unas garantías mínimas de seguridad, las leyes que regulen estos productos deben establecer los siguientes elementos.

En primer lugar, los productos deben cumplir con una serie de *requisitos mínimos técnicos* que los fabricantes han de implementar. Esos requisitos esenciales tienen como objetivo mitigar los principales riesgos que esos productos pueden ocasionar una vez éstos se introduzcan en el mercado. En el caso del RIA, esos requisitos los encontramos en los artículos 8 a 15, entre otros requisitos, calidad de los datos, niveles adecuados de transparencia, métricas de precisión o solidez, etc. El objetivo es que en el diseño se implementen todas las medidas técnicas adecuadas para mitigar los

5. Actualmente ello se regula en el apartado f) del artículo 16 del RIA que establece que los proveedores «se asegurarán de que los sistemas de IA de alto riesgo sean sometidos al procedimiento pertinente de evaluación de la conformidad a que se refiere el artículo 43 antes de su introducción en el mercado o puesta en servicio;»
6. Las tres textos legales que conforman el Nuevo Marco Legislativo son: el Reglamento (CE) n.º 765/2008 del Parlamento Europeo y del Consejo por el que se establecen los requisitos de acreditación y vigilancia del mercado de los productos; la Decisión n.º 768/2008/CE del Parlamento Europeo y del Consejo sobre un marco común para la comercialización de los productos y; el Reglamento (UE) 2019/1020 del Parlamento Europeo y del Consejo relativo a la vigilancia del mercado y la conformidad de los productos.
7. Además del Nuevo Marco Legislativo se ha de tener en cuenta también el llamado Nuevo Enfoque y el Enfoque Global. Un análisis histórico de todo ello puede verse en: V. Álvarez García, *Industria*, Iustel, 2010, pp. 47 y ss. Véase también la Guía azul sobre la aplicación de la normativa europea relativa a los productos de 2022, pp.7 y ss.
8. Así se indica expresamente en diferentes considerandos del RIA. Véase los considerandos 46, 64, 83, 84, 87, 124.

efectos perversos que pueden afectar a los derechos fundamentales, la seguridad o la salud de las personas una vez el sistema de IA comience a utilizarse.

En segundo lugar, se contempla la posibilidad de que los fabricantes puedan utilizar determinadas normas técnicas elaboradas por organismos europeos de normalización⁹ o por la Comisión Europea que específicamente están diseñadas para cumplir con los requisitos técnicos exigidos por la legislación de ese producto. Nos estamos refiriendo a las *normas armonizadas* y *las especificaciones comunes* respectivamente¹⁰. Estas normas se analizan en otro capítulo de esta obra colectiva.

En tercer lugar, una vez el fabricante ha implementado los requisitos, tomando o no como referencia las normas armonizadas o especificaciones comunes, el siguiente paso es que el producto pase una *evaluación de la conformidad* que asegure que éste cumple con las exigencias legales marcadas por la legislación. Cada producto tiene contemplada un tipo de evaluación de la conformidad, si bien, existen ciertos procesos de evaluación de la conformidad similares con especificaciones propias aplicables a cada producto¹¹. Esta evaluación en algunos supuestos se llevará a cabo por una tercera entidad distinta al fabricante.

En cuarto lugar, una vez el producto ha superado la evaluación de la conformidad, corresponde al fabricante *declarar la conformidad* del producto y en su caso establecer el *marcado* de éste. Es en ese momento cuando el producto se podrá comercializar.

Finalmente, en quinto lugar, una vez el producto se ponga en el mercado, el fabricante estará obligado a seguir cumpliendo con los requisitos técnicos exigidos por la legislación aplicable al producto. Además, las autoridades públicas correspondientes tendrán la potestad de vigilar y supervisar que efectivamente ello es así, es lo que se denomina la *vigilancia del mercado*¹².

Nuevo Marco legislativo común	Reglamento Europeo de IA
Cumplimiento mínimo de requisitos	Artículos 8-15
Aplicación de normas armonizadas/ especificaciones comunes	Artículos 40 y 41
Evaluación de la conformidad	Artículo 43
Declaración de la conformidad y marcado CE	Artículos 47 y 48
Vigilancia del Mercado	Artículos 70 y 74 a 84

9. Estos organismos europeos de normalización son CEN, CENELEC y ETSI.

10. Véase entre otros preceptos los artículos 40 y 41 del Reglamento de IA, así como las definiciones indicadas en la Guía azul sobre la aplicación de la normativa europea relativa a los productos de 2022. p. 49.

11. Guía azul sobre la aplicación de la normativa europea relativa a los productos de 2022. pp.74 y ss.

12. La vigilancia del mercado tiene por objeto garantizar que los productos cumplan los requisitos aplicables que proporcionan un alto nivel de protección de los intereses públicos protegidos por la legislación de armonización de la UE.

Productos como las máquinas, los juguetes, los ascensores o los productos sanitarios entre otros siguen la estructura previamente indicada¹³, la cual, también está presente en el RIA para los sistemas de IA.

IV. LAS FORMAS DE EVALUACIÓN DE LA CONFORMIDAD EN EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL

La evaluación de la conformidad es el proceso por el que se demuestra que un producto cumple con los requisitos especificados en una norma o estándar¹⁴. En nuestro caso, a través de la evaluación de la conformidad los proveedores o un organismo externo verifican que un sistema de IA cumple con los requisitos mínimos contemplados en el RIA¹⁵.

El procedimiento de evaluación de la conformidad está integrado por toda una serie de procesos y fases a través del cual se verifica que un producto es conforme con los requisitos de la legislación de armonización exigidos para que tal producto se pueda poner en el mercado con ciertas garantías. En este sentido, si el sistema de IA se somete a una modificación sustancial¹⁶, será necesario volver a someterlo a un nuevo proceso de verificación de la conformidad¹⁷.

El RIA contempla dos grandes procesos de evaluación de la conformidad. La diferencia esencial estriba en saber quién se encarga de verificar que el sistema de IA cumple con las exigencias del RIA antes de su puesta en el mercado. De esta manera, o bien la evaluación de la conformidad la realiza el propio proveedor que ha diseñado el sistema, o bien la lleva a cabo un tercero, el llamado organismo notificado. Cada uno de estos procesos se encuentran descritos en los Anexos VI y VII respectivamente.

Optar por uno u otro procedimiento de evaluación de la conformidad no queda en manos de los proveedores sino que dependerá del tipo de sistema de IA que hayan desarrollado¹⁸.

Es turno de analizar cada uno de los procesos de evaluación de la conformidad que contempla el Reglamento de AI.

1. LA AUTOEVALUACIÓN DE CONFORMIDAD

A través de la autoevaluación de la conformidad o control interno, el proveedor verifica que su sistema de IA cumple con los requisitos del RIA. Todo este proceso

13. El listado completo de productos y componentes de seguridad de productos se mencionan en el considerando 50 del RIA.

14. ISO/IEC 17000:2004. Conformity assessment — Vocabulary and general principles.

15. Evaluación de la conformidad: *el proceso por el que se demuestra si se han cumplido los requisitos establecidos en el capítulo II, sección 2, en relación con un sistema de IA de alto riesgo*. Artículo 3.20. RIA.

16. En este contexto una modificación sustancial será por ejemplo un cambio de sistema operativo, un cambio en la arquitectura del software, un cambio en la finalidad prevista del sistema, etc. Considerando 128.

17. Artículo 43.4 del Reglamento de IA.

18. El artículo 43 del Reglamento de IA describe a qué tipo de evaluación de la conformidad ha de someterse cada uno de los sistemas de IA que entran dentro del ámbito de aplicación del Reglamento de IA.

es enteramente realizado por el proveedor sin intervención de terceros organismos notificados o autoridades públicas.

Conforme al Anexo VI, la autoevaluación de la conformidad está integrada por tres fases o procesos.

En primer lugar, el proveedor ha de verificar que el sistema de gestión de calidad implementado cumple con todos los requisitos exigidos por el RIA. Así, el artículo 17 de esta norma obliga al proveedor a desarrollar todo un conjunto de documentación y la implementación de procedimientos que aseguren que efectivamente dicho sistema de gestión de la calidad es adecuado.

En segundo lugar, corresponde al proveedor evaluar que el sistema de IA cumple con los requisitos esenciales previstos por el RIA tomando como referencia la documentación técnica que se haya elaborado sobre el producto¹⁹. La evaluación de los requisitos obligará al proveedor a desplegar diferentes medidas de evaluación tales como la comprobación de documentación, ensayos, testeos del sistema de IA, etc. Entendemos que todo este proceso deberá documentarse debidamente.

En tercer lugar, y como fase final del proceso de autoevaluación, el proveedor deberá comprobar que el proceso de diseño y la vigilancia post comercialización del sistema de IA a la que se refiere el artículo 72 del RIA son coherentes con la parte de la documentación técnica que hace referencia a dichos procesos²⁰.

Una vez el proveedor haya realizado estos tres procesos de forma óptima, hay que entender que la evaluación de la conformidad se ha superado. Como es lógico, todas estas actuaciones internas deberán documentarse y estar siempre a disposición de la autoridad de vigilancia del mercado que requiera tal información.

La documentación de todo el proceso de autoevaluación es muy importante, ya que por un lado se demuestra que ésta se ha realizado adecuadamente y por otro se confirma que el sistema de IA cumple con las exigencias establecidas por el RIA.

2. LA EVALUACIÓN REALIZADA POR UN ORGANISMO NOTIFICADO

Junto a la autoevaluación, el otro proceso de evaluación de la conformidad previsto por el RIA es aquel en el que interviene un organismo notificado. Recordemos que un organismo notificado es aquel que ha sido acreditado para poder realizar evaluaciones de conformidad sobre sistemas de IA sometidos al RIA y notificado como tal a la Comisión Europea²¹.

El Anexo VII del RIA establece las fases que integran la participación del organismo notificado en el proceso de verificación de los sistemas de IA. Esta evaluación comprende esencialmente el examen del sistema de gestión de la calidad y la documentación técnica del sistema de IA²².

19. Para más información sobre el contenido y los elementos esenciales de la documentación técnica consúltese el artículo 11 y el Anexo IV del Reglamento de IA.

20. Véase los apartados 2 y 9 del Anexo IV del Reglamento de IA.

21. Véase el trabajo de esta obra colectiva que analiza el papel y funciones de los organismos notificados y en su caso el de las autoridades notificantes.

22. Se habilita a la Comisión Europea para que mediante acto delegado pueda modificar alguna de estas fases. Artículo 43.5 Reglamento de IA.

A) La evaluación del sistema de gestión de la calidad

Por lo que se refiere a la gestión del sistema de calidad, el proveedor deberá dirigir al organismo notificado una solicitud de evaluación del sistema de IA. Entre el contenido que ha de integrar esa solicitud se encuentran los datos identificativos del proveedor, la documentación técnica del sistema de IA elaborada, la documentación relativa al sistema de gestión de la calidad, así como los procedimientos establecidos que aseguren que dicho sistema de calidad se cumplirá. Entendemos que cada organismo notificado dispondrá de diferentes modelos de solicitud o formularios que deberán presentar los proveedores²³.

Presentada la solicitud, corresponde al organismo notificado evaluar si dicho sistema cumple o no con los requisitos del artículo 17 del RIA. Esta decisión ha de notificarse al proveedor o en su caso al representante autorizado de éste. Dicha decisión deberá motivarse e incluirá las conclusiones de la evaluación.

Una vez el sistema de gestión de la calidad ha sido aprobado, éste puede sufrir modificaciones, en estos supuestos, antes de llevar a cabo esas modificaciones, el proveedor habrá de comunicárselo al organismo notificado para que examine los cambios propuestos y en su caso decida si dichos cambios siguen cumpliendo con las exigencias establecidas por el RIA. El organismo notificado comunicará su decisión al proveedor. Esta resolución incluirá las conclusiones del examen de los cambios y la decisión de evaluación motivada.

Además de decidir sobre las potenciales modificaciones que pretenda llevar a cabo el proveedor sobre el sistema de gestión de la calidad, el organismo notificado podrá llevar a cabo diferentes actuaciones de vigilancia de control de dicho sistema de gestión. Entre otras, se autoriza al organismo notificado a acceder a las instalaciones del proveedor, a realizar periódicamente auditorías y a llevar a cabo cuantas pruebas adicionales considere necesarias a efectos de asegurarse que el sistema de IA cumple con el RIA.

B) Análisis de la documentación técnica

Al igual que ocurría con el sistema de gestión de la calidad, para que el organismo notificado pueda evaluar la documentación técnica, el proveedor deberá plantear a éste la solicitud correspondiente. Esta solicitud deberá contener los datos identificativos del proveedor, la documentación técnica, así como una declaración de que éste no ha presentado esta solicitud ante otro organismo notificado. Para el caso de las PYMES, y siempre que estas lo soliciten, los organismos notificados deberán facilitarles el formulario simplificado de documentación técnica elaborado por la Comisión Europea²⁴.

Recibida la solicitud, el organismo notificado se encargará de evaluar la documentación técnica. El RIA contempla diferentes situaciones donde se habilita

23. El Centro Nacional de Certificación de Productos Sanitarios es el único organismo notificado en España para realizar la evaluación de la conformidad de los productos sanitarios conforme al Reglamento 2017/745. La solicitud para la verificación del sistema de gestión de la calidad se encuentra en la siguiente web:

https://certificaps.gob.es/wp-content/uploads/CertificacionMDR/R_DEX_05-Solicitud-de-evaluaci%C3%B3n-del-sistema-de-gesti%C3%B3n-de-calidad.pdf

24. Artículo 11.1 Reglamento Europeo de Inteligencia Artificial.

al organismo notificado a llevar a cabo otras actuaciones cuando estime que la documentación facilitada no es suficiente. En este sentido, cuando el organismo lo considere necesario éste podrá: acceder al conjunto de datos de entrenamiento, validación y prueba utilizados, acceder al modelo de entrenamiento y al modelo entrenado del sistema de IA, obligar al proveedor a realizar pruebas adicionales o en su caso realizarlas él mismo. Entendemos que en todos estos supuestos el organismo notificado deberá justificar las razones por las que considera necesario realizar esas actuaciones que van más allá del acceso a la documentación técnica facilitada inicialmente por el proveedor.

Todo este conjunto de actividades ayudarán al organismo notificado a verificar adecuadamente el cumplimiento de los requisitos previstos en el RIA por parte de un sistema de IA.

Una vez realizada la evaluación de la documentación técnica por parte del organismo notificado de acuerdo a los procesos descritos anteriormente, éste notificará al proveedor o en su caso al representante autorizado la decisión adoptada. Esta decisión indicará si el sistema de IA evaluado cumple o no con los requisitos del RIA. En los casos en los que la decisión sea positiva, el organismo notificado expedirá el certificado de la UE de la documentación técnica. En cambio, si el organismo notificado no considera que el sistema de IA cumple con el Reglamento, éste lo indicará. Cuando la denegación se haya producido porque se considera que los datos utilizados no cumplen con las exigencias del Reglamento, el organismo notificado establecerá consideraciones específicas sobre dichos datos utilizados durante el entrenamiento del sistema de IA y obligará al proveedor a realizar un nuevo entrenamiento antes de que éste último vuelva a solicitar una nueva evaluación de su sistema de IA²⁵.

Una vez se hayan realizado todas las actuaciones previamente mencionadas y se haya obtenido una decisión favorable para todas ellas, la evaluación de la conformidad se entenderá superada.

V. LA EVALUACIÓN DE LA CONFORMIDAD EN FUNCIÓN DEL TIPO DE SISTEMA DE INTELIGENCIA ARTIFICIAL

El artículo 43 del RIA establece el proceso de evaluación de la conformidad que ha de realizar cada proveedor teniendo en cuenta el tipo de sistema de IA que haya diseñado o esté diseñando.

El RIA establece dos grandes grupos de sistemas de IA de alto riesgo. Por un lado aquellos sistemas de IA que se utilizan para una serie de finalidades (finalidades de alto riesgo, Anexo III), y por otro lado aquellos sistemas de IA que son productos o componentes de seguridad de productos que están sometidos a legislación de armonización y en dicha legislación se contempla que la evaluación de la conformidad de esos productos la lleve a cabo un organismo notificado (productos de alto riesgo, Anexo I)²⁶.

Dependiendo del tipo de sistema de IA se aplicará un proceso u otro de evaluación de la conformidad.

25. Véase el Apartado 4.6 Anexo VII del Reglamento de IA.

26. Véase el artículo 6 junto con los Anexos I y III del Reglamento de IA.

1 SISTEMAS DE INTELIGENCIA ARTIFICIAL CUYA FINALIDAD ES CONSIDERADA DE ALTO RIESGO (FINALIDADES DE ALTO RIESGO)

El anexo III del Reglamento Europeo de Inteligencia Artificial establece un listado de finalidades que las considera de alto riesgo cuando se llevan a cabo por un sistema de IA. El proceso de evaluación de la conformidad difiere en parte según el tipo de finalidad para la cual se pretenda utilizar el sistema de Inteligencia Artificial. Así, hemos de distinguir la evaluación de la conformidad que se contempla por un lado para los sistemas de alto riesgo cuya finalidad es la identificación biométrica y, por otro lado, el resto de finalidades contempladas en dicho Anexo.

Es turno de analizar esas diferencias que mostramos en la siguiente tabla.

Formas de evaluación de la conformidad.	Finalidades de alto riesgo (Anexo III)
Autoevaluación (Anexo VI) o presencia de organismo notificado (Anexo VII)	Identificación biométrica.
Autoevaluación (Anexo VI)	Infraestructuras. Educación y formación profesional. Empleo, gestión de trabajadores y acceso al autoempleo. Acceso y disfrute de los servicios privados esenciales y de los servicios prestaciones públicas esenciales. Aplicación de la ley por parte de las autoridades policiales. Gestión de la Migración, asilo y control de fronteras. Administración de Justicia y procesos democráticos.

A) Sistemas de Inteligencia Artificial cuya finalidad de alto riesgo sea la identificación biométrica

De acuerdo al artículo 43.1 del RIA, el proceso de evaluación de la conformidad de los sistemas de identificación biométrica podrá ser, o bien el de la autoevaluación, o bien el de la presencia del organismo notificado. Ello dependerá de si el proveedor ha aplicado o no normas armonizadas o especificaciones comunes en su sistema de IA para cumplir con las exigencias del RIA.

Por tanto, cuando el proveedor haya aplicado normas armonizadas o especificaciones comunes, éste optará entre realizar la autoevaluación de la conformidad (Anexo VI) o solicitar que ésta la realice un organismo notificado (Anexo VII).

Ahora bien, si el proveedor no utiliza normas armonizadas o estas las aplica de forma parcial o en su caso no dispone de especificaciones comunes para cumplir con los requisitos del Reglamento, éste necesariamente deberá acudir al proceso de evaluación de la conformidad con presencia de organismo notificado.

B) Sistemas de Inteligencia Artificial cuya finalidad de alto riesgo sea distinta a la identificación biométrica

Para el resto de finalidades consideradas de alto riesgo distintas a la identificación biométrica los proveedores realizarán la autoevaluación de conformidad contemplada en el Anexo VI que previamente ya hemos explicado²⁷.

A pesar de que en estos casos el proceso de verificación se lleva a cabo en su totalidad por parte de los proveedores, dicho proceso deberá quedar totalmente documentado y siempre a disposición de la autoridad de vigilancia del mercado. En este sentido, los proveedores están obligados a demostrar, previa solicitud motivada de la autoridad competente, la conformidad de su sistema de IA con los requisitos esenciales del RIA²⁸.

2. SISTEMAS DE INTELIGENCIA ARTIFICIAL DE PRODUCTOS O COMPONENTES DE SEGURIDAD DE PRODUCTOS CONSIDERADOS DE ALTO RIESGO

El Anexo I hace referencia a una serie de productos y componentes de seguridad de productos cuyo diseño, puesta en el mercado y vigilancia de éstos están regulados por leyes de armonización que siguen el NML. Como ya hemos señalado anteriormente, cada una de estas leyes presenta una estructura similar. Dentro de esos elementos comunes todos estos productos han de superar la evaluación de la conformidad que en su caso esté contemplada en dichos textos legales. Cada legislación contempla varios procesos de evaluación de la conformidad en función de las características de cada uno de estos productos.

Cuando un sistema de IA sea producto en sí mismo o componente de seguridad de uno de estos productos, la evaluación de la conformidad de estos sistemas de IA estará conformada por dos grandes procesos.

Por un lado, el proceso de evaluación de la conformidad de ese sistema de IA será el que se contemple para ese producto o componente de seguridad del producto en la legislación aplicable. Por otro lado, dentro de ese proceso de evaluación de la conformidad se deberán incorporar algunas de las actuaciones que ya hemos comentado relacionadas con la documentación técnica que han de revisar los organismos notificados y que se derivan directamente del propio RIA.

A) La evaluación de la conformidad de los sistemas de inteligencia artificial en la estructura de la legislación de armonización

La verificación de los requisitos del RIA se enmarcará dentro del proceso de evaluación de la conformidad previsto en cada una de las leyes de armonización de esos productos o componentes de seguridad de productos. El objetivo es que si una organización pretende desarrollar un juguete, un producto sanitario o un ascensor que lleva integrado un sistema de IA, ésta no tenga que realizar dos evaluaciones de conformidad, sino únicamente una, esto es, la que se indica en la legislación de armonización aplicable a esos productos²⁹.

27. Artículo 43.2 del Reglamento Europeo de Inteligencia Artificial.

28. Artículo 16.k) del Reglamento Europeo de Inteligencia Artificial.

29. Considerando 124 del Reglamento de la IA.

El proceso de verificación previsto en cada acto de legislación de armonización depende en la mayoría de los casos del producto. Es decir, dentro de un mismo acto de legislación se contemplan diferentes procesos de evaluación de la conformidad³⁰.

Procesos de evaluación de la conformidad contemplados en los diferentes actos de armonización de la UE	
Módulo A	Control interno de la producción. (autoevaluación) Control interno de la producción más ensayo supervisado de los productos Control interno de fabricación más control supervisado de los productos a intervalos aleatorios
Módulo B	Examen CE de tipo
Módulo C	Conformidad con el tipo basada en el control interno de la producción
Módulo C1	Conformidad con el tipo basada en el control interno de la producción más ensayo supervisado de los productos
Módulo C2	Conformidad con el tipo basada en el control interno de la producción más control supervisado de los productos a intervalos aleatorios
Módulo D	Conformidad con el tipo basada en el aseguramiento de la calidad del proceso de producción
Módulo D1	Aseguramiento de la calidad del proceso de producción
Módulo E	Conformidad con el tipo basada en el aseguramiento de la calidad del producto
Módulo E1	Aseguramiento de la calidad de la inspección y el ensayo del producto acabado
Módulo F	Conformidad con el tipo basada en la verificación del producto
Módulo F1	Conformidad basada en la verificación de los productos
Módulo G	Conformidad basada en la verificación por unidad
Módulo H	Conformidad basada en el pleno aseguramiento de la calidad
Módulo H1	Conformidad basada en el pleno aseguramiento de la calidad más el examen del diseño

La forma de integrar la verificación de los requisitos previstos en el RIA a la metodología de evaluación de la conformidad diseñada por cada legislación de armonización de los diferentes productos no se explicita en esta norma a lo largo de su articulado.

30. Anexo II. Decisión n.º 768/2008/CE del Parlamento Europeo y del Consejo, de 9 de julio de 2008, sobre un marco común para la comercialización de los productos y por la que se deroga la Decisión 93/465/CEE del Consejo.

No obstante, el Considerando 64 aboga por una aplicación simultánea y complementaria de los diversos actos legislativos que pueden resultar aplicables en estos casos. El objetivo no es otro que el de evitar cargas o costes innecesarios. La verificación de estos requisitos deberá responder a la misma filosofía.

Es muy probable que la integración de estos requisitos acabe desplegándose a través del desarrollo de normas armonizadas o especificaciones comunes. También es posible que las diferentes leyes de armonización que en su caso se vayan modificando o actualizando comiencen a contemplar la integración de los requisitos del RIA en esas legislaciones. En este sentido, el Reglamento 2023/1230 relativo a las máquinas ya ha realizado en parte ese proceso de entendimiento entre los diferentes textos normativos³¹.

A falta de la redacción de esas normas armonizadas u otros instrumentos que ayuden o faciliten en parte la integración de los requisitos exigidos por el RIA en el proceso de evaluación de los productos o componentes de seguridad de productos que ya contemplan un proceso de evaluación de la conformidad, la conclusión a la que podemos llegar es que la integración de los requisitos del RIA no podrá afectar a la lógica, la metodología o a la estructura indicada en dichas leyes de armonización.

B) Las obligaciones de verificación derivadas del Reglamento en la evaluación de la conformidad de productos o componentes de seguridad de productos

Si bien el proceso de evaluación de la conformidad a seguir será el marcado por la legislación de armonización, a efectos de asegurar que el sistema de IA cumple con los requisitos del RIA, éste habilita a los organismos notificados conforme a la legislación de armonización del producto en cuestión a llevar a cabo una serie de actuaciones que se derivan del proceso de evaluación de la conformidad previsto en el RIA.

Concretamente, el organismo notificado que pretenda llevar a cabo la evaluación de un sistema de IA que es en sí mismo un producto o un componente de seguridad llevará a cabo las siguientes actuaciones: accederá a la documentación técnica del sistema de IA, podrá acceder al conjunto de datos de entrenamiento, validación y prueba utilizados y podrá acceder al modelo de entrenamiento y al modelo entrenado del sistema de IA. Además, el organismo notificado denegará el certificado UE de la evaluación de la documentación técnica cuando entienda que el sistema no cumple los requisitos relativos a los datos utilizados para el entrenamiento³².

Con esta última apreciación se busca garantizar que los organismos notificados puedan llevar a cabo un procedimiento de evaluación de conformidad adecuado que si bien ha de respetar la estructura de la legislación de armonización en cuestión, resulte adecuado y en sintonía con las exigencias del RIA.

31. Considerando 54, artículo 25.2 y Anexo I. Parte A del Reglamento (UE) 2023/1230 del Parlamento Europeo y del Consejo, de 14 de junio de 2023, relativo a las máquinas, y por el que se derogan la Directiva 2006/42/CE del Parlamento Europeo y del Consejo y la Directiva 73/361/CEE del Consejo.

32. Artículo 43.3.

Formas de evaluación de la conformidad	Evaluación de la conformidad en función del tipo de sistema de IA (Artículo 43)
Autoevaluación	Finalidades de alto riesgo. Anexo III. Apartados 2 a 8. (todas salvo identificación biométrica)
Autoevaluación o Presencia de organismo notificado	Finalidades de alto riesgo. Anexo III. Apartado 1. (solo identificación biométrica)
Evaluación de la conformidad según legislación de armonización del producto	Productos de Alto Riesgo. Anexo I

C) Posibilidad de prescindir de la presencia del organismo notificado

Cuando un sistema de IA sea un producto o componente de seguridad de un producto de los actos legislativos de armonización, el fabricante podrá prescindir de la evaluación de la conformidad con presencia de organismo notificado siempre que se den dos condiciones acumulativas.

En primer lugar, que el acto de legislación de armonización prevea la posibilidad de que el fabricante pueda prescindir del proceso de verificación realizado por tercero si éste ha aplicado normas armonizadas que cubren los requisitos exigidos por esa legislación.

En segundo lugar, que el fabricante haya aplicado normas armonizadas o especificaciones comunes que cubran los requisitos exigidos a los sistemas de IA de alto riesgo.

En definitiva, se podrá prescindir de la presencia de organismo notificado cuando así lo contemple el acto legislativo al que se somete el producto y se hayan aplicado normas armonizadas o especificaciones comunes que cubran todos los requisitos exigidos tanto por el acto legislativo al que se somete el producto como los requisitos exigidos al sistema de IA conforme al RIA.

Actualmente no existen normas armonizadas o especificaciones comunes que cubran los requisitos establecidos en el RIA para los sistemas de IA, de manera que esta opción de eludir la evaluación de la conformidad de terceros indicada no es posible aún.

Para estos casos, entendemos que, aunque el fabricante pudiera evitar la evaluación de la conformidad de su producto por parte de un organismo notificado en atención a la legislación de armonización por aplicar normas armonizadas que cubren los requisitos de esa norma, dicho organismo notificado habría de intervenir si no hay normas armonizadas o especificaciones comunes que cubran los requisitos del RIA.

Por ejemplo, el artículo 19.2 de la Directiva 2009/48/CE sobre la seguridad de los juguetes, contempla la posibilidad de que el fabricante prescinda de la evaluación de la conformidad realizada por organismos notificados si éste ha aplicado normas

armonizadas que abarquen todos los requisitos pertinentes previstos en dicha Directiva.

Por tanto, si un sistema de IA es un juguete o un componente de seguridad de un juguete, dado que por ahora no existen normas armonizadas ni especificaciones comunes que cubran los requisitos que establece el RIA, el fabricante de ese juguete no podrá eludir el organismo notificado a pesar de que el acto legislativo de armonización de juguetes lo permita.

Producto/Sistema de IA	<i>¿Ha aplicado normas armonizadas que cubre la legislación?</i>	<i>¿Es necesario organismo notificado?</i>
Juguete	No cubre los requisitos de la Directiva de juguetes	Sí
Juguete	Sí cubre los requisitos de la Directiva de juguetes	No
Juguete que a su vez es un sistema de IA	Sí cubre los requisitos de la Directiva de juguetes	Sí
	No cubre requisitos del RIA	
Juguete que a su vez es un sistema de IA	Sí cubre los requisitos de la Directiva de juguetes	No
	Sí cubre los requisitos del RIA	

VI. LOS ORGANISMOS NOTIFICADOS ENCARGADOS DE REALIZAR LA EVALUACIÓN DE LA CONFORMIDAD DE LOS DIFERENTES SISTEMAS DE INTELIGENCIA ARTIFICIAL

Como regla general, los proveedores de sistemas de IA que hayan de someterse a un proceso de evaluación de la conformidad con presencia de organismo notificado podrán elegir aquel que estimen adecuado, siempre claro está, que éste haya sido acreditado para poder realizar la evaluación de la conformidad de sistemas de IA.

Ahora bien, se contemplan algunas reglas específicas donde los proveedores ven limitado su ámbito de libertad a la hora de elegir uno u otro organismo notificado.

En primer lugar, conforme al último párrafo del artículo 43.1 del RIA, el sistema de IA que tenga como finalidad la identificación biométrica y cuyo uso se prevea que sea realizado por parte de autoridades en aplicación de la ley o autoridades de inmigración, el organismo notificado será³³: a) o bien la autoridad de protección de datos competente con arreglo al Reglamento (UE) 2016/679 o a la Directiva (UE) 2016/680 (actualmente es la AEPD)³⁴, b) o bien cualquier otra autoridad designada con arreglo a las mismas condiciones establecidas en los artículos 41 a 44 de la

33. Artículo 74.8 del Reglamento de Inteligencia Artificial.

34. Tanto para el caso de las actividades de policía como en su caso las de inmigración y control de fronteras.

Directiva (UE) 2016/680³⁵. Entendemos por tanto que, el organismo notificado para estos sistemas de IA será la AEPD u organismo análogo en materia de protección de datos como ocurre con el CGPJ en materia de justicia³⁶, ya que muy probablemente no se creó en España una autoridad con el grado de independencia y con las mismas facultades de supervisión y acceso a los datos que cuenta la actual AEPD en estos contextos.

No pensamos que la Agencia Española de Supervisión de Inteligencia Artificial (AESIA) cuente actualmente con el nivel de independencia exigido en los artículos 41 a 44 de la Directiva 2016/680 y por el que aboga el RIA para estos supuestos. En este sentido, a diferencia de la AEPD que se ha configurado como una autoridad administrativa independiente en nuestro derecho interno³⁷, la AESIA está constituida como una agencia estatal que tiene autonomía e independencia técnica³⁸, sin embargo, de sus estatutos se puede atisbar que el Gobierno central tienen mucha capacidad de control, sobre todo a la hora de elegir a los órganos esenciales de esta agencia³⁹.

En segundo lugar, también conforme al último párrafo del artículo 43.1 del RIA, los proveedores cuyo sistema de IA tenga como finalidad la identificación biométrica y tal sistema se prevea su puesta en servicio por parte de instituciones o agencias de la UE, el organismo notificado será el Supervisor Europeo de Protección de Datos⁴⁰.

En tercer lugar, el artículo 41.3 establece que los organismo notificados que son competentes para realizar la evaluación de la conformidad de los productos sometidos a los actos legislativos de armonización dispondrán de la facultad para verificar la conformidad de los requisitos del RIA cuando dicho producto sea un sistema de IA. Estos organismos notificados no sólo deberán cumplir con los requisitos previstos en los actos legislativos de armonización que le habilitan para poder realizar las evaluaciones de conformidad, sino que además habrán de implementar otra serie de exigencias que se derivan del Reglamento Europeo IA para poder ser evaluadores de los requisitos previstos en esta norma. Entre estas exigencias encontramos la

35. Estos artículos hacen referencia a la necesidad de que esa autoridad pública en el marco de sus actividades goce de un alto grado de independencia. Artículo 42. Directiva (UE) 2016/680 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por parte de las autoridades competentes para fines de prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de sanciones penales, y a la libre circulación de dichos datos y por la que se deroga la Decisión Marco 2008/977/JAI del Consejo.

36. La autoridad de protección de datos en el ámbito de la Administración de Justicia es el Consejo General del Poder Judicial. Artículo 236. Nonies. Ley Orgánica 6/1985, de 1 de julio, del Poder Judicial.

37. Artículo 109 de la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público.

38. Artículo 108 bis de la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público.

39. Tal y como puede ser el presidente de la AESIA o la dirección de ésta. Véase el Real Decreto 729/2023, de 22 de agosto, por el que se aprueba el Estatuto de la Agencia Española de Supervisión de Inteligencia Artificial.

40. Artículo 74.9 del RIA.

necesidad de que estos organismos ostente el personal y medios suficientemente competentes para poder verificar adecuadamente los requisitos del RIA⁴¹.

A priori, resulta lógico pensar que el organismo notificado que realiza las evaluaciones de conformidad de un ascensor o un juguete conforme a su normativa sea el más capacitado para llevar a cabo las evaluaciones de la conformidad de ese mismo ascensor o juguete cuando éste lleve integrado un sistema de IA como componente de seguridad. Ello es así porque estos organismos notificados tienen una amplia experiencia en verificar los requisitos de los ascensores, sin embargo, será necesario que dichos organismos implementen los medios humanos y técnicos para que efectivamente también pueda verificar los requisitos derivados del RIA.

El RIA habilita a estos organismos a que puedan subcontratar con terceras entidades las labores de verificación de los requisitos del RIA⁴². Es probable por tanto que los organismos notificados de productos o componentes de seguridad de productos sometidos a legislación de armonización que pretendan extender sus procesos de verificación a sistemas de IA que sean su vez productos o componentes de seguridad de productos sobre los que vienen realizando evaluaciones de conformidad acaben subcontratando esas actividades a entidades especializadas en la verificación de los requisitos exigidos por el RIA.

Sistema de IA	Organismo notificado
Identificación biométrica utilizada por parte de autoridades de inmigración o aplicación de la ley	Muy posiblemente la AEPD
Identificación biométrica utilizada por parte de autoridades e instituciones europeas	Supervisor Europeo de Protección de Datos
Producto o componente de seguridad de producto sometido a legislación de armonización	Organismo notificado del producto o componente de seguridad del producto.

VII. SUPUESTOS DONDE NO ES NECESARIO REALIZAR LA EVALUACIÓN DE LA CONFORMIDAD O EXISTEN PRESUNCIONES DE CONFORMIDAD DE SU CUMPLIMIENTO

El RIA contempla una serie de situaciones donde no es necesario que un sistema de IA deba pasar por un proceso de verificación de la conformidad previo a su puesta en el mercado. Estos supuestos son excepcionales y obedecen a razones justificadas. A su vez, también regula algunos supuestos donde se presume la conformidad del sistema de IA con los requisitos exigidos por el RIA.

1. AUTORIZACIÓN PREVIA DE PUESTA EN EL MERCADO DE SISTEMAS DE INTELIGENCIA ARTIFICIAL

De acuerdo al artículo 46.1 del RIA, cualquier autoridad de vigilancia del mercado podrá autorizar la introducción en el mercado de un sistema de IA sin que tal sistema haya pasado un proceso de evaluación de la conformidad.

41. Apartados 4,9 y 10 del artículo 31 del Reglamento de IA.

42. Artículo 33 del Reglamento de IA.

Esta situación se permitirá cuando el uso de este sistema de IA sea necesario para proteger la vida o la salud de las personas, el medio ambiente, la seguridad pública o activos fundamentales de la industria y de las infraestructuras. Como se puede comprobar, las razones por las que se puede conceder este tipo de autorizaciones abarcan un gran número supuestos. Ahora bien, el RIA contempla ciertas previsiones que tratan de garantizar la excepcionalidad de este tipo de medidas.

En primer lugar, la autoridad de vigilancia del mercado deberá asegurarse que el sistema de IA cumple con los requisitos mínimos esenciales exigidos a los sistemas de IA⁴³ antes de conceder la autorización. La autorización deberá estar debidamente motivada tomando en consideración las razones por las que se ha permitido ésta.

En segundo lugar, una vez concedida la autorización, la autoridad de vigilancia del mercado la comunicará a la Comisión Europea y al resto de Estados Miembros. En un plazo de 15 días estos últimos podrán plantear alguna objeción a la autorización concedida por entender que no está suficientemente justificada o es contraria al derecho de la Unión Europea. Si no se presentan objeciones se entenderá que la autorización está justificada, en caso contrario, la Comisión iniciará los trámites oportunos para comunicarse con la autoridad de vigilancia que concedió la autorización y los operadores del sistema para que puedan expresar su opinión. La Comisión finalmente decidirá si la autorización está o no justificada.

La participación de la Comisión o de los Estados Miembros durante este proceso obedece al hecho de que cualquier autoridad de vigilancia de cualquier Estado Miembro puede autorizar que un sistema de IA se utilice en cualquier parte de la Unión Europea. Pensamos por tanto que se trata de una medida de control por parte de los Estados Miembros y de la Comisión Europea hacia las diferentes autoridades de vigilancia del mercado que pueden en su caso otorgar con cierta facilidad o poco rigor este tipo de autorizaciones excepcionales.

2. LA PUESTA EN EL MERCADO DEL SISTEMA DE INTELIGENCIA ARTIFICIAL SIN AUTORIZACIÓN PREVIA

En situaciones de urgencia debidamente justificadas por razones excepcionales de seguridad pública o en caso de amenaza específica, sustancial e inminente para la vida o la seguridad física de las personas físicas, las autoridades encargadas del orden público o las autoridades de protección civil podrán poner en servicio un sistema de IA de alto riesgo sin necesidad de obtener la autorización previamente mencionada.

En estos supuestos las autoridades que utilicen estos sistemas deberán necesariamente solicitar sin demora indebida la autorización previamente explicada en el apartado anterior a la autoridad de vigilancia del mercado. Si la autoridad de vigilancia del mercado deniega la autorización, el uso del sistema deberá suspenderse y los resultados e informaciones de salida derivados de dicho uso deberán descartarse.

43. El artículo 46.3 señala que solo se expedirá la autorización «*si la autoridad de vigilancia del mercado llega a la conclusión de que el sistema de IA de alto riesgo cumple los requisitos establecidos en la sección 2*». Hemos de entender que se refiere a la Sección 2 del Capítulo III del Reglamento de IA, sección que abarca los requisitos esenciales, es decir, artículos 8 a 15.

3. LAS EXENCIONES DE EVALUACIÓN DE LA CONFORMIDAD DE LOS PRODUCTOS SOMETIDOS A LEGISLACIÓN DE ARMONIZACIÓN

Cuando un sistema de IA sea un producto o componente de seguridad de un producto sometido a legislación de armonización, los procedimientos de excepción de la evaluación de la conformidad solo se permitirán si así lo contempla la legislación de armonización aplicable.

4. LAS PRESUNCIONES DE CONFORMIDAD

El RIA establece varias presunciones de conformidad que hacen referencia a varios requisitos exigibles a los sistemas de IA. De esta manera, si un proveedor aplica ese requisito conforme a lo establecido por dicha presunción, hay que entender que cumple con el requisito exigido por el RIA.

Son dos las presunciones que se contemplan en el artículo 42, por un lado, cuando los sistemas de IA hayan sido entrenados y probados con datos que reflejen el entorno geográfico, conductual o funcional específico de su uso, se presumirá que dicho sistema es conforme al requisito previsto en el artículo 10.4 del RIA. Entendemos que la presunción de conformidad prevista en este precepto tiene los mismos efectos que la presunción de conformidad de las normas armonizadas o especificaciones comunes, es decir, el uso de éstas no supone la elusión de la evaluación de la conformidad de los requisitos cubiertos por esa presunción pero sí asegura un proceso de verificación mucho más rápido que en los supuestos en los que dichos requisitos se tuvieran que demostrar con la aplicación de otras técnicas o estándares.

Por otro lado, cuando un sistema de IA haya sido certificado o se le haya expedido una declaración de conformidad con un esquema de ciberseguridad y cuyas referencias se hayan publicado en el DOUE⁴⁴, se presumirá la conformidad con los requisitos previstos en el artículo 15 del RIA sobre ciberseguridad en la medida en que dicho certificado de ciberseguridad o declaración de conformidad se prevean tales requisitos.

Finalmente, y aunque no se trata de una presunción de conformidad en sentido estricto, el RIA en su artículo 57.7 contempla la posibilidad de que las autoridades de vigilancia del mercado o los organismos notificados tengan en cuenta positivamente a efectos de acelerar el procedimiento de evaluación de la conformidad de un sistema de IA los informes que se hayan emitido sobre ese sistema por su participación en un espacio controlado de pruebas.

VIII. REFLEXIONES SOBRE LA REGULACIÓN DE LA EVALUACIÓN DE LA CONFORMIDAD ESTABLECIDA EN EL REGLAMENTO

Como se ha señalado a lo largo de este capítulo, el proceso de evaluación de la conformidad forma parte de una batería de medidas e instrumentos que tienen como

44. Reglamento (UE) 2019/881 del Parlamento Europeo y del Consejo, de 17 de abril de 2019, relativo a ENISA (Agencia de la Unión Europea para la Ciberseguridad) y a la certificación de la ciberseguridad de las tecnologías de la información y la comunicación y por el que se deroga el Reglamento (UE) n.º 526/2013 (Reglamento sobre la Ciberseguridad).

objetivo asegurar que un producto que se pone en el mercado cumple con unos mínimos de seguridad y de garantía para las personas.

El proceso de verificación de los requisitos exigidos a los sistemas de IA se muestra por tanto como una fase elemental para asegurar que dichos productos son adecuados para poder utilizarse.

El legislador europeo ha considerado que para un número relevante de sistemas de IA de alto riesgo esa verificación de los requisitos la lleve a cabo el propio proveedor⁴⁵. El ámbito laboral, educativo, gran parte del sector público y actuaciones policiales y judiciales salvo identificación biométrica quedan al amparo del control interno de los proveedores.

Resulta cuanto menos sorprendente esta decisión legislativa. A pesar de los esfuerzos por diseñar un marco legal claramente innovador y respetuoso con los derechos fundamentales presentes durante el ciclo de vida de los sistemas de IA a través de la obligación de imponer requisitos que reducen los principales riesgos asociados a dicho sistemas, la autoevaluación por parte de los proveedores puede generar importantes sospechas en cuanto a la integración efectiva de esos requisitos en los sistemas de IA que operen en la Unión Europea.

El legislador europeo confía por tanto en la industria en estos primeros años para que sea ésta la que en su caso se auto verifique a la espera de que se vayan desarrollando productos cada vez más maduros y acordes con las exigencias legislativas.

Es cierto no obstante que, pese a la autoevaluación como regla general, el RIA otorga gran peso a los Estados Miembros y a la Comisión Europea para asegurar un cumplimiento efectivo de esta novedosa norma.

Así, en primer lugar, será esencial el papel más o menos protagonista que los Estados Miembros concedan a las autoridades de vigilancia del mercado. Estas autoridades públicas deberán contar con los medios humanos y técnicos suficientes para realizar una supervisión efectiva de los sistemas de IA que se pongan en el mercado, tanto si han pasado un proceso de autoevaluación como si han participado organismos notificados. La independencia de estas autoridades respecto de los respectivos gobiernos como de las entidades del sector privado deberá ser adecuada, no olvidemos que estas autoridades supervisarán sistemas tanto del sector público como del privado.

En segundo lugar, el propio RIA otorga a la Comisión Europea la posibilidad de que a través de actos delegados pueda cambiar la autoevaluación exigible actualmente a los sistemas de IA de alto riesgo del Anexo III por la evaluación por parte de un organismo notificado⁴⁶.

IX. CONCLUSIONES

1. Cualquier sistema de IA considerado de alto riesgo que pretenda introducirse en el mercado de la Unión Europea deberá pasar un proceso de verificación previo del cumplimiento e integración de los requisitos esenciales exigidos por el RIA.

45. Véase el Anexo III salvo el apartado 1 respecto a la identificación biométrica.

46. Artículo 43.6 del Reglamento Europeo de Inteligencia Artificial.

2. El RIA contempla dos procesos de evaluación de la conformidad, por un lado, la autoevaluación, la cual será realizada por el propio proveedor del sistema de IA y, por otro lado, la evaluación con presencia de un organismo notificado, ese decir, una entidad tercera que ha sido acreditada para realizar evaluaciones de la conformidad de sistemas de IA conforme a este reglamento europeo.

3. La elección de un proceso de evaluación de la conformidad u otro no queda en manos del proveedor sino que dependerá del tipo de sistema de IA que se haya desarrollado.

4. Cuando el sistema de IA tenga como finalidad la identificación biométrica, el proveedor podrá optar por uno u otro proceso de evaluación de la conformidad siempre que haya aplicado normas armonizadas o especificaciones comunes para implementar los requisitos esenciales exigidos por el RIA. En caso contrario, el proveedor deberá acudir al proceso de evaluación de la conformidad con presencia de organismo notificado.

5. Cuando el sistema de IA tenga como finalidad algunos de los fines considerados de alto riesgo por el RIA (salvo identificación biométrica), será el propio proveedor el que llevará a cabo la evaluación de la conformidad del sistema de IA (la autoevaluación).

6. Cuando el sistema de IA sea un producto o un componente de seguridad de un producto que está regulado por la normativa europea, el proceso de evaluación de la conformidad será aquél que esté previsto en la norma que regula el producto. La incorporación de los requisitos esenciales del RIA no alterará la estructura del proceso de evaluación previsto en la norma que regula el producto, si bien, en ese proceso se deberán verificar los requisitos exigidos por el RIA.

7. Como regla general, los proveedores tienen la libertad para elegir el organismo notificado que evaluará sus sistemas de IA, sin embargo, en determinadas ocasiones donde entra en juego el sector público, no tendrán margen de elección.

8. Como regla general, todo sistema de IA que se pretenda poner en el mercado de la UE deberá haber pasado un proceso de evaluación de la conformidad, no obstante, existen supuestos excepcionales donde no se requerirá tal proceso de verificación inicial.

Régimen general de obligaciones de proveedores y responsables del despliegue en el Reglamento de inteligencia artificial

ADRIÁN PALMA ORTIGOSA

Profesor Ayudante Doctor del Departamento de Derecho Administrativo de la Universitat de València¹

I. INTRODUCCIÓN

El presente trabajo estudia diferentes preceptos del RIA. En primer lugar, se profundiza en los diferentes operadores que intervienen a lo largo de la cadena de valor de la IA, se define cada uno de estos agentes y se analizan sus principales funciones y obligaciones. En segundo lugar, se estudia el papel y funciones asignadas a unas de las autoridades competentes del cumplimiento del RIA, esta es, la autoridad notificante. En tercer lugar, se analizan las medidas que contempla esta norma europea en favor de las pymes y las empresas emergentes con el objetivo de facilitar una correcta adaptación de estas últimas a dicha norma. En cuarto lugar, se estudia el procedimiento sobre notificación de incidentes graves de los sistemas de IA regulado en el RIA².

1. Este trabajo se ha llevado a cabo en el marco de los siguientes proyectos de investigación:
«Algorithmical Law» (PROMETEO/2021/009. Financiado por la Generalitat Valenciana. «La regulación de la economía digital: tutela publica de la igualdad y herramientas algorítmicas» (PID2019-108745GB-I00). Ministerio de Ciencia e Innovación. «Derechos y garantías públicas frente a las decisiones automatizadas y el sesgo y discriminación algorítmicas» 2023-2025 (PID2022-136439OB-I00) financiado por el Ministerio de Ciencia e Innovación. «Herramientas algorítmicas para ciudadanos y Administraciones Públicas» (Proyectos de Generación de Conocimiento, Ministerio de ciencia e Innovación, convocatoria 2021, PID2021-126881OB-I00). «Algorithmic Decisions and the Law: Opening the Black Box» (TED2021-131472A-I00) del Plan de Recuperación, Transformación y Resiliencia. Convenio de Derechos Digitales-SEDIA Ámbito 5 y 6 (2023/C046/00228673).
2. Este trabajo se ha llevado a cabo en el marco de los siguientes proyectos de investigación: «Algorithmical Law» (PROMETEO/2021/009. Financiado por la Generalitat Valenciana. «La regulación de la economía digital: tutela publica de la igualdad y herramientas algorítmicas» (PID2019-108745GB-I00). Ministerio de Ciencia e Innovación. «Derechos y garantías públicas frente a las decisiones automatizadas y el sesgo y discriminación algorítmicas» 2023-2025 (PID2022-136439OB-I00) financiado por el

II. EVOLUCIÓN, TRAMITACIÓN Y CONTENIDO FINAL DE LOS ARTÍCULOS DEL REGLAMENTO IMPLICADOS

Algunos de los preceptos que se analizan en esta parte de la obra colectiva han sufrido importantes cambios a lo largo del proceso legislativo del RIA. Entre los principales cambios podemos destacar los siguientes.

En primer lugar, por lo que se refiere al contenido, se han aumentado de forma considerable las obligaciones exigidas a los diferentes operadores presentes durante la cadena de valor de los sistemas de IA. Destaca sobre todo el aumento de exigencias al responsable del despliegue. Este agente inicialmente se denominaba usuario, si bien, en las últimas versiones conocidas del Reglamento se sustituyó tal denominación. El cambio resulta en nuestra opinión acertado, ya que era frecuente que el término de usuario se confundiera con las personas afectadas por las decisiones. Respecto de las obligaciones, cabe destacar la obligación de realizar una evaluación de impacto en materia de derechos fundamentales, esta exigencia se implementó en los últimos acuerdos del Reglamento, pero su origen deviene de las enmiendas aportadas por el Parlamento Europeo³.

En Segundo lugar, en cuanto a la forma, algunos de estos preceptos han cambiado de lugar o se han refundido, si bien, el lugar donde estaban encuadrados en la primera versión del RIA no se ha visto altamente afectado.

III. LOS OPERADORES PRESENTES DURANTE LA CADENA DE VALOR DE LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL EN EL REGLAMENTO.

A lo largo de la compleja cadena de valor de los sistemas de IA pueden estar presentes todo tipo de agentes que participan de una manera u otra. El RIA ha tratado de poner nombre y apellidos a todos estos agentes a los que denomina operadores. Por lo que se refiere al nombre, establece una definición de cada uno de ellos, en cuanto a los apellidos, a cada uno de estos operadores le asigna una serie de funciones y obligaciones.

El objetivo del Reglamento es doble, por un lado se asignan responsabilidades claras, de manera que cada agente conozca qué tiene o no tiene que hacer en virtud del Reglamento, y por otro lado, se reduce o mitiga la posibilidad de eludir potenciales responsabilidades cuando el sistema cometa algún daño a terceros.

Es turno de analizar cada uno de estos operadores y sus funciones⁴.

1. EL PROVEEDOR Y SUS OBLIGACIONES

El proveedor es toda persona física, jurídica o autoridad pública que desarrolle un sistema de IA o un modelo de IA de uso general o para el que se desarrolle un sistema

Ministerio de Ciencia e Innovación. «Herramientas algorítmicas para ciudadanos y Administraciones Públicas» (Proyectos de Generación de Conocimiento, Ministerio de ciencia e Innovación, convocatoria 2021, PID2021-126881OB-I00).

3. Véase la enmienda 413 de la resolución del Parlamento Europeo el 14 de junio de 2023 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión.

4. Estos son: proveedor, fabricante del producto, responsable del despliegue, representante autorizado, importador o distribuidor. Artículo 3.8. Reglamento de IA.

de IA o un modelo de IA de uso general y lo introduzca en el mercado o ponga en servicio el sistema de IA con su propio nombre o marca comercial⁵.

El proveedor es el eje central sobre el que descansan el mayor número de obligaciones y responsabilidades en el cumplimiento del RIA. Ello obedece al hecho de que es el sujeto encargado de desarrollar o mandar desarrollar el sistema de IA, el cual, posteriormente se utilizará en la toma de decisiones.

Esas obligaciones aparecen mencionadas a lo largo de todo el texto normativo, si bien, aparecen concentradas en los preceptos iniciales del reglamento.

Principales obligaciones	Artículo
Requisitos esenciales de los sistemas de IA	8 a 15
Obligaciones esenciales: documentación técnica, evaluación de la conformidad, marcado CE, etc.	16
Sistema de la gestión de la calidad	17
Conservación de la documentación	18
Archivos de registros	19
Medidas correctoras y obligaciones de informar	20
Cooperación con las autoridades competentes	21

En este trabajo sólo estudiaremos las obligaciones mencionadas en los artículos 20 y 21, estas son, las medidas correctoras, las obligaciones de información y la cooperación con las autoridades competentes. El resto de las obligaciones se tratan en otros apartados de esta obra colectiva.

En primer lugar, por lo que se refiere a la aplicación de medidas correctoras, el proveedor que considere o tenga motivos para considerar que un sistema de IA que ha introducido no es conforme al RIA, adoptará las medidas adecuadas para desactivarlo o recuperarlo. Además, ha de informar de dicha situación a los distribuidores del sistema y en su caso, a los responsables del despliegue de éste, así como a los representantes autorizados y a los importadores.

En segundo lugar, si el proveedor considera que un sistema de IA presenta un riesgo tras una notificación del responsable del despliegue de tal sistema, el proveedor deberá investigar las causas de dicho riesgo e informará a las autoridades de vigilancia del mercado competentes y, cuando proceda, al organismo notificado que haya emitido el correspondiente certificado de evaluación de la conformidad.

En tercer lugar, los proveedores están obligados a facilitar toda la información y documentación que les solicite la autoridad competente para demostrar el cumplimiento del RIA. Siempre que haya una petición motivada por parte de la autoridad competente, el proveedor también le dará acceso al archivo de registros generados automáticamente del sistema en la medida que dichos registros estén bajo su control. Estos deberes de información y cooperación resultan esenciales, ya que en muchos casos, los sistemas de IA que se introducen en el mercado no pasan por

5. Artículo 3.3. Reglamento de IA.

ningún tipo de control más a allá de la propia autoevaluación interna que realice el propio proveedor del sistema, de ahí que el papel de las autoridades de vigilancia del mercado una vez que el sistema esté adoptando decisiones sea muy relevante.

2. EL RESPONSABLE DEL DESPLIEGUE Y SUS OBLIGACIONES

El responsable del despliegue es toda persona física, jurídica o autoridad pública que utilice un sistema de IA bajo su propia autoridad, salvo cuando su uso se enmarque en una actividad personal de carácter no profesional⁶.

Junto con el proveedor, el responsable del despliegue es el operados al cual el RIA impone más obligaciones de cumplimiento normativo. Ello resulta del todo lógico teniendo en cuenta que será el que utilizará el sistema gran parte del tiempo que esté éste funcionando.

Las obligaciones del responsable del despliegue aparecen sobre todo concentradas en los artículos 26 y 27 del RIA. Con el objetivo de facilitar la lectura de estos preceptos en este trabajo agrupamos el contenido de estas obligaciones en 4 grandes grupos.

Obligación	Artículo y apartado
Cumplimiento de requisitos esenciales	Artículo 26 apartados 1, 2, 3,4 y 6.
Otras obligaciones de cumplimiento	Artículo 26 apartados 5,7,8,9,11 y 12
Obligaciones relacionadas con el uso de sistemas de identificación biométrica	Artículo 26 apartado 10.
La evaluación de impacto relativa a los derechos fundamentales	Artículo 27

2.1. Obligaciones relacionadas con el cumplimiento de los requisitos esenciales del Reglamento

El RIA establece toda una serie de requisitos mínimos esenciales que han de ser integrados por parte del proveedor durante el desarrollo de los sistemas de IA. Estos requisitos mínimos se encuentran reconocidos en los artículos 8 a 15 del RIA. Parte del contenido de estos requisitos está diseñado para que el responsable del despliegue pueda utilizar adecuadamente el sistema de IA que ha diseñado el proveedor.

En primer lugar, el responsable del despliegue deberá adoptar las medidas técnicas y organizativas adecuadas para garantizar que utilizan el sistema conforme a las instrucciones de uso de éste. Las instrucciones de uso son una herramienta habitual prevista en las normas europeas sobre productos que ayudan a reducir la opacidad y favorece la transparencia del funcionamiento esos productos⁷, en nuestro caso,

6. Artículo 3.4. Reglamento de IA.

7. Por instrucción de uso hay que entender «la información facilitada por el proveedor para informar al responsable del despliegue, en particular, de la finalidad prevista y de la correcta utilización de un sistema de IA». Artículo 3.15 del Reglamento de IA.

los sistemas de IA⁸. Si los responsables del despliegue no siguen adecuadamente estas instrucciones es probable que se les atribuyan responsabilidades por los daños generados por el sistema de IA.

En segundo lugar, se establece que los responsables del despliegue encomendarán las tareas de supervisión de los sistemas de IA a personas que tengan la competencia⁹, la formación y las autoridades necesarias. A pesar de que el proveedor deberá haber diseñado el sistema de manera que se facilite una vigilancia efectiva del mismo¹⁰, corresponde al usuario designar al personal competente y capacitado para ello. Estas medidas de supervisión humana se implementarán sin perjuicio de que el responsable del despliegue deba implementar otras en virtud de otras normas del derecho nacional o europeo¹¹. Por ejemplo, el artículo 22 del Reglamento General de Protección de Datos obliga a las entidades que utilicen sistemas de IA en la toma de decisiones totalmente automatizados a implementar cauces de supervisión humana una vez se ha adoptado la decisión por parte del algoritmo¹². En la medida de lo posible, y siempre que resulte compatible, ambas obligaciones se podrían complementar.

En tercer lugar, el responsable del despliegue debe asegurarse que los datos de entrada que utilice el sistema sean pertinentes y suficientemente representativos para la finalidad prevista del sistema. Se matiza que esta obligación entrará en juego en la medida que dicho responsable del despliegue ejerza el control de esos datos. Hay que destacar que esta misma obligación se exige a los proveedores de los sistemas de IA¹³, no obstante, en el momento que el responsable del despliegue pasa a utilizar el sistema de IA y éste es el que introduce los datos, hay que entender que las obligaciones en esa materia pueden alterarse, sobre todo, si el responsable del despliegue tiene el control de esos datos y deja de utilizar datos que presentan las características específicas indicadas para la finalidad del sistema de IA.

En cuarto lugar, el responsable del despliegue deberá conservar los archivos de registro que el sistema genere automáticamente durante al menos 6 meses desde que se van generando siempre que tales archivos estén bajo su control. Corresponde al proveedor diseñar el sistema para que éste genere automáticamente dichos registros¹⁴, el elemento clave será comprobar quién tiene el control sobre los mismos.

2.2. Otras obligaciones derivadas del cumplimiento del Reglamento

Además de las obligaciones específicas referidas a los requisitos esenciales de los sistemas de IA, el RIA exige a los usuarios toda una serie de obligaciones que

-
8. El contenido mínimo de las instrucciones de uso aparece definido en el artículo 13.3 del Reglamento de IA.
 9. Sobre la supervisión humana véase entre otros: Lazcoz Moratinos, G y Obregón Fernández, A., «La supervisión humana de los sistemas de inteligencia artificial de alto riesgo. Aportaciones desde el Derecho Internacional Humanitario y el Derecho de la Unión Europea». *Revista electrónica de estudios internacionales*, n.º 42, 2021.
 10. Véase el artículo 14 del Reglamento de IA. (Vigilancia humana).
 11. Artículo 26.3 del Reglamento de IA.
 12. Palma Ortigosa, A., *Decisiones automatizadas y protección de datos. Especial atención a los sistemas de inteligencia artificial*. Dykinson. Madrid. 2022, p.286 y ss.
 13. Artículo 10.3 y 4 del Reglamento de IA. (Datos).
 14. Artículo 12. Reglamento de IA. (Registros).

tratan de asegurar un adecuado cumplimiento de ese sistema de IA respecto de esta norma europea.

En primer lugar, se contemplan la obligación de vigilar el funcionamiento del sistema de IA basándose en las instrucciones de uso, y cuando proceda, informar al proveedor sobre el sistema de vigilancia pos-comercialización¹⁵. Además, si el responsable del despliegue detecta que el sistema podría presentar un riesgo, deberá informar de ello al proveedor o distribuidor y a la autoridad de vigilancia del mercado pertinente¹⁶. A su vez, si el responsable del despliegue detecta un incidente grave en el sistema de IA, deberá informar de dicho incidente al proveedor, seguidamente al importador o al distribuidor y a la autoridad de vigilancia del mercado competente. Si el responsable del despliegue no logra contactar con el proveedor, el primero deberá realizar todas las comunicaciones y actuaciones que se exigen al segundo cuando sucede un incidente grave previstas en el RIA¹⁷.

En segundo lugar, el responsable del despliegue, antes de utilizar el sistema de IA en el lugar de trabajo, deberá informar de esa puesta en marcha a los representantes de los trabajadores y a los trabajadores expuestos a la utilización del sistema de IA. Esta información se facilitará, cuando proceda, conforme a los cauces previstos en el derecho nacional o europeo que regulan los procesos de información en favor de los trabajadores y sus representantes. En este sentido, existen ya algunas normas que obligan a los empleadores a informar a los representantes de los trabajadores¹⁸ y en algunos casos también a los propios trabajadores sobre el uso de sistemas algorítmicos y las consecuencias de su uso¹⁹. En la medida de lo posible estas obligaciones de información deberán integrarse y complementarse.

En tercer lugar, corresponde a los responsables del despliegue que sean autoridades públicas registrar los sistemas de IA en la base de datos de la UE creada por el RIA²⁰. En el caso de que dicho sistema no esté registrado en esa base de datos deberán informar de tal hecho al proveedor o al distribuidor.

En cuarto lugar, los responsables del despliegue utilizarán la información que el proveedor les haya facilitado sobre el sistema de IA para llevar a cabo la evaluación de impacto regulada en la normativa europea sobre protección de datos de carácter personal²¹. Como es lógico, esta obligación entrará en juego cuando el responsable del despliegue vaya a utilizar datos de carácter personal. En este sentido, gran parte

15. El sistema de vigilancia pos comercialización se encuentra regulado en el artículo 72 del Reglamento de IA.

16. Para más información sobre el riesgo que puede presentar un sistema de IA, véase el Artículo 79 del Reglamento de IA.

17. Para más información sobre esto último, consúltese el aparatado VI de este trabajo y el artículo 73 del Reglamento de IA.

18. Así se contempla en el Artículo 64.4.d). Real Decreto Legislativo 2/2015, de 23 de octubre, por el que se aprueba el texto refundido de la Ley del Estatuto de los Trabajadores.

19. Propuesta de Directiva del Parlamento Europeo y del Consejo relativa a la mejora de las condiciones laborales en el trabajo en plataformas digitales.

20. Para más información sobre ese registro véase el artículo 49 del Reglamento de IA. Para más información sobre la base de datos véase el artículo 71.

21. Véase los artículos 35 del Reglamento (UE) 2016/679 o el artículo 27 de la Directiva (UE) 2016/680.

de la información que se haya ido documentando durante el proceso de diseño del sistema de IA por parte del proveedor será esencial para cumplir no sólo está con esta concreta obligación, sino con otras que se contemplan en la normativa de protección de datos, tal y como ocurre con el principio de privacidad desde el diseño.

En quinto lugar, los responsables del despliegue de sistemas de IA cuya finalidad se considera de alto riesgo²² deberán informar a las personas físicas que dichos sistemas están tomando decisiones total o parcialmente automatizadas sobre esas personas. A diferencia del Reglamento General de Protección de Datos que prevé un régimen más garantista para las decisiones plenamente automatizadas respecto de las parcialmente automatizadas²³, el RIA pone el foco de atención sobre todo en el tipo de sistema de IA que se utilizan y no en la participación más o menos activa de la persona en el proceso decisorio del sistema de IA.

En el caso de sistemas de IA cuya finalidad sea la aplicación de la ley, la información que se facilitará será la indicada en el artículo 13 de la Directiva de datos de carácter personal policiales²⁴, texto europeo que en España ha sido transpuesto por la Ley Orgánica 7/2021²⁵.

En sexto lugar, se establece un deber general de cooperación con las autoridades nacionales competentes con relación al cumplimiento del RIA.

2.3. Obligaciones específicas cuando se utilice un sistema de inteligencia artificial con finalidad de identificación biométrica

El RIA contempla una serie de actuaciones que ha de llevar a cabo el responsable del despliegue cuando pretenda utilizar un sistema de IA de alto riesgo de identificación biométrica selectiva en diferido.

Estas actuaciones tratan de asegurar que el uso de estos sistemas de IA, que llevan aparejados importantes riesgos de uso, se realice con un mínimo de garantías. Entre esas garantías cabe destacar: la necesidad de solicitar autorización previa para usar estos sistemas con esa finalidad, el uso no indiscriminado del sistema de IA sobre personas, la necesidad de presentar informes anuales a las autoridades de vigilancia del mercado, así como las de protección de datos sobre el uso que han hecho de estas herramientas²⁶.

22. Artículo 6.2 y Anexo III del Reglamento de IA.

23. Véanse los artículos 13.2.f), 14.2.g), 15.1.h) y 22. Un análisis de estos preceptos se puede ver en: Cotino Hueso, L., «Derechos y garantías ante el uso público y privado de inteligencia artificial, robótica y big data», en Bauzá, Marcelo (dir.), *El Derecho de las TIC en Iberoamérica, Obra Colectiva de FIADI*. La Ley— Thompson-Reuters, Montevideo, pp. 917-952.

24. Artículo 13 Directiva (UE) 2016/680 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por parte de las autoridades competentes para fines de prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de sanciones penales, y a la libre circulación de dichos datos y por la que se deroga la Decisión Marco 2008/977/JAI del Consejo.

25. Artículo 21 de la Ley Orgánica 7/2021, de 26 de mayo, de protección de datos personales tratados para fines de prevención, detección, investigación y enjuiciamiento de infracciones penales y de ejecución de sanciones penales.

26. Artículo 26.10 del Reglamento de IA.

2.4. La obligación de realizar una evaluación de impacto relativa a los derechos fundamentales

Como se ha indicado inicialmente, esta obligación no estaba contenida en las primeras versiones del texto originario del RIA. Fue el Parlamento Europeo el que inicialmente apostó por introducir esta medida como garantía general en favor de los afectados sobre los cuales los sistemas de IA adoptarán decisiones²⁷. Aunque con algunos cambios, la redacción final de este precepto es muy parecida a la que propuso el Parlamento Europeo, si bien, el nivel de exigencias se ha reducido²⁸.

La obligación de llevar a cabo una evaluación de impacto recae sobre determinados responsables del despliegue, estos son:

Por un lado, cualquier responsable del despliegue que sea autoridad pública o entidades privadas que preste sus servicios a dichas autoridades cuando el sistema de IA de alto riesgo tenga como finalidad algunas de las indicadas en el Anexo III del Reglamento. Quedan excluidos únicamente aquellos sistemas de IA de alto riesgo que se utilicen como componente de seguridad de gestión y funcionamiento de infraestructuras digitales críticas²⁹.

Por otro lado, cualquier responsable del despliegue, independientemente de si es o no autoridad pública, que utilice su sistema de IA para evaluar la solvencia de personas físicas o establecer su calificación crediticia o utilice dicho sistema para la evaluación de riesgos y la fijación de precios en relación con las personas físicas en el caso de los seguros de vida y de salud.

En ambos supuestos, sólo estarán obligados a realizar esta evaluación de impacto los responsables del despliegue que apliquen el primer uso del sistema de IA. En posteriores usos, el responsable del despliegue podrá basarse en la evaluación de impacto inicial que se realizó salvo que alguno de los elementos que han de estar presentes en la evaluación se hayan modificado o alterado con el uso del sistema. En este último supuesto, el responsable del despliegue deberá actualizar la evaluación de impacto en la medida que ésta se haya visto modificada por los cambios o alteraciones sufridas en el sistema de IA.

La evaluación de impacto consistirá en la realización de toda una serie de actuaciones que están estrechamente vinculadas a la información que en su caso haya facilitado el proveedor al responsable del despliegue del sistema de IA.

En primer lugar, se deberá realizar una descripción de los procesos que llevará a cabo el responsable del despliegue en los que manejará el sistema de IA, así como el tiempo durante el cual pretende utilizarlo y su frecuencia de uso.

27. Véase la enmienda 413 de la resolución del Parlamento Europeo el 14 de junio de 2023 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión.

28. Una aproximación a diferentes modelos de evaluaciones de impacto de sistemas de IA se puede ver en: Simón Castellano, P., *La evaluación de impacto algorítmico en los derechos fundamentales*, Aranzadi, 2023.

29. Infraestructuras críticas como el del tráfico rodado o del suministro de agua, gas, calefacción o electricidad, etc. Anexo III. Punto 2.

En segundo lugar, se establecerán las categorías de personas físicas y grupos que puedan verse afectados, así como los riesgos específicos que pueden afectar a dichas personas durante el uso del sistema de IA. Naturalmente, esta información podrá ser en parte elaborada con la documentación que le haya facilitado el proveedor en virtud de las obligaciones de transparencia que el RIA le impone a este último³⁰.

En tercer lugar, se deberá realizar una descripción de las medidas de vigilancia humana que en su caso pretenda desplegar para utilizar el sistema, así como cualquier otra medida que tenga a como objetivo reducir los potenciales riesgos que el uso del sistema pueda generar. Sobre esto último, el RIA obliga explícitamente a los responsables del despliegue a establecer acuerdos de gobernanza interna y mecanismos de reclamación³¹.

Para facilitar la elaboración de la evaluación de impacto, la Oficina de la IA desarrollará un cuestionario simplificado. Además, si el responsable del despliegue ya ha realizado una evaluación de impacto derivada de la normativa de protección de datos de carácter personal, la evaluación de impacto contemplada en el RIA será complementaria a la de protección de datos. Curiosamente, el Parlamento Europeo apostó porque dicha evaluación de impacto sobre protección de datos personales se publicara, sin embargo, esta propuesta no prosperó³².

Realizada la evaluación de impacto, el responsable del despliegue notificará sus resultados a la autoridad de vigilancia del mercado competente. Esta notificación no se llevará a cabo en aquellos supuestos excepcionales en los que la autoridad de vigilancia del mercado haya autorizado el uso del sistema de IA aunque tal sistema no haya pasado un proceso de evaluación de la conformidad³³.

La configuración de esta evaluación de impacto presenta en nuestra opinión dos objetivos claros.

Por un lado, se materializan específicamente los potenciales riesgos que ese sistema de IA pueda generar para los derechos fundamentales de las personas, así como las medidas para mitigarlos en el contexto específico donde se utilizará dicho sistema de IA. Recordemos que el proveedor ya habrá implementado un sistema de gestión de riesgos y habrá tenido en cuenta las potenciales afectaciones a los derechos fundamentales³⁴. Ese sistema de riesgos elaborado por el proveedor será en muchos casos la base esencial para elaborar la evaluación de impacto. Ello sobre todo ocurrirá cuando el sistema de IA haya sido desarrollado específicamente para el responsable del despliegue y el proveedor de ante mano conozca los usos potenciales e incluso los colectivos específicos a los que irá dirigido el uso del sistema. Sin embargo, la evaluación de impacto sobre todo aportará un valor añadido cuando ese sistema de

30. Véase el artículo 13.

31. Artículo 27.1.f) Reglamento de IA.

32. Enmienda 419 de la resolución del Parlamento Europeo el 14 de junio de 2023 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión.

33. En supuestos excepcionales el Reglamento de IA autoriza a las autoridades de vigilancia de mercado a al uso de sistemas de IA de alto riesgo a pesar de que estos no hayan pasado un proceso de evaluación de la conformidad. Artículo 46.1 Reglamento de IA.

34. Artículo 9.2.a) Reglamento de IA.

IA sea implementado por diferentes responsables del despliegue que, si bien tendrán en cuenta el sistema de riesgos presentado por el proveedor, deberán adaptarlo a su contexto específico respetando siempre la finalidad prevista del sistema de IA.

Por otro lado, la evaluación de impacto también será importante a efectos de la información que el responsable del despliegue facilitará a las autoridades de vigilancia del mercado competentes sobre el sistema. Ello es relevante porque recordemos que la mayoría de los sistemas de IA que están regulados en el anexo III tienen contemplada la autoevaluación de la conformidad, lo que supone que más allá del control interno del proveedor, no existe ningún ente previo al despliegue del sistema de IA que evalúe de algún modo la conformidad del sistema de IA. A través de las obligaciones de registros de los sistemas que se contemplan en el Reglamento y el deber de notificación de la evaluación de impacto se logra al menos que las autoridades de vigilancia del mercado tengan en el radar los sistemas de IA de alto riesgo que se están utilizando, así como la información esencial sobre estos sistemas y las entidades que los utilizan.

3. EL REPRESENTANTE AUTORIZADO Y SUS OBLIGACIONES

El representante autorizado es definido por el RIA como toda persona física o jurídica ubicada o establecida en la UE que haya recibido y aceptado el mandato por escrito de un proveedor de un sistema de IA para cumplir las obligaciones y llevar a cabo los procedimientos establecidos en esta norma europea en representación de dicho proveedor³⁵.

Antes de que un proveedor que está establecido fuera de la UE quiera comercializar su sistema de IA en la UE, deberá nombrar mediante un mandato escrito a un representante autorizado que esté establecido en la UE. En este sentido, destacar que no todas las legislaciones europeas de productos contemplan esta figura como obligatoria³⁶, es más, algunas directamente ni hacen mención al representante autorizado.

El mandato establecido entre el proveedor y el representante autorizado resulta esencial a la hora de valorar las tareas que ha de realizar éste último.

En primer lugar, estas tareas habilitan al representante autorizado a verificar que el proveedor ha realizado la evaluación de la conformidad pertinente, así como la documentación técnica y la declaración de conformidad del sistema de IA.

En segundo lugar, deberán conservar al menos durante 10 años los datos de contacto del proveedor, una copia de la declaración de conformidad, la documentación técnica del sistema de IA y, en su caso, el certificado expedido por el organismo notificado cuando haya intervenido en el proceso de evaluación de la conformidad³⁷.

35. Artículo 3.5 del Reglamento de IA.

36. Ello ocurre por ejemplo en la legislación de productos como las embarcaciones de recreo o de juguetes. Artículo 8. Directiva 2013/53/UE del Parlamento Europeo y del Consejo, de 20 de noviembre de 2013, relativa a las embarcaciones de recreo y a las motos acuáticas, y por la que se deroga la Directiva 94/25/CE. Artículo 5. Directiva 2009/48/CE del Parlamento Europeo y del Consejo, de 18 de junio de 2009, sobre la seguridad de los juguetes.

37. Dependiendo del tipo de sistema de IA, la evaluación de la conformidad requerirá o no la participación de un organismo notificado. Véase el artículo 43 del Reglamento

En tercer lugar, o bien deberán registrar el sistema de IA y la información sobre éste en la base de datos de la UE establecida por el RIA³⁸, o bien, si el sistema de IA ya está registrado por el proveedor, garantizar que la información que está incorporada a dicho registro cumple con las exigencias previstas³⁹.

En cuarto lugar, los representantes autorizados, previa petición de las autoridades de vigilancia del mercado, están obligados a facilitar la información que tengan sobre el sistema de IA a efectos de demostrar la conformidad del sistema de IA. En concreto se establece expresamente que se deberá facilitar el acceso a los archivos de registros que esté generando automáticamente el sistema de IA cuando dichos archivos estén bajo control del proveedor. Ello tiene sentido, ya que recordemos que el proveedor se encuentra ubicado fuera de la UE. Además, los proveedores deberán cooperar en todas las actuaciones que lleven a cabo estas últimas para reducir o mitigar los riesgos que el sistema de IA pueda presentar.

Si el representante autorizado tiene motivos para considerar que el proveedor está incumpliendo las obligaciones establecidas en el RIA, pondrá fin al mandato e informará de ello a la autoridad de vigilancia del mercado competente y, en su caso, al organismo notificado que hubiera evaluado la conformidad del sistema de IA.

4. EL IMPORTADOR Y SUS OBLIGACIONES

El importador es toda persona física o jurídica ubicada o establecida en la UE que introduzca en el mercado un sistema de IA que lleve el nombre o la marca comercial de una persona física o jurídica establecida en un tercer país⁴⁰.

Como ahora se verá, las funciones que el Reglamento asigna a los importadores reflejan que este operador no hay que considerarlo como mero revendedor de sistemas de IA, sino que desempeña un papel crucial a la hora de garantizar la conformidad de los productos importados⁴¹.

En primer lugar, el importador, antes de introducir en el mercado un sistema de IA deberá verificar que el proveedor ha realizado una serie de actuaciones obligatorias sobre dicho sistema, estas son: haber realizado la evaluación de la conformidad pertinente, haber elaborado la documentación técnica, comprobar que el sistema lleva el Marcado CE y la declaración de conformidad y que se ha nombrado a un representante autorizado. Como se puede comprobar, se trata de toda una serie de obligaciones que el RIA exige a cualquier proveedor de un sistema de IA antes de que éste quiera introducirlo en el mercado. En estos supuestos, dado que la introducción la realiza el importador, éste último debe corroborar que el sistema se encuentra en condiciones para ello.

-
- de IA y el capítulo de esta obra colectiva que analiza la evaluación de la conformidad.
38. Las obligaciones de registro se contemplan para todos los sistemas de IA del Anexo III, salvo los sistemas de IA del punto 2 de dicho Anexo (infraestructuras críticas). Artículo 49.1. Reglamento de IA.
 39. La información que se ha de incorporar a la base de datos de la UE (artículo 71) aparece indicada en la sección A del Anexo VIII del Reglamento de IA.
 40. Artículo 3.6. RIA.
 41. Guía azul sobre la aplicación de la normativa europea relativa a los productos de 2022. p. 33.

Si el importador comprueba que el sistema de IA no es conforme a derecho, no lo introducirá en el mercado hasta que no se haya conseguido la conformidad de dicho sistema. Será necesario que el importador se ponga en contacto con el proveedor para aclarar cualquier duda acerca de la conformidad del producto.

En segundo lugar, si los importadores tienen motivos para considerar que un sistema presenta un riesgo⁴², deberán comunicar tal situación al proveedor del sistema, a los representantes autorizados y a las autoridades de vigilancia del mercado competentes.

En tercer lugar, los importadores deberán asegurarse que las condiciones de almacenamiento o transporte de los sistemas de IA no afecten a los requisitos esenciales de éste⁴³. Además, cuando proceda, en la documentación o en el embalaje se indicará el nombre del fabricante o la marca comercial, así como la dirección de contacto.

En cuarto lugar, los importadores deberán conservar durante al menos 10 años desde que se introdujo en el mercado el sistema de IA sus instrucciones de uso, la declaración de conformidad y, cuando proceda, la copia del certificado de conformidad emitido por un organismo notificado⁴⁴.

En quinto lugar, los importadores están obligados a facilitar la información que tengan sobre el sistema de IA a las autoridades de vigilancia del mercado cuando éstas se lo requieran y a cooperar en todas las actuaciones que lleven a cabo estas últimas para reducir o mitigar los riesgos que el sistema de IA pueda presentar.

5. EL DISTRIBUIDOR Y SUS OBLIGACIONES

El distribuidor es toda persona física o jurídica que forme parte de la cadena de suministro, distinta del proveedor o el importador, que comercialice un sistema de IA en el mercado de la UE⁴⁵.

El RIA establece una serie de obligaciones a los distribuidores.

En primer lugar, antes de comercializar un sistema de IA, los distribuidores verificarán que el sistema lleve el marcado CE exigido, una copia de la declaración de conformidad⁴⁶, las instrucciones de uso y el nombre comercial del importador o del proveedor. En este último caso, el proveedor también deberá haber facilitado al distribuidor el sistema de gestión de la calidad.

42. El concepto de riesgo se encuentra definido en el Artículo 79.1 del Reglamento de IA.

43. Estos requisitos se encuentran indicados en los artículos 8 a 15 del Reglamento de IA. Son: Sistema de gestión de riesgos, datos y gobernanza de datos, documentación técnica, conservación de registros, transparencia y comunicación de información, vigilancia humana, precisión, solidez y ciberseguridad.

44. Dependiendo del tipo de sistema de IA, la evaluación de la conformidad requerirá o no la participación de un organismo notificado. Véase el artículo 43 del Reglamento de IA y el capítulo de esta obra colectiva que analiza la evaluación de la conformidad.

45. Artículo 3.7 del Reglamento de IA.

46. En diversas ocasiones se han sancionado a los distribuidores de productos que estaban comercializando productos que no disponían del marcado CE o la declaración de conformidad. AN de 20 de mayo de 2010. (JUR 2010\182746).

Si con la información facilitada el distribuidor considera que el sistema de IA no es conforme con los requisitos esenciales del RIA, éste no lo comercializará hasta que el sistema haya obtenido la conformidad. Si además detecta que el sistema presenta un riesgo⁴⁷, el distribuidor informará de ello al proveedor o al importador. Es decir, el distribuidor no debería suministrar sistemas de IA cuando conozca o suponga, debido a la información que está en su poder y a su experiencia profesional, que no son conformes al RIA⁴⁸.

En segundo lugar, una vez que el sistema de IA sea comercializado, si el distribuidor considera que dicho sistema no es conforme con los requisitos esenciales del RIA, adoptará las actuaciones oportunas para corregir tal situación. Entre esas actuaciones estará la de retirar el sistema, recuperarlo o velar porque el proveedor o el importador adopte las medidas adecuadas para revertir el problema.

Además, si el distribuidor que ha comercializado el sistema detecta que éste presenta un riesgo, deberá avisar de ello al proveedor o al importador y a las autoridades de vigilancia del mercado donde haya comercializado dicho sistema de IA informándoles del riesgo y de las medidas correctoras adoptadas.

En tercer lugar, los distribuidores, al igual que los importadores, deberán asegurarse que las condiciones de almacenamiento o transporte de los sistemas de IA no comprometen los requisitos esenciales de éstos. Por tanto, la persona o personas a cargo de las condiciones de distribución deberán adoptar las medidas necesarias para proteger la conformidad del sistema de IA⁴⁹.

En cuarto lugar, los distribuidores están obligados a facilitar la información que tengan sobre el sistema de IA a las autoridades de vigilancia del mercado cuando éstas se lo requieran y a cooperar en todas las actuaciones que lleven a cabo estas últimas para reducir o mitigar los riesgos que el sistema de IA pueda presentar. Nótese que la figura del distribuidor difiere entre otras cosas de la del proveedor o del importador en la medida que el primero tiene mucha menos información sobre el sistema de IA que los segundos, sin embargo, su papel también es relevante a lo largo de toda la cadena de suministros del sistema de IA.

Obligación	Representante autorizado	Importador	Distribuidor
Asegurar que el sistema de IA cumple con el Reglamento	x	x	x
No introducir en el mercado si hay dudas		x	x
Informar en caso de que el sistema presente riesgos		x	x

47. El concepto de riesgo se encuentra definido en el Artículo 79.1 del Reglamento de IA.

48. Guía azul sobre la aplicación de la normativa europea relativa a los productos de 2022. P.35.

49. Guía azul sobre la aplicación de la normativa europea relativa a los productos de 2022. P.36.

Obligación	Representante autorizado	Importador	Distribuidor
Indicar el nombre o marca comercial		x	
Conservación de documentación	x	x	
Deber de cooperar con las autoridades	x	x	x
Obligaciones de registro	x	x	

6. POSIBLES ALTERACIONES DE LAS RESPONSABILIDADES DE LOS OPERADORES

El RIA contempla diferentes situaciones en las que los operadores⁵⁰ presentes durante la cadena de valor de la IA distintos al proveedor pasan a ser considerados proveedores y asumen las obligaciones exigidas a éste último sobre el sistema de IA⁵¹. Estas situaciones son:

En primer lugar, cuando dichos operadores pongan su nombre o marca comercial en un sistema de IA de alto riesgo que previamente se haya introducido en el mercado o puesto en servicio. En estos supuestos el proveedor y el operador podrán establecer acuerdos contractuales que estipulen el reparto de obligaciones de otro modo.

En segundo lugar, cuando el operador modifique sustancialmente un sistema de IA de alto riesgo que haya sido introducido en el mercado o puesto en servicio siempre que tal sistema siga considerándose de alto riesgo tras esa modificación sustancial⁵².

En tercer lugar, cuando el operador modifique la finalidad prevista del sistema de IA de tal manera que pase a considerarse de alto riesgo cuando inicialmente y sin el cambio de finalidad, tal sistema que estaba en el mercado no era considerado de alto riesgo.

En todos estos supuestos, el proveedor inicial dejará de ser considerado proveedor a efectos del RIA. Ahora bien, este proveedor inicial deberá cooperar estrechamente con el nuevo proveedor y facilitará la información necesaria para que éste último pueda cumplir con las obligaciones exigidas por el Reglamento⁵³. Se trata por tanto de una obligación a futuro que no depende *a priori* del destinatario de ésta (proveedor inicial) sino de los operadores posteriores que en su caso pasen a considerarse proveedores.

50. Cualquier distribuidor, importador, responsable del despliegue o tercero.

51. Véase las obligaciones exigidas al proveedor del artículo 16 del Reglamento de IA.

52. Por modificación sustancial hay que entender «un cambio en un sistema de IA tras su introducción en el mercado o puesta en servicio que no haya sido previsto o proyectado en la evaluación de la conformidad inicial realizada por el proveedor y a consecuencia del cual se vea afectado el cumplimiento por parte del sistema de IA de los requisitos establecidos en el capítulo II, sección 2, o que dé lugar a una modificación de la finalidad prevista para la que se haya evaluado el sistema de IA en cuestión». Artículo 3.23 Reglamento de IA.

53. Artículo 25.2. Reglamento de IA.

Estas obligaciones de cooperación no se exigirán al proveedor inicial cuando éste hubiera dejado claro que su sistema de IA que inicialmente no era de alto riesgo no debía ser alterado de tal manera que pasara a considerarse como tal.

A su vez, en los supuestos en los que un sistema de IA de alto riesgo que es componente de seguridad de productos contemplados en la legislación de armonización de la UE, el fabricante del producto pasará a considerarse proveedor cuando: o bien el sistema se introduzca en el mercado bajo el nombre o la marca comercial del fabricante del producto, o bien, el sistema se ponga en servicio bajo el nombre o la marca comercial del fabricante del producto después de que el producto haya sido introducido en el mercado.

Por último, cuando un tercero facilite a un proveedor de un sistema de IA de alto riesgo herramientas, servicios, componentes o procesos que integren el sistema de IA de alto riesgo, ese tercero y el proveedor deberán realizar un acuerdo escrito donde especificarán la información y documentación que en su caso sea necesaria para que el proveedor pueda cumplir con el RIA. Esta obligación no se exigirá a terceros que pongan a disposición del público herramientas, servicios o componentes distintos de modelos de IA de uso general en el marco de una licencia gratuita y abierta.

IV. LAS AUTORIDADES NOTIFICANTES

1. CONCEPTO DE AUTORIDAD NOTIFICANTE

El RIA obliga a los Estados Miembros a designar diferentes autoridades con el objetivo de asegurar su cumplimiento. Entre esas autoridades encontramos la autoridad notificante.

De acuerdo al apartado 19 del artículo 3 del RIA, la autoridad notificante es la responsable de establecer y llevar a cabo los procedimientos necesarios para la evaluación, la designación y la notificación de los organismos de evaluación de la conformidad, así como de su supervisión⁵⁴.

El concepto de autoridad notificante no es novedoso del RIA, existen otros textos europeos que ya hacen mención a esta figura. El RIA sigue la estructura configurada por el Nuevo Marco Legislativo (NML). El NML está integrado por varios textos europeos que establecen unas bases comunes sobre la comercialización, evaluación y vigilancia de productos en la Unión Europea⁵⁵. Todos estos textos contemplan la figura de la autoridad notificante.

Cada Estado Miembro decidirá si nombran una o varias autoridades notificantes, debiendo existir al menos una por país⁵⁶. Hasta la fecha, como regla general, las

54. Artículo 3.19 Reglamento de IA.

55. Las tres textos legales que conforman el Nuevo Marco Legislativo son: el Reglamento (CE) n.º 765/2008 del Parlamento Europeo y del Consejo por el que se establecen los requisitos de acreditación y vigilancia del mercado de los productos; la Decisión n.º 768/2008/CE del Parlamento Europeo y del Consejo sobre un marco común para la comercialización de los productos y; el Reglamento (UE) 2019/1020 del Parlamento Europeo y del Consejo relativo a la vigilancia del mercado y la conformidad de los productos.

56. Artículo 28.1 y 70.1 del Reglamento de IA.

autoridades nacionales que han sido designadas para llevar a cabo estas actividades en virtud de otras legislaciones que regulan la comercialización y supervisión de productos que también siguen el NML son Direcciones Generales o Subdirecciones Generales integradas dentro de un Ministerio determinado⁵⁷. Ello ocurre por ejemplo en el caso de productos como los juguetes, ascensores, equipos radioeléctricos, entre otros⁵⁸.

Producto	Autoridad Notificante
Máquinas	SG Calidad y Seguridad industrial (Ministerio de Industria)
Aparatos que queman combustible gaseoso	
Equipos a presión	
Instalaciones de transporte por cable	
Protección individual	
Ascensores	
Protección para uso en atmósferas potencialmente explosivas	
Equipos radioeléctricos	Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales (Ministerio de Transformación digital)
Juguetes	DG Consumo (Ministerio de Consumo)
Embarcaciones de recreo	DG Marina mercante (Ministerio de Transportes)
Productos sanitarios	Ministerio de Sanidad

Teniendo en cuenta lo señalado anteriormente, la autoridad notificante en España para el cumplimiento del RIA será posiblemente la Secretaría de Estado de Digitalización de Inteligencia Artificial o algún organismo administrativo dependiente de ésta⁵⁹. Si bien, podrá haber otras autoridades notificantes si así lo

57. El listado completo de autoridades notificantes se puede consultar en: <https://webgate.ec.europa.eu/single-market-compliance-space/#/notified-bodies/notifying-authorities?filter=countryId:724>

58. Esos productos son, en concreto, máquinas, juguetes, ascensores, equipo y sistemas de protección para uso en atmósferas potencialmente explosivas, equipos radioeléctricos, equipos a presión, equipos de embarcaciones de recreo, instalaciones de transporte por cable, aparatos que queman combustibles gaseosos, productos sanitarios y productos sanitarios para diagnóstico *in vitro*. Considerando 50 del RIA.

59. Dentro de la Secretaría de Estado para la Digitalización y la Inteligencia Artificial se encuentra la Dirección General de Digitalización e Inteligencia Artificial y dentro de ésta última la Subdirección General de Inteligencia Artificial y Tecnologías Habilitadoras Digitales. Real Decreto 210/2024, de 27 de febrero, por el que se establece

considera España en ámbitos específicos como justicia o aplicación de la ley entre otras.

2. ACTIVIDADES DE LA AUTORIDAD NOTIFICANTE

Las actividades que se asignan a estas autoridades se focalizan en garantizar que una entidad pública o privada cuenta con los medios humanos, organizativos y técnicos suficientes para poder realizar la evaluación de la conformidad de sistemas de IA regulados por el RIA. Esas entidades son conocidas como organismo de evaluación de la conformidad⁶⁰.

Antes de que un organismo de evaluación de la conformidad pueda realizar evaluaciones de conformidad de sistemas de IA, la autoridad notificante deberá corroborar que efectivamente ese organismo de evaluación de la conformidad puede realizar dichas evaluaciones. Una vez que la autoridad notificante comprueba que ese organismo de evaluación de la conformidad cumple con los requisitos para poder realizar evaluaciones de conformidad de sistemas de IA⁶¹, la autoridad notificante «notificará» a la Comisión Europea y al resto de Estados Miembros tal situación. Desde ese momento, ese organismo será «organismo notificado» habilitado para realizar evaluaciones de conformidad de sistemas de IA conforme al RIA.

Entre las actividades que se le asignan a la autoridad notificante, la evaluación y la supervisión podrán ser otorgadas a un organismo nacional de acreditación⁶², en el caso de España, ese organismo de acreditación es actualmente ENAC⁶³.

A la hora de llevar a cabo las actividades asignadas, la autoridad notificante no podrá ejercer ninguna actividad que efectúen los organismos notificados, ni ningún servicio de consultoría de carácter comercial o competitivo. Además, deberán evitar cualquier conflicto de intereses que pueda surgir entre los organismos de evaluación de la conformidad y éstas. Con ello se pretende asegurar que la evaluación y notificación de los organismos de evaluación de la conformidad se lleve a cabo de la forma más imparcial y objetiva posible por parte de la autoridad notificante.

la estructura orgánica básica del Ministerio para la Transformación Digital y de la Función Pública.

60. Organismo de evaluación de la conformidad: «*un organismo independiente que desempeña actividades de evaluación de la conformidad, como el ensayo, la certificación y la inspección*». Artículo 3.21. Reglamento de IA.
61. El proceso de notificación de los organismos de evaluación de la conformidad se encuentra en los artículos 29 y 30 del Reglamento de inteligencia artificial. Existe un apartado de esta obra que lo analiza expresamente.
62. Organismo nacional de acreditación: «el único organismo de un Estado miembro con potestad pública para llevar a cabo acreditaciones». Artículo 2.11. Reglamento (CE) n.º 765/2008 del Parlamento Europeo y del Consejo, de 9 de julio de 2008, por el que se establecen los requisitos de acreditación y vigilancia del mercado relativos a la comercialización de los productos y por el que se deroga el Reglamento (CEE) n.º 339/93.
63. Artículo 1. Real Decreto 1715/2010, de 17 de diciembre, por el que se designa a la Entidad Nacional de Acreditación (ENAC) como organismo nacional de acreditación de acuerdo con lo establecido en el Reglamento (CE) n.º 765/2008 del Parlamento Europeo y el Consejo, de 9 de julio de 2008, por el que se establecen los requisitos de acreditación y vigilancia del mercado relativos a la comercialización de los productos y por el que se deroga el Reglamento (CEE) n.º 339/93.

La exigencia de independencia y objetividad exigida en sus actividades a estas autoridades no se contemplaba en la primera versión del RIA⁶⁴, sin embargo, en los diferentes cambios que se han ido introduciendo se han incorporado. Estas exigencias de imparcialidad y objetividad son plenamente aplicables respecto de los potenciales organismos de evaluación de la conformidad que evaluarán, tanto si son públicos como privados. En este sentido, la mayoría de los organismos notificados que han sido notificados para realizar evaluaciones de conformidad de otros productos son privados, sin embargo, nada impide que pueda ser un organismo público, tal y como ocurre con los procesos de verificación de los productos sanitarios⁶⁵.

Por otro lado, la autoridad notificante deberá organizarse de tal manera que las decisiones relativas a la notificación serán adoptadas por personas competentes distintas a las que realizaron la evaluación de los organismos de evaluación de la conformidad. Estas personas deberán tener la suficiente competencia para realizar de forma adecuada las tareas asignadas, debiendo en su caso ostentar conocimientos especializados en ámbitos como la tecnología de la información, la inteligencia artificial y los derechos fundamentales. Esto último resulta esencial, ya que a diferencia de otras normas sobre productos que se centran en reducir o mitigar los riesgos relacionados con la seguridad y la salud de las personas, en el caso de los sistemas de IA hay que añadir además los potenciales derechos fundamentales que se pueden ver afectados.

Actividades de las autoridades notificantes ⁶⁶	Articulado
Regulación general de las Autoridades Notificantes	Artículo 28
Evaluación y notificación de organismos de evaluación de la conformidad	Artículo 29 y 30
Supervisión de organismos notificados	Artículo 33.4, 34.3, 36, 37, 38 y 45

V. MEDIDAS DIRIGIDAS A LOS PROVEEDORES Y USUARIOS A PEQUEÑA ESCALA

Es indudable que el RIA tiene como objetivo reducir o mitigar los riesgos que pueden generar y generarán el uso de sistemas de inteligencia artificial. Para lograr ese objetivo, los diferentes operadores presentes durante el ciclo de vida de los sistemas de IA han de cumplir con todas las obligaciones que esta norma les impone.

64. Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de inteligencia artificial) y se modifican determinados actos legislativos de la Unión. De 21 de abril de 2021.

65. El Centro Nacional de Certificación de Productos Sanitarios es el único organismo notificado en España para realizar la evaluación de la conformidad de los productos sanitarios conforme al Reglamento 2017/745. Artículo 35.bis. Real Decreto 1275/2011, de 16 de septiembre, por el que se crea la Agencia estatal «Agencia Española de Medicamentos y Productos Sanitarios» y se aprueba su Estatuto.

66. Estas actividades se analizan de forma más pormenorizada en esta obra en el apartado que estudia los organismos notificados.

La implementación y adaptación a esta norma será más fácil para aquellas entidades privadas que cuenta con más personal y recursos. Son estas entidades además las que en muchos casos están desarrollando técnicas para cumplir con las exigencias normativas. A modo de ejemplo, las grandes empresas del sector como Microsoft, IBM, Google entre otras, no paran de realizar fuertes inversiones para implementar o facilitar la llamada inteligencia artificial explicable. Es decir, sus técnicas de cumplimiento normativo marcarán las pautas a seguir por entidades menores.

Para evitar que las innumerables exigencias normativas previstas en esta norma acaben asfixiando a las empresas que no tienen tantos recursos para adaptarse a ésta, el RIA contempla toda una serie de medidas destinadas a los proveedores y responsables del despliegue de sistemas de IA que sean pymes y empresas emergentes.

Estas medidas obligan esencialmente a tres sujetos, estos son: los Estados Miembros, los organismos notificados y la Oficina de la IA.

Por lo que se refiere a los *Estados Miembros*, en primer lugar, éstos deberán priorizar la participación de las pymes en los espacios controlados de pruebas que pretendan llevar a cabo. Así se contempla por ejemplo en el entorno controlado de pruebas de España⁶⁷. En este sentido, el propio RIA establece que una de las finalidades de estos espacios controlados de pruebas será la de facilitar y acelerar el acceso al mercado de la UE de sistemas de IA desarrollados por pymes⁶⁸.

En segundo lugar, se deberán establecer canales adecuados de asesoramiento que ayuden a estas empresas a implementar adecuadamente el RIA, así como formaciones sobre esta misma norma y su implementación. Así, algunas de estas funciones ya se atribuyen a la AESIA sin distinción del tamaño o tipología de empresa⁶⁹.

En tercer lugar, los Estados Miembros deberán fomentar la participación de las pymes en el proceso de desarrollo de la normalización. Esta medida resulta esencial. Recordemos que una forma que facilita a los fabricantes de productos el cumplimiento de una directiva o un reglamento europeo aplicable a ese producto es el uso de normas armonizadas creadas por organismos de normalización europeos⁷⁰. De esta manera, aunque las normas armonizadas no son de obligada cumplimiento, su uso otorga presunción de conformidad a los productos que se han diseñado tomándolas como referencia, de ahí que los fabricantes normalmente acudan a éstas⁷¹.

67. Artículo 8.1.j). Real Decreto 817/2023, de 8 de noviembre, que establece un entorno controlado de pruebas para el ensayo del cumplimiento de la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial. Así también se contempla en el Estatuto de la AESIA. Artículo 25.a). 4º. Real Decreto 729/2023, de 22 de agosto, por el que se aprueba el Estatuto de la Agencia Española de Supervisión de Inteligencia Artificial.

68. Artículo 57.9.e) Reglamento de IA.

69. Artículo 4.3. a) y b). Real Decreto 729/2023, de 22 de agosto, por el que se aprueba el Estatuto de la Agencia Española de Supervisión de Inteligencia Artificial.

70. Estos organismos europeos de normalización son CEN, CENELEC y ETSI.

71. Álvarez García, V y Tahirí Moreno, J., «La regulación de la inteligencia artificial en Europa a través de la técnica armonizadora del nuevo enfoque». *Revista General de Derecho Administrativo*, núm 63 (2023).

Será esencial que en el proceso de elaboración de estas normas armonizadas tengan un papel relevante el sector de las pymes y empresas emergente⁷².

En cuanto a los *organismos notificados*, en primer lugar, se obliga a éstos a establecer diferentes tasas en función de la tipología y tamaño de empresa por los servicios realizados durante la evaluación de la conformidad de sistemas de IA regulada en el RIA. Se ha estimado que un proceso de evaluación de la conformidad de los contemplados en el RIA puede llegar a costar entre 16.800 y 23.000€ a las entidades que pretendan poner en el mercado sus sistemas de IA⁷³.

En segundo lugar, también se agiliza el proceso de evaluación de la conformidad relacionado con la presentación de la documentación técnica que deberán facilitar las pymes a los organismos notificados. Así, la Comisión Europea desarrollará un formulario simplificado que ayudará a las Pymes a la hora de documentar la documentación técnica que éstas han de presentar a los organismos notificados cuando vayan a verificar que su sistema de IA cumple con el RIA⁷⁴.

Para finalizar, la *Oficina de la IA* también desarrollará actuaciones que fomenten la correcta adecuación de las Pymes al RIA. Entre estas actuaciones encontramos la de sensibilización del cumplimiento normativo, la creación de una plataforma única de información sobre este texto legal, así como el diseño de modelos normalizados que ayuden a implementar las diferentes obligaciones previstas en el RIA.

Sujeto obligado	Obligación	Artículo
Estado Miembro	Priorización en la participación de pilotos de pruebas Asesoramiento particularizado Formaciones específicas Participación en el proceso de normalización	62.1.
	Priorización en la participación de pilotos de pruebas	57.9.e)
Organismo notificado	Tasas adaptadas sobre evaluaciones de conformidad	62.2
	Formularios simplificados	11.1
Oficina de la IA	Sensibilización sobre la aplicación del Reglamento. Creación de plataforma informativa Diseño de modelos normalizados	62.3

VI. LA NOTIFICACIÓN DE INCIDENTES GRAVES

1. CONCEPTO DE INCIDENTE GRAVE

De acuerdo al artículo 3.49 del RIA, un incidente grave es aquel incidente o defecto de funcionamiento de un sistema de IA que, directa o indirectamente genere: el

72. McFadden/Jones/Taylor/Osborn, *Harmonising Artificial Intelligence: The role of standards in the EU AI Regulation*, Oxford Commission on AI & Good Governance (2021). P.20.
73. Comisión Europea. *Study to support an impact assessment of regulatory requirements for Artificial Intelligence in Europe*. 2021. P.12.
74. Artículo 11.1 Reglamento Europeo de Inteligencia Artificial.

fallecimiento de una persona o un perjuicio grave para su salud, una alteración grave del funcionamiento de las infraestructuras críticas, el incumplimiento de alguna obligación prevista en el derecho de la Unión que tenga como objetivo proteger derechos fundamentales o cualquier daño grave a la propiedad o al medio ambiente.

Como puede comprobarse, la consideración de incidente grave está pensada para las afectaciones más relevantes que puede generar un sistema de IA en la esfera de las personas, la propiedad o el medio ambiente. En este sentido, los sistemas de IA están implementados en diferentes productos sanitarios que se utilizan para la detección de enfermedades o en las operaciones de pacientes, en infraestructuras críticas como puede ser la canalización del agua o el despliegue de la línea eléctrica, o el manejo ingente de datos personales. En todos estos supuestos, un incidente en el funcionamiento de un sistema de IA se considerará grave.

En todos estos supuestos, la notificación tendrá como objetivo reducir el riesgo de que ese incidente grave se pueda repetir o, en el caso de que se repita, mitigar los posibles daños que se hayan generado.

Entendemos relevante analizar los supuestos en los que un incidente se considera grave cuando éste suponga un incumplimiento de las obligaciones derivadas del Derecho de la Unión destinadas a proteger los derechos fundamentales⁷⁵.

Son dos las condiciones acumulativas que se prevén para considerar este incidente grave, por un lado, que se genere un incumplimiento de una obligación presente en el Derecho de la UE, y por otro lado, que ese incumplimiento derivado de la norma tenga como objetivo proteger un derecho fundamental.

Por lo que se refiere a la obligación derivada del Derecho de la UE, hemos de incluir cualquier obligación que se reconozca en los diferentes textos normativos previsto en el ordenamiento jurídico de la UE, así como los textos nacionales que en su caso se hayan dictado en virtud de esa legislación europea. Piénsese por ejemplo en una directiva que se ha de transponer o en un reglamento europeo que exige de la colaboración de los Estados Miembros para desarrollar ciertos elementos de éste. A su vez, por derecho fundamental hay que entender todos los derechos reconocidos en la Carta de Derechos Fundamentales de la Unión Europea⁷⁶.

En definitiva, cualquier incidente en un sistema de IA que genere el incumplimiento de una obligación reconocida en derecho de la UE que afecte a un derecho fundamental reconocido en la Carta se considerará incidente grave.

2. ¿QUIÉN, ANTE QUIÉN Y CUÁNDO SE HA DE NOTIFICAR?

Corresponde al proveedor del sistema de IA que ha sufrido el incidente grave notificarlo a las autoridades de vigilancia del mercado de los Estados miembros en los que se haya producido⁷⁷. No obstante, cuando el incidente lo haya detectado el responsable del despliegue, éste último lo notificará al proveedor, y, a continuación, al importador o distribuidor y a la autoridad de vigilancia del mercado pertinente⁷⁸.

75. Artículo 3.49.c) del Reglamento de IA.

76. Así lo interpretamos teniendo en cuenta los considerandos 1, 2 y 48 del Reglamento de Inteligencia Artificial.

77. Artículo 73.1. Reglamento de IA.

78. Artículo 28.5. Reglamento de IA.

Nótese que el número de autoridades de vigilancia del mercado a las que se ha de notificar el incidente variará en función del número de Estados miembros en los que haya podido generarse el incidente y del tipo de sistema de IA que haya podido generarlo. En este sentido, el RIA contempla diferentes autoridades de vigilancia del mercado para los distintos tipos de sistemas de IA que regula⁷⁹.

Como regla general, el proveedor, o en su caso el responsable del despliegue, tienen como máximo un plazo de 15 días para notificar el incidente desde que éste se produjo. Ahora bien, ese plazo de 15 días será más reducido cuando se den determinadas circunstancias.

En primer lugar, la notificación se realizará desde el momento en el que el proveedor haya establecido un nexo causal o una posibilidad razonable de dicho vínculo entre el sistema de IA y el incidente grave.

En segundo lugar, la notificación se realizará al día siguiente del incidente cuando éste genere una alteración grave e irreversible a la gestión o al funcionamiento de infraestructuras críticas o tal incidente de lugar a una infracción generalizada⁸⁰, es decir, un acto u omisión contrario a la legislación de la UE que afecte o pueda afectar a un colectivo de personas en varios Estados Miembros⁸¹.

En tercer lugar, cuando el incidente de lugar al fallecimiento de una persona, la notificación se realizará en el momento que se establezca un nexo causal entre dicho incidente y el funcionamiento del sistema de IA. En todo caso, la notificación no podrá postergarse más allá del plazo de 10 días desde que sucedió dicho incidente.

Tipo de incidente	Plazo	
	Mínimo	Máximo
Regla general	En cuanto se conozca el nexo causal	15
Infraestructuras críticas o infracción generalizada	Día siguiente	
Fallecimiento de persona	En cuanto se conozca el nexo causal	10

Junto a los plazos máximos de notificación, el RIA contempla dos supuestos en los que, dependiendo del tipo de sistema de IA, la notificación de incidentes queda reducida solo a determinados supuestos.

Por un lado, para los sistemas de IA de alto riesgo que sean productos o componentes de seguridad de productos sanitarios o productos de diagnóstico

79. Por ejemplo, ello ocurre en los casos de sistemas de IA cuya finalidad sea el sector bancario, el ámbito judicial o los sistemas de IA utilizados para fines policiales, entre otros. Artículo 74 Reglamento de IA.

80. El concepto de infraestructura crítica se define en el artículo 3.62 del Reglamento de IA.

81. El concepto de infracción generalizada se describe en el artículo 3.61 del Reglamento de IA.

in vitro⁸², la notificación de incidentes graves se limitará a los supuestos en los que tales incidentes hayan supuesto el incumplimiento de obligaciones derivadas del Derecho de la Unión destinadas a proteger los derechos fundamentales⁸³. Entendemos que esta previsión se debe al hecho de que las legislaciones de estos productos ya contemplan sus propias notificaciones de incidentes graves que presentan una estructura muy similar a la establecida en el RIA, pero adaptada a la realidad de los productos sanitarios⁸⁴.

Por otro lado, para los sistemas de IA de alto riesgo enumerados en el Anexo III del RIA que se encuentren sujetos a instrumentos legislativos de la Unión que establezcan obligaciones equivalentes sobre notificación de incidentes graves, la notificación de estos incidentes se limitará al supuesto indicado previamente, es decir, a aquellos casos en los que tales incidentes hayan supuesto el incumplimiento de obligaciones derivadas del Derecho de la Unión destinadas a proteger los derechos fundamentales⁸⁵.

3. ACTUACIONES POSTERIORES A LA NOTIFICACIÓN DEL INCIDENTE

Una vez que se ha notificado el incidente grave a la Autoridad de Vigilancia del mercado competente, el RIA contempla diferentes actuaciones que los proveedores y dichas autoridades han de llevar a cabo.

Por parte de los *proveedores*, éstos procederán a realizar las investigaciones necesarias para esclarecer el incidente sucedido. Consideramos que estas investigaciones se habrán iniciado previamente a la notificación del incidente cuando en un primer momento se haya indagado en el posible el nexo causal entre el incidente sucedido y el fallo en el sistema de IA.

A su vez, los proveedores deberán realizar una evaluación de riesgos de tal incidente y las medidas correctoras para reducir o mitigarlos⁸⁶. Es posible que los diferentes potenciales incidentes graves que pudieran suceder hayan sido contemplados en el sistema de riesgos que necesariamente todo proveedor ha de

82. Estos productos se regulan en el Reglamentos 2017/745 (productos sanitarios) y en el Reglamento 2017/746. (productos sanitarios para diagnóstico in vitro).

83. Artículo 73.11 del Reglamento de IA.

84. Véase el artículo 2.65 sobre la definición de incidente grave y el artículo 87.1 sobre la notificación de dichos incidentes en el Reglamentos 2017/745 (productos sanitarios). A su vez, véase el artículo 2.68 sobre la definición de incidente grave y el artículo 82 sobre notificación de incidentes graves en el Reglamento 2017/746. (productos sanitarios para diagnóstico in vitro).

85. En las versiones iniciales del Reglamento de IA esta previsión estaba específicamente diseñada para aquellos supuestos en los que una entidad financiera utilizaba un sistema de IA con la finalidad de evaluar la solvencia financiera de las personas (Anexo III. Punto 5 b) del Reglamento de IA). Sin embargo, en la última versión conocida esta mención expresa ha sido retirada. Pese a ello, seguimos pensando que este percepto está diseñado esencialmente para estos supuestos. Véase la versión inicial del Artículo 62.3 de la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (ley de inteligencia artificial) y se modifican determinados actos legislativos de la Unión. Resolución de 21 de abril de 2021.

86. Artículo 73.7 Reglamento de IA.

desarrollar de su sistema de IA⁸⁷. Si tales incidentes se previeron en dicho sistema de riesgos, el proveedor o el responsable del despliegue puede aplicar las medidas previstas en él.

Además, deberán cooperar con las diferentes autoridades, y cuando sea el caso, con el organismo notificado que haya evaluado la conformidad de su sistema de IA⁸⁸. No pudiendo realizar modificaciones del sistema que puedan repercutir en futuras investigaciones sobre las causas de dicho incidente.

Por parte de las *autoridades de vigilancia del mercado*⁸⁹, una vez hayan sido notificadas de ese incidente grave, en un plazo máximo de 7 días adoptarán diferentes medidas en función de la gravedad de éste. Estas medidas pueden ir desde la prohibición de venta de los sistemas de IA, pasando por la retirada o la recuperación de éstos⁹⁰. Además, deberán informar de ello a la Comisión Europea cuando las medidas que pretendan adoptar rebasen las fronteras del Estado Miembro donde la autoridad de vigilancia del mercado ejerce su competencia. Algo que ocurre frecuentemente con los sistemas de IA, los cuales, no suelen estar diseñados únicamente para ofrecer sus servicios en un solo Estado Miembro.

Independientemente de que se adopten o no las medidas previamente indicadas, la autoridad de vigilancia del mercado deberá informar del incidente grave a la Comisión Europea a través del sistema de intercambio rápido de información previsto en el Reglamento Europeo de Vigilancia del Mercado⁹¹.

Por otro lado, cuando a una autoridad de vigilancia del mercado le notifiquen un incidente grave que suponga un incumplimiento de una obligación de derecho europeo destinada a proteger un derecho fundamental, ésta deberá informar a la autoridad competente encargada de supervisar y velar por el cumplimiento del derecho fundamental afectado por tal incidente⁹². Entre esas autoridades encontramos por ejemplo las encargadas de vigilar y supervisar la normativa de protección de datos de carácter personal.

VII. CONCLUSIONES

1. El RIA define y establece las diferentes funciones y obligaciones a los distintos operadores que intervienen durante la cadena de valor de los sistemas de IA.

87. Artículo 9. Reglamento de IA.

88. Dependiendo del tipo de sistema de IA, la evaluación de la conformidad requerirá o no la participación de un organismo notificado. Véase el artículo 43 del Reglamento de IA y el capítulo de esta obra colectiva que analiza la evaluación de la conformidad.

89. Artículos 73.8, 9 y 12 del Reglamento de IA.

90. Artículo 19.1. Reglamento (UE) 2019/1020 del Parlamento Europeo y del Consejo, de 20 de junio de 2019, relativo a la vigilancia del mercado y la conformidad de los productos y por el que se modifican la Directiva 2004/42/CE y los Reglamentos (CE) n.º 765/2008 y (UE) n.º 305/2011.

91. Artículo 73.12 del Reglamento de IA y el artículo 20 del Reglamento (UE) 2019/1020 del Parlamento Europeo y del Consejo, de 20 de junio de 2019, relativo a la vigilancia del mercado y la conformidad de los productos y por el que se modifican la Directiva 2004/42/CE y los Reglamentos (CE) n.º 765/2008 y (UE) n.º 305/2011.

92. Artículo 73.8 77.1 del Reglamento de IA.

2. Esas obligaciones sobre todo están destinadas a los proveedores y en menor medida a los responsables del despliegue.

3. Los importadores, distribuidores y representantes autorizados también tienen asignadas ciertas responsabilidades propias de las funciones que llevan a cabo.

4. Cualquier operador distinto al proveedor puede pasar a considerarse proveedor cuando realice determinadas actuaciones respecto del sistema de IA.

5. Las autoridades notificantes tiene como función principal supervisar, evaluar y designar a los organismos de evaluación de la conformidad. Notificado un organismo de evaluación de la conformidad para realizar tales evaluaciones conforme al RIA por parte de una autoridad notificante, éste pasará a considerarse organismo notificado. Las autoridades notificantes deberán ostentar los recursos humanos y técnicos necesarios para llevar a cabo estas actividades.

6. El RIA contempla toda una serie de medidas destinadas a las pymes que sean proveedores o responsables del despliegue de sistemas de IA. El objetivo es claro, debido a la complejidad que puede suponer para éstas la adaptación e integración de los requisitos exigidos por esta norma, se obliga a ciertos sujetos a implementar esas medidas de apoyo que ayuden a una correcta adaptación de la norma.

7. El RIA contempla un proceso de notificación de los incidentes grave que pueda sufrir un sistema de IA. Se pretende con ello establecer un procedimiento que mitigue o se reduzca en la medida de los posible los efectos que dichos incidentes hayan causado o puedan causar en un futuro en otros sistemas de IA que presenten las mismas características.

Sujetos y agentes en evaluaciones de conformidad (organismos notificados)

IGNACIO ALAMILLO DOMINGO
Doctor en Derecho¹

I. INTRODUCCIÓN

Los sistemas de IA de alto riesgo se encuentran sujetos obligatoriamente a evaluación de la conformidad conforme a lo establecido en el artículo 16.f) del RIA, que remite al artículo 43 del propio Reglamento, evaluación que, en determinados casos, debe ser realizada por un organismo notificado a tal efecto.

En este capítulo se presenta el régimen jurídico aplicable a dichos organismos notificados para la realización de actividades de evaluación de la conformidad,² abordándose fundamentalmente el contenido de los artículos 29 a 39, y 44 a 46, del RIA.

El RIA define al organismo notificado como un organismo de evaluación de la conformidad notificado conforme al Reglamento y otra legislación de armonización relevante de la Unión (artículo 3.22), donde el organismo de evaluación

1. EID, trust and security legal freak. With a PhD in Law about eIDAS. CISA, CISM, CDPSE.
2. La literatura en esta materia es llamativamente escasa, me permito señalar los estudios más relevantes: De Lucia L., «One and Triune — Mutual Recognition and the Circulation of Goods in the EU», *Review of European Administrative Law*, 13 (3), 2020, pp. 7-35; De Vries S., Kanevskaia O., De Jager R., «Internal Market 3.0: The Old “New Approach” for Harmonising AI Regulation», *European Papers — A Journal on Law and Integration*, 8 (2), 2023, pp. 583-610; Demetzou, K., «Introduction to the conformity assessment under the draft EU AI Act, and how it compares to DPIAs», *Future of Privacy Forum*, August 12, 2022; Galland J.-P., «The difficulties of regulating markets and risks in Europe through notified bodies», *European Journal of Risk Regulation*, 4 (3), 2013, pp. 365-373 y «La difficile construction d’une expertise européenne indépendante», *Revue d’anthropologie des connaissances*, 7-1, 2013; Holder C., Hawes C., Hatzel, J., «The Commission’s proposed Artificial Intelligence Regulation», *Computer and Telecommunications Law Review*, 27 (5), 2021, pp. 130-134 y Lohbeck, D., «Chapter 4 — Notified Bodies and Certification», *CE Marking Handbook*, Newnes, 1998, pp. 53-63; Tricker, R., «2 — Structure of new approach directives», *CE Conformity Marking*, Butterworth-Heinemann, 2000, pp. 46-54; Veale, M., Borgesius, F.Z., «Demystifying the Draft EU Artificial Intelligence Act. Analysing the good, the bad, and the unclear elements of the proposed approach», *Computer Law Review International*, 4/2021.

de la conformidad es un organismo que realiza actividades de evaluación de la conformidad de tercera parte, incluyendo prueba, certificación e inspección (artículo 3.21 del RIA).

Si bien nos encontramos ante un régimen específico establecido por el RIA, el Considerando 46 del RIA aclara que, como parte de la legislación de armonización de la Unión, las normas aplicables a la comercialización, puesta en servicio y uso de sistemas de IA de alto riesgo deben establecerse de forma coherente con «Nuevo marco legislativo para la comercialización de los productos», contenido en el Reglamento (CE) n.º 765/2008 del Parlamento Europeo y del Consejo por el que se establecen los requisitos de acreditación y vigilancia del mercado de los productos, la Decisión n.º 768/2008/CE del Parlamento Europeo y del Consejo sobre un marco común para la comercialización de los productos y el Reglamento (UE) 2019/1020 del Parlamento Europeo y del Consejo sobre la vigilancia del mercado y la conformidad de los productos; por lo que las normas pertinentes de dicho marco serán aplicables de forma supletoria, en los términos descritos en la Comunicación de la Comisión (2022/C 247/01), «Guía azul» sobre la aplicación de la normativa europea relativa a los productos, de 2022.

La evaluación de la conformidad se refiere al proceso de demostrar que se cumplen los requisitos establecidos en la sección 2 del capítulo III del RIA, relativos a los sistemas de IA de alto riesgo, a cuyo análisis detallado nos remitimos, definición que especifica la definición general contenida en el artículo 2.12 del Reglamento (CE) n.º 765/2008; a saber, el proceso por el que se demuestra si se cumplen los requisitos específicos relativos a un producto, un proceso, un servicio, un sistema, una persona o un organismo.

En este sentido, no resulta ocioso recordar que, conforme a lo establecido en el artículo 6.1 del RIA, un sistema de IA se considera de alto riesgo cuando el sistema de IA está destinado a utilizarse como componente de seguridad de un producto, o el sistema de IA es en sí mismo un producto, cubierto por la legislación de armonización de la Unión enumerada en el anexo I, siempre que dicho producto cuyo componente de seguridad es el sistema de IA, o el propio sistema de IA como producto, debe someterse a una evaluación de la conformidad por terceros, con vistas a la comercialización o puesta en servicio de dicho producto con arreglo a la legislación de armonización de la Unión enumerada en el anexo I.

El anexo I incluye tanto legislación de armonización de la Unión basada en el nuevo marco legislativo, 12 normas jurídicas, como a otra legislación de armonización de la Unión, 8 normas jurídicas. En los 12 casos previstos en la sección A del anexo I, el proveedor llevará a cabo la evaluación de la conformidad pertinente con arreglo a lo dispuesto en dichos actos jurídicos.

Asimismo, se consideran sistemas de IA de alto riesgo los referidos en el anexo III del RIA, entre los que se encuentran los sistemas de IA en determinados ámbitos, como la biometría, las infraestructuras críticas, la educación, el empleo, la gestión del personal y el acceso al autoempleo, el acceso y disfrute de determinados servicios esenciales, y beneficios, públicos y privados, el cumplimiento de la ley, el control de fronteras, de la migración y del derecho de asilo, y la administración de justicia y los procesos democráticos.

Aunque la evaluación de conformidad es siempre obligatoria en sistemas de IA de alto riesgo, sólo en algunos casos resulta exigible la intervención de un organismo notificado:

— En el caso de sistemas de IA de alto riesgo a los que se apliquen los actos jurídicos enumerados en la sección A del anexo I del RIA; esto es, máquinas, seguridad de los juguetes, embarcaciones de recreo y motos acuáticas, ascensores y componentes de seguridad para ascensores, aparatos y sistemas de protección para uso en atmósferas potencialmente explosivas, comercialización de equipos radioeléctricos, comercialización de equipos a presión, instalaciones de transporte por cable, equipos de protección individual, aparatos de gas, productos sanitarios y productos sanitarios para diagnóstico *in vitro*.

— En el caso de sistemas de IA para sistemas de identificación biométrica remota, sistemas de IA destinados a ser utilizados para la categorización biométrica, en función de atributos o características sensibles o protegidos basada en la inferencia de dichos atributos o características, o sistemas de IA destinados al reconocimiento de emociones (anexo III.1). Sin embargo, la intervención del organismo notificado sólo es necesaria cuando no existan normas armonizadas y no se disponga de especificaciones comunes; o cuando el proveedor no haya aplicado o sólo haya aplicado parcialmente la norma armonizada; o cuando existan las especificaciones comunes pero el proveedor no las haya aplicado; o cuando una o varias de las normas armonizadas se hayan publicado con una restricción y sólo en la parte de la norma que estaba restringida.

Además, cuando el sistema esté destinado a ser puesto en servicio por autoridades policiales, de inmigración o asilo, así como por instituciones, órganos u organismos de la UE, necesariamente actuará como organismo notificado la autoridad de vigilancia del mercado prevista a tal efecto en el propio RIA.

II. EL PROCEDIMIENTO DE NOTIFICACIÓN

Para actuar como organismo de evaluación de la conformidad de sistemas de IA de alto riesgo resulta necesario cumplir una serie de exigencias y ser notificado como tal frente a la Comisión Europea y a los restantes Estados Miembros por parte de una autoridad notificante nacional, responsable de establecer y aplicar los procedimientos necesarios para la evaluación, designación y notificación de los organismos de evaluación de la conformidad y de su supervisión, sin perjuicio de la posibilidad de que la evaluación y el seguimiento de los citados organismos de evaluación de la conformidad sean realizados por un organismo nacional de acreditación en el sentido del Reglamento (CE) n.º 765/2008 y de conformidad con el mismo.

Sin perjuicio de lo que se acaba de indicar, el artículo 39 del RIA prevé que los organismos de evaluación de la conformidad establecidos conforme al Derecho de un tercer país con el que la Unión haya celebrado un acuerdo puedan ser autorizados para desempeñar las actividades de los organismos notificados con arreglo al RIA, siempre que, conforme al texto acordado durante los diálogos tripartitos, dichos organismos de evaluación de la conformidad de cumplan las exigencias del artículo 31 o que garanticen un nivel de cumplimiento equivalente, lo que quedará a la concreción del acuerdo a celebrar entre la Comisión y cada concreto tercer país. Así ocurre ya en otras legislaciones de armonización, incluyendo Australia, Canadá,

EEUU, Japón, Nueva Zelanda y Suiza. Cabe imaginar que, al menos en aquellos casos en que un sistema de IA de alto riesgo se integre en un producto sujeto a evaluación de conformidad por organismo notificado, esta previsión pueda ser particularmente relevante.

En todo caso, el RIA no agota todos los aspectos precisos para la implementación del procedimiento de notificación, por lo que resulta previsible que la legislación nacional correspondiente concrete y adecue el procedimiento, por ejemplo, a efectos de la solicitud de designación, o de los requisitos lingüísticos, como ha ocurrido en otros reglamentos, como el Reglamento (UE) 2017/745 del Parlamento Europeo y del Consejo, de 5 de abril de 2017, sobre los productos sanitarios. Todas las actuaciones relativas al procedimiento de notificación y sus cambios deberán, por tanto, realizarse conforme a la normativa de procedimiento administrativo común, en los términos que se establezcan en la normativa nacional, como sucede en otros casos. Dado que normalmente nos encontraremos ante solicitantes con la condición de persona jurídica, se tratará de un procedimiento de tramitación íntegramente electrónica. También deberá atribuirse la competencia al órgano correspondiente y concretarse cuestiones como el régimen del silencio administrativo o de recursos en caso de denegación de la designación.

1. LA SOLICITUD DE DESIGNACIÓN

El procedimiento de notificación, regulado en el artículo 29 del RIA, se inicia mediante la presentación, por parte del organismo de evaluación de la conformidad candidato, de una solicitud de designación a la autoridad notificante del Estado miembro en el que esté establecido el citado organismo candidato.

La solicitud de designación debe acompañarse de una descripción de las actividades de evaluación de la conformidad a realizar, del módulo o módulos de evaluación de la conformidad y de los tipos de sistemas de IA para los que el organismo de evaluación de la conformidad se considere competente, así como de un certificado de acreditación, si lo hay, expedido por un organismo nacional de acreditación, que declare que el organismo de evaluación de la conformidad cumple los requisitos establecidos en el artículo 31 del RIA, al que nos referiremos posteriormente.

Además, el organismo candidato podrá añadir cualquier documento válido relacionado con las designaciones existentes del organismo notificado solicitante con arreglo a cualquier otra legislación de armonización de la Unión, lo que especialmente tiene sentido en aquellos casos en que el sistema de IA de alto riesgo se incorpora como componente de un producto sujeto a legislación de armonización. No es el primer caso en que una legislación de armonización de la Unión se refiere a los programas informáticos integrados en un producto, desde luego, significativamente en la Directiva 2006/42/CE, relativa a las máquinas o la Directiva 2014/53/UE, sobre los equipos radioeléctricos.

En este sentido, el RIA prevé que los organismos que hayan sido notificados con arreglo a los actos jurídicos previstos en el anexo I del RIA (legislación de armonización conforme al Nuevo Marco Legislativo u otras leyes de armonización) estarán facultados para controlar la conformidad de los sistemas de IA de alto riesgo con los requisitos establecidos al efecto, pero siempre que la conformidad de dichos

organismos notificados con los requisitos establecidos en el artículo 31, apartados 4, 9 y 10, haya sido evaluada en el contexto del procedimiento de notificación con arreglo a dichos actos jurídicos.

Las actividades de evaluación de la conformidad previstas en la Decisión 768/2008 incluyen la calibración, el ensayo, la certificación y la inspección, mientras que los módulos de evaluación de la conformidad son los siguientes: control interno de la producción (más ensayo supervisado de los productos o más control supervisado de los productos a intervalos aleatorios), examen CE de tipo, conformidad con el tipo basada en el control interno de la producción (más ensayo supervisado de los productos o más control supervisado de los productos a intervalos aleatorios), conformidad con el tipo basada en el aseguramiento de la calidad del proceso de producción, conformidad con el tipo basada en el aseguramiento de la calidad del producto, conformidad con el tipo basada en la verificación del producto, conformidad basada en la verificación por unidad, y conformidad basada en el pleno aseguramiento de la calidad.

El RIA no apunta directamente a ninguno de estos módulos, pero serán aplicables en función de lo que establezca la legislación de armonización correspondiente, en los casos en que la evaluación de la conformidad del sistema de IA de alto riesgo se realice de forma conjunta con la evaluación de la conformidad del producto en que el mismo se integra. En estos casos, además, deberán realizarse necesariamente también las actividades previstas en los epígrafes 4.3 a 4.5 y párrafo quinto del epígrafe 4.6 del anexo VII del RIA.

En los restantes casos, simplemente se aplicará el procedimiento de evaluación de la conformidad con intervención del organismo notificado, contenido en el anexo VII del RIA, referido a la evaluación del sistema de gestión de la calidad y la evaluación de la documentación técnica, a cuyo análisis nos remitimos.

Finalmente, cuando el organismo de evaluación de la conformidad solicitante no haya sido previamente acreditado, facilitará a la autoridad notificante todas las pruebas documentales necesarias para la verificación, el reconocimiento y el seguimiento periódico del cumplimiento de los requisitos establecidos en el artículo 31, algo que normalmente ocurrirá en aquellos casos donde la autoridad notificante haya decidido no encomendar la evaluación y seguimiento a un organismo nacional de acreditación.

En el caso de los organismos notificados designados en virtud de cualquier otra legislación de armonización de la Unión, todos los documentos y certificados vinculados a dichas designaciones podrán utilizarse para respaldar su procedimiento de designación. La orientación general del Consejo ha introducido la obligación de que el organismo notificado actualice la documentación referida en los apartados 2 y 3 del artículo 29, si se producen cambios pertinentes, para que la autoridad responsable de los organismos notificados supervise y verifique el cumplimiento continuo de los requisitos establecidos en el artículo 31.

2. EL PROCEDIMIENTO DE NOTIFICACIÓN

El procedimiento de notificación, como tal, se encuentra regulado en el artículo 30 del RIA, ciertamente alineado con el artículo R23 de la Decisión (UE) n.º 768/2008, que inicialmente ordena que las autoridades notificantes solo podrán notificar

organismos de evaluación de la conformidad que hayan cumplido los requisitos establecidos en el artículo 31.

Precisamente el Considerando 126, en la versión propuesta por la Comisión, ha sido objeto de enmienda tanto por el Parlamento Europeo como por el Consejo, a cuya redacción han incorporado, a las exigencias inicialmente previstas de independencia, competencia y ausencia de conflicto de interés, la necesidad de que los organismos notificados cumplan las exigencias relevantes de ciberseguridad. Además, precisamente la redacción final del Considerando se refiere de forma expresa al uso de la herramienta prevista en el citado artículo R23 de la Decisión (UE) n.º 768/2008.

Por ello, se establece que las autoridades notificantes notificarán a la Comisión y a los demás Estados miembros mediante la herramienta de notificación electrónica desarrollada y gestionada por la Comisión, en que la actualidad es el sistema de información NANDO (New Approach Notified and Designated Organisations), accesible en <https://webgate.ec.europa.eu/single-market-compliance-space/#/notified-bodies> A propuesta del Parlamento Europeo, y en cierto modo en línea con la orientación general del Consejo, se ha aclarado que se deberá notificar cada organismo de conformidad que haya satisfecho las exigencias del artículo 31.

Con respecto al contenido de la notificación, la propuesta original de la Comisión Europea se limitaba a la información pormenorizada de las actividades de evaluación de la conformidad, el módulo o los módulos de evaluación de la conformidad y las tecnologías de inteligencia artificial afectadas. Sin embargo, tanto el Parlamento Europeo como el Consejo propusieron incluir también la correspondiente declaración de competencia del organismo de evaluación de la conformidad, que podrá ser, en su caso, el resultado del procedimiento de acreditación. Y, a propuesta del Consejo, se ha añadido también la previsión de que, cuando una notificación no se base en un certificado de acreditación, la autoridad notificante deberá facilitar pruebas documentales que acrediten la competencia del organismo de evaluación de la conformidad y las disposiciones adoptadas para garantizar que dicho organismo será objeto de un seguimiento periódico y seguirá cumpliendo los requisitos establecidos en el artículo 33. Se trata de una exigencia análoga a la que se contiene en otras normas reglamentarias aplicables a organismos notificados, como el ya citado Reglamento (UE) 2017/745.

Realizada la notificación a través del sistema de información NANDO, el organismo de evaluación de la conformidad en cuestión únicamente podrá realizar las actividades de un organismo notificado si la Comisión o los demás Estados miembros no formulan ninguna objeción en el plazo de dos semanas tras la validación de una notificación que incluya un certificado de acreditación, plazo que será de dos meses cuando se aporten pruebas documentales que acrediten la competencia del organismo de evaluación de la conformidad. Se trata de régimen diferente del inicialmente propuesto por la Comisión Europea, que era de un mes con carácter general y que no se encontraba, por cierto, alineado con el artículo R23 de la Decisión (UE) n.º 768/2008.

A propuesta del Parlamento Europeo, se ha previsto, en caso de que se formulen objeciones, la obligación de la Comisión de consultar sin demora a los Estados miembros pertinentes y al organismo de evaluación de la conformidad; debiendo

decidir la Comisión si la autorización está justificada o no, decisión que será dirigida al Estado miembro de que se trate y al organismo de evaluación de la conformidad pertinente.

Finalmente, a propuesta del Consejo, se ha trasladado al artículo 36 del RIA la previsión, contenida en la propuesta original de la Comisión, de que las autoridades notificantes notificaran a la Comisión y a los demás Estados miembros todo cambio posterior de la notificación que resultare pertinente.

3. IDENTIFICACIÓN Y PUBLICIDAD DE LOS ORGANISMOS NOTIFICADOS

El artículo 35 del RIA, que no ha sufrido modificaciones durante su tramitación, se ocupa de la identificación de cada organismo notificado, a quien se asignará un número único para todas sus actividades de evaluación de la conformidad, independientemente de que haya sido notificado con arreglo a varios actos de la Unión. Dicho número se gestiona en el ya mencionado sistema de información NANDO.

Asimismo, el mismo artículo ordena que la Comisión hará pública la lista de organismos notificados con arreglo al Reglamento, junto con los números de identificación que les hayan sido asignados y las actividades para las que hayan sido notificados, debiendo asegurarse de que la lista se mantenga actualizada, todo lo cual se gestiona en el sistema de información NANDO, del que cabe esperar sea ampliado con la nueva legislación armonizada.

4. CAMBIOS EN LA NOTIFICACIÓN

El artículo 36 del RIA se ocupa del tratamiento de los cambios en las notificaciones realizadas, y es uno de los referidos a los organismos notificados que han sufrido mayores modificaciones desde la propuesta original de la Comisión.

En primer lugar, y como ya se dijo, se ha incorporado al inicio del artículo un primer epígrafe con la obligación de que las autoridades notificantes notifiquen a la Comisión y a los demás Estados Miembros todo cambio posterior de la notificación que resultare pertinente, especificándose que dicha notificación de cambios deberá realizarse a través de la herramienta electrónica indicada en el artículo 30.2 del RIA; esto es, el sistema de información NANDO.

En segundo lugar, y de nuevo a propuesta del Consejo, el epígrafe 2 del artículo 36 del RIA aclara que las ampliaciones en el alcance de una notificación ya realizada implicarán la realización de un nuevo procedimiento de notificación, incluida la correspondiente solicitud, previsión lógica atendiendo a que dicha ampliación sin control previo permitiría fraudes de ley. Para otros cambios en la notificación, resultarán aplicables los procedimientos previstos en los restantes apartados del propio artículo, y que son el cese de actividad por parte del organismo notificado, o la limitación, suspensión o retirada de la notificación, en caso de incumplimiento de las exigencias aplicables al organismo notificado.

El régimen del cese, contenido en el epígrafe 3 del artículo 36 del RIA, introducido a propuesta del Consejo, resulta aplicable cuando un organismo notificado decida poner fin a sus actividades de evaluación de la conformidad, en

cuyo caso deberá informar de ello a la autoridad notificante y a los proveedores afectados lo antes posible y, en el caso de un cese previsto, un año antes del cese de sus actividades. Cuando el organismo notificado haya cesado su actividad, la autoridad notificante retirará la designación.

Por lo que se refiere al régimen de limitación, suspensión o retirada de la notificación, la propuesta de la Comisión contenía dos epígrafes, que han sido objeto de una significativa ampliación durante la tramitación del reglamento, con especial protagonismo del Consejo, en particular en relación con el detalle de las medidas oportunas que deben adoptar las autoridades notificantes para asegurarse de que los expedientes de dicho organismo notificado sean asumidos por otro organismo notificado o se pongan a disposición de las autoridades notificantes responsables cuando estas los soliciten. El texto originalmente propuesto por la Comisión resulta muy similar al que se encuentran en otras legislaciones de armonización, a pesar de lo cual en este caso se considerado oportuno incrementar el nivel de detalle.

En este sentido, conforme al epígrafe 4 del artículo 36 del RIA, cuando una autoridad notificante tenga motivos suficientes para considerar que un organismo notificado ha dejado de cumplir los requisitos establecidos en el artículo 31 o que está incumpliendo sus obligaciones, deberá investigar sin demora el asunto con la máxima diligencia, informando al organismo notificado de que se trate acerca de las objeciones planteadas y dándole la posibilidad de dar a conocer sus puntos de vista, procedimiento administrativo que deberá realizarse con plena sujeción a la normativa de procedimiento administrativo común y con todas las garantías legalmente previstas.

En cualquier caso, si la autoridad notificante llegara a la conclusión de que el organismo notificado ha dejado de cumplir los requisitos establecidos en el artículo 31 o de que está incumpliendo sus obligaciones, deberá necesariamente resolver restringiendo, suspendiendo o retirando la notificación, según proceda, en función de la gravedad del incumplimiento de dichos requisitos o de dichas obligaciones, debiendo informar inmediatamente de ello a la Comisión y a los demás Estados miembros, obviamente a través del sistema de información NANDO.

Como se ha indicado ya, todas estas actuaciones deberán realizarse conforme a la normativa de procedimiento administrativo común, pero debiendo considerar las particularidades previstas en los restantes epígrafes del artículo 36, que obviamente resultan de aplicación directa.

En primer lugar, el nuevo epígrafe 5 del artículo 36 del RIA ordena al organismo notificado cuya designación haya sido suspendida, restringida o retirada total o parcialmente, informar de ello a los fabricantes afectados a más tardar en un plazo de 10 días.

En segundo término, conforme a los nuevos epígrafes 6 y 7 del artículo 36 del RIA, en caso de limitación, suspensión o retirada de una notificación, la autoridad notificante deberá realizar las siguientes actuaciones:

— Adoptar las medidas adecuadas para garantizar que se conserven los archivos del organismo notificado de que se trate y ponerlos a disposición de

las autoridades notificantes de otros Estados miembros y de las autoridades de vigilancia del mercado que lo soliciten.

— Evaluar el impacto en los certificados expedidos por el organismo notificado.

— Presentar un informe sobre sus conclusiones a la Comisión y a los demás Estados miembros en un plazo de tres meses a partir de la notificación de los cambios introducidos.

— Exigir al organismo notificado que suspenda o retire, en un plazo razonable determinado por la autoridad, todos los certificados que se hayan expedido indebidamente para garantizar la conformidad de los sistemas de IA en el mercado.

— Informar a la Comisión y a los Estados miembros de los certificados cuya suspensión o retirada haya exigido.

— Facilitar a las autoridades nacionales competentes del Estado miembro en el que el prestador tenga su domicilio social toda la información pertinente sobre los certificados para los que haya exigido la suspensión o retirada. Dicha autoridad competente adoptará las medidas adecuadas, cuando sea necesario, para evitar un riesgo potencial para la salud, la seguridad o los derechos fundamentales.

Finalmente, el último párrafo del nuevo epígrafe 9 del artículo 36 del RIA ordena a la autoridad nacional competente o al organismo notificado que asuma las funciones del organismo notificado afectado por el cambio de notificación informar inmediatamente de ello a la Comisión, a los demás Estados miembros y a los demás organismos notificados.

5. CUESTIONAMIENTO DE LA COMPETENCIA DE LOS ORGANISMOS NOTIFICADOS

El artículo 37 del RIA se ocupa de la impugnación de la competencia de los organismos notificados, ordenando en su epígrafe primero, en redacción propuesta por el Parlamento Europeo, que la Comisión investigará, cuando sea necesario, todos los casos en los que existan motivos para dudar de la competencia de un organismo notificado o del cumplimiento continuado por parte de un organismo notificado de los requisitos establecidos en el artículo 31 y de sus correspondientes responsabilidades.

Para ello, la autoridad notificante facilitará a la Comisión, previa solicitud, toda la información pertinente relativa a la notificación o el mantenimiento de la competencia del organismo notificado de que se trate (epígrafe 2), la cual deberá ser tratada de forma confidencial, en atención a su sensibilidad (epígrafe 3).

Como resultado de esta investigación, cuando la Comisión compruebe que un organismo notificado no cumple o ha dejado de cumplir los requisitos para su notificación, informará de ello al Estado miembro notificante y le pedirá que adopte las medidas correctoras necesarias, incluida la suspensión o retirada de la notificación si es necesario, pudiendo hacerlo la propia Comisión, en caso de inactividad del Estado miembro, como se determine en un acto de ejecución. Así se dispone en el epígrafe 4, que ha sido objeto de una importante modificación con respecto a la propuesta inicial, y que ha adoptado la posición del Parlamento.

III. LA ACTUACIÓN DE LOS ORGANISMOS NOTIFICADOS

1. LOS REQUISITOS APLICABLES A LOS ORGANISMOS NOTIFICADOS

El importante artículo 31 del RIA detalla el elenco de requisitos exigibles a los organismos de evaluación de la conformidad que opten a la notificación, o a sus filiales o subcontratistas (artículo 33), que incluyen los siguientes:

— Encontrarse establecidos conforme a la legislación nacional y disponer de personalidad jurídica (epígrafe 1, redactado a propuesta del Consejo), lo que debe entenderse si perjuicio del reconocimiento de organismos establecidos en terceros Estados.

— Satisfacer los requisitos organizativos, de gestión de calidad, de recursos y de procesos que sean necesarios para el cumplimiento de sus tareas, así como los requisitos de ciberseguridad adecuados (epígrafe 2). La adición de los requisitos de ciberseguridad procede del Consejo.

— Disponer de una estructura organizativa, de asignación de responsabilidades, líneas jerárquicas y un funcionamiento que garantice la confianza en el desempeño y en los resultados de las actividades de evaluación de la conformidad que lleven a cabo (epígrafe 3).

— Ser independientes del proveedor de un sistema de IA de alto riesgo en relación con el cual realicen actividades de evaluación de la conformidad, y de cualquier otro operador que tenga un interés económico en el sistema de IA de alto riesgo que se evalúe, así como de cualquier competidor del proveedor; lo que no impedirá el uso de sistemas de IA evaluados que sean necesarios para el funcionamiento del organismo de evaluación de la conformidad o el uso de dichos sistemas con fines personales (epígrafe 4, inciso final añadido a propuesta del Consejo).

— No intervenir directamente en el diseño, el desarrollo, la comercialización o el uso de sistemas de IA de alto riesgo, ni representará a las partes que participen en dichas actividades, ni en ninguna actividad que pueda entrar en conflicto con su independencia de juicio o integridad en relación con las actividades de evaluación de la conformidad para las que se les notifique, lo que se aplicará, en particular, a los servicios de consultoría (nuevo epígrafe 5, añadido a propuesta del Parlamento).

— Estar organizados y operados de manera que se salvguarde la independencia, objetividad e imparcialidad de sus actividades, para lo que documentarán y aplicarán una estructura y procedimientos para salvaguardar la imparcialidad y promover y aplicar los principios de imparcialidad en todas sus actividades de organización, personal y evaluación (epígrafe 6).

— Disponer de procedimientos documentados que garanticen que su personal, comités, filiales, subcontratistas y cualquier organismo asociado o personal de organismos externos respeten la confidencialidad de la información que, de conformidad con el artículo 78 del RIA, obren en su poder durante la realización de las actividades de evaluación de la conformidad, excepto cuando la ley exija su divulgación. Por ello, se establece que el personal de los organismos notificados estará obligado a guardar secreto profesional con respecto a toda la información obtenida en el ejercicio de sus funciones, excepto en relación con las autoridades notificantes del Estado miembro en el que se desarrollen sus actividades (epígrafe 7, cuya única

modificación ha sido, a propuesta del Consejo, la adición de la referencia al artículo 78 del RIA).

— Disponer de procedimientos para la realización de actividades que tengan debidamente en cuenta el tamaño de una empresa, el sector en el que opera, su estructura y el grado de complejidad del sistema de IA de que se trate (epígrafe 8).

— Suscribir un seguro de responsabilidad civil adecuado para sus actividades de evaluación de la conformidad, a menos que la responsabilidad sea asumida por el Estado miembro en el que estén establecidos de conformidad con la legislación nacional o que dicho Estado miembro sea directamente responsable de la evaluación de la conformidad (epígrafe 9, ligeramente modificado a propuesta del Consejo para especificar que el Estado concernido será el de establecimiento del organismo notificado).

— Ser capaces de llevar a cabo todas las tareas que les incumben en virtud del presente Reglamento con el máximo grado de integridad profesional y la competencia requerida en el ámbito específico, tanto si dichas tareas son realizadas por los propios organismos notificados como en su nombre y bajo su responsabilidad (epígrafe 10).

— Tener las competencias internas suficientes para poder evaluar eficazmente las tareas realizadas por partes externas en su nombre, para lo cual el organismo notificado dispondrá permanentemente de suficiente personal administrativo, técnico, jurídico y científico que posea experiencia y conocimientos relativos a los tipos pertinentes de sistemas de IA, datos y computación de datos, así como a los requisitos legalmente establecidos (epígrafe 11, en redacción a propuesta del Consejo).

— Participar en las actividades de coordinación a que se refiere el artículo 38 y participar en las organizaciones europeas de normalización, mediante su intervención directa o representación, o asegurándose de conocer y estar al día de las normas pertinentes (epígrafe 12).

— Cuando subcontrate tareas específicas relacionadas con la evaluación de la conformidad o recurra a una filial, asegurarse de que el subcontratista o la filial cumplan los requisitos establecidos en el artículo 31 e informar a la autoridad notificante en consecuencia (artículo 33.1).

— Asumir la plena responsabilidad de las tareas realizadas por los subcontratistas o filiales (artículo 33.2).

— Subcontratar o llevar a cabo actividades de evaluación de la conformidad mediante una filial sólo con el acuerdo del proveedor, y poner a disposición del público una lista de sus filiales (artículo 33.3).

— Mantener los documentos pertinentes relativos a la evaluación de las cualificaciones del subcontratista o de la filial y de los trabajos realizados por ellos a disposición de la autoridad notificante durante un período de cinco años a partir de la fecha de finalización de la actividad de subcontratación (artículo 33.4, en redacción a propuesta del Consejo).

— Encontrarse establecidos conforme a la legislación nacional y disponer de personalidad jurídica (epígrafe 1, redactado a propuesta del Consejo), lo que debe entenderse si perjuicio del reconocimiento de organismos establecidos en terceros Estados.

— Satisfacer los requisitos organizativos, de gestión de calidad, de recursos y de procesos que sean necesarios para el cumplimiento de sus tareas, así como los requisitos de ciberseguridad adecuados (epígrafe 2). La adición de los requisitos de ciberseguridad procede del Consejo.

— Disponer de una estructura organizativa, de asignación de responsabilidades, líneas jerárquicas y un funcionamiento que garantice la confianza en el desempeño y en los resultados de las actividades de evaluación de la conformidad que lleven a cabo (epígrafe 3).

— Ser independientes del proveedor de un sistema de IA de alto riesgo en relación con el cual realicen actividades de evaluación de la conformidad, y de cualquier otro operador que tenga un interés económico en el sistema de IA de alto riesgo que se evalúe, así como de cualquier competidor del proveedor; lo que no impedirá el uso de sistemas de IA evaluados que sean necesarios para el funcionamiento del organismo de evaluación de la conformidad o el uso de dichos sistemas con fines personales (epígrafe 4, inciso final añadido a propuesta del Consejo).

— No intervenir directamente en el diseño, el desarrollo, la comercialización o el uso de sistemas de IA de alto riesgo, ni representará a las partes que participen en dichas actividades, ni en ninguna actividad que pueda entrar en conflicto con su independencia de juicio o integridad en relación con las actividades de evaluación de la conformidad para las que se les notifique, lo que se aplicará, en particular, a los servicios de consultoría (nuevo epígrafe 5, añadido a propuesta del Parlamento).

— Estar organizados y operados de manera que se salvaguarde la independencia, objetividad e imparcialidad de sus actividades, para lo que documentarán y aplicarán una estructura y procedimientos para salvaguardar la imparcialidad y promover y aplicar los principios de imparcialidad en todas sus actividades de organización, personal y evaluación (epígrafe 6).

— Disponer de procedimientos documentados que garanticen que su personal, comités, filiales, subcontratistas y cualquier organismo asociado o personal de organismos externos respeten la confidencialidad de la información que, de conformidad con el artículo 78 del RIA, obren en su poder durante la realización de las actividades de evaluación de la conformidad, excepto cuando la ley exija su divulgación. Por ello, se establece que el personal de los organismos notificados estará obligado a guardar secreto profesional con respecto a toda la información obtenida en el ejercicio de sus funciones, excepto en relación con las autoridades notificantes del Estado miembro en el que se desarrollen sus actividades (epígrafe 7, cuya única modificación ha sido, a propuesta del Consejo, la adición de la referencia al artículo 78 del RIA).

— Disponer de procedimientos para la realización de actividades que tengan debidamente en cuenta el tamaño de una empresa, el sector en el que opera, su estructura y el grado de complejidad del sistema de IA de que se trate (epígrafe 8).

— Suscribir un seguro de responsabilidad civil adecuado para sus actividades de evaluación de la conformidad, a menos que la responsabilidad sea asumida por el Estado miembro en el que estén establecidos de conformidad con la legislación nacional o que dicho Estado miembro sea directamente responsable de la evaluación de la conformidad (epígrafe 9, ligeramente modificado a propuesta del Consejo)

para especificar que el Estado concernido será el de establecimiento del organismo notificado).

— Ser capaces de llevar a cabo todas las tareas que les incumben en virtud del presente Reglamento con el máximo grado de integridad profesional y la competencia requerida en el ámbito específico, tanto si dichas tareas son realizadas por los propios organismos notificados como en su nombre y bajo su responsabilidad (epígrafe 10).

— Tener las competencias internas suficientes para poder evaluar eficazmente las tareas realizadas por partes externas en su nombre, para lo cual el organismo notificado dispondrá permanentemente de suficiente personal administrativo, técnico, jurídico y científico que posea experiencia y conocimientos relativos a los tipos pertinentes de sistemas de IA, datos y computación de datos, así como a los requisitos legalmente establecidos (epígrafe 11, en redacción a propuesta del Consejo).

— Participar en las actividades de coordinación a que se refiere el artículo 38 y participar en las organizaciones europeas de normalización, mediante su intervención directa o representación, o asegurándose de conocer y estar al día de las normas pertinentes (epígrafe 12).

— Cuando subcontrate tareas específicas relacionadas con la evaluación de la conformidad o recurra a una filial, asegurarse de que el subcontratista o la filial cumplan los requisitos establecidos en el artículo 31 e informar a la autoridad notificante en consecuencia (artículo 33.1).

— Asumir la plena responsabilidad de las tareas realizadas por los subcontratistas o filiales (artículo 33.2).

— Subcontratar o llevar a cabo actividades de evaluación de la conformidad mediante una filial sólo con el acuerdo del proveedor, y poner a disposición del público una lista de sus filiales (artículo 33.3).

— Mantener los documentos pertinentes relativos a la evaluación de las cualificaciones del subcontratista o de la filial y de los trabajos realizados por ellos a disposición de la autoridad notificante durante un período de cinco años a partir de la fecha de finalización de la actividad de subcontratación (artículo 33.4, en redacción a propuesta del Consejo).

— Encontrarse establecidos conforme a la legislación nacional y disponer de personalidad jurídica (epígrafe 1, redactado a propuesta del Consejo), lo que debe entenderse si perjuicio del reconocimiento de organismos establecidos en terceros Estados.

— Satisfacer los requisitos organizativos, de gestión de calidad, de recursos y de procesos que sean necesarios para el cumplimiento de sus tareas, así como los requisitos de ciberseguridad adecuados (epígrafe 2). La adición de los requisitos de ciberseguridad procede del Consejo.

— Disponer de una estructura organizativa, de asignación de responsabilidades, líneas jerárquicas y un funcionamiento que garantice la confianza en el desempeño y en los resultados de las actividades de evaluación de la conformidad que lleven a cabo (epígrafe 3).

— Ser independientes del proveedor de un sistema de IA de alto riesgo en relación con el cual realicen actividades de evaluación de la conformidad, y de cualquier otro operador que tenga un interés económico en el sistema de IA de alto riesgo que se

evalúe, así como de cualquier competidor del proveedor; lo que no impedirá el uso de sistemas de IA evaluados que sean necesarios para el funcionamiento del organismo de evaluación de la conformidad o el uso de dichos sistemas con fines personales (epígrafe 4, inciso final añadido a propuesta del Consejo).

— No intervenir directamente en el diseño, el desarrollo, la comercialización o el uso de sistemas de IA de alto riesgo, ni representará a las partes que participen en dichas actividades, ni en ninguna actividad que pueda entrar en conflicto con su independencia de juicio o integridad en relación con las actividades de evaluación de la conformidad para las que se les notifique, lo que se aplicará, en particular, a los servicios de consultoría (nuevo epígrafe 5, añadido a propuesta del Parlamento).

— Estar organizados y operados de manera que se salvaguarde la independencia, objetividad e imparcialidad de sus actividades, para lo que documentarán y aplicarán una estructura y procedimientos para salvaguardar la imparcialidad y promover y aplicar los principios de imparcialidad en todas sus actividades de organización, personal y evaluación (epígrafe 6).

— Disponer de procedimientos documentados que garanticen que su personal, comités, filiales, subcontratistas y cualquier organismo asociado o personal de organismos externos respeten la confidencialidad de la información que, de conformidad con el artículo 78 del RIA, obren en su poder durante la realización de las actividades de evaluación de la conformidad, excepto cuando la ley exija su divulgación. Por ello, se establece que el personal de los organismos notificados estará obligado a guardar secreto profesional con respecto a toda la información obtenida en el ejercicio de sus funciones, excepto en relación con las autoridades notificantes del Estado miembro en el que se desarrollen sus actividades (epígrafe 7, cuya única modificación ha sido, a propuesta del Consejo, la adición de la referencia al artículo 78 del RIA).

— Disponer de procedimientos para la realización de actividades que tengan debidamente en cuenta el tamaño de una empresa, el sector en el que opera, su estructura y el grado de complejidad del sistema de IA de que se trate (epígrafe 8).

— Suscribir un seguro de responsabilidad civil adecuado para sus actividades de evaluación de la conformidad, a menos que la responsabilidad sea asumida por el Estado miembro en el que estén establecidos de conformidad con la legislación nacional o que dicho Estado miembro sea directamente responsable de la evaluación de la conformidad (epígrafe 9, ligeramente modificado a propuesta del Consejo para especificar que el Estado concernido será el de establecimiento del organismo notificado).

— Ser capaces de llevar a cabo todas las tareas que les incumben en virtud del presente Reglamento con el máximo grado de integridad profesional y la competencia requerida en el ámbito específico, tanto si dichas tareas son realizadas por los propios organismos notificados como en su nombre y bajo su responsabilidad (epígrafe 10).

— Tener las competencias internas suficientes para poder evaluar eficazmente las tareas realizadas por partes externas en su nombre, para lo cual el organismo notificado dispondrá permanentemente de suficiente personal administrativo, técnico, jurídico y científico que posea experiencia y conocimientos relativos a los tipos pertinentes de sistemas de IA, datos y computación de datos, así como a los requisitos legalmente establecidos (epígrafe 11, en redacción a propuesta del Consejo).

— Participar en las actividades de coordinación a que se refiere el artículo 38 y participar en las organizaciones europeas de normalización, mediante su intervención directa o representación, o asegurándose de conocer y estar al día de las normas pertinentes (epígrafe 12).

— Cuando subcontrate tareas específicas relacionadas con la evaluación de la conformidad o recurra a una filial, asegurarse de que el subcontratista o la filial cumplan los requisitos establecidos en el artículo 31 e informar a la autoridad notificante en consecuencia (artículo 33.1).

— Asumir la plena responsabilidad de las tareas realizadas por los subcontratistas o filiales (artículo 33.2).

— Subcontratar o llevar a cabo actividades de evaluación de la conformidad mediante una filial sólo con el acuerdo del proveedor, y poner a disposición del público una lista de sus filiales (artículo 33.3).

— Mantener los documentos pertinentes relativos a la evaluación de las cualificaciones del subcontratista o de la filial y de los trabajos realizados por ellos a disposición de la autoridad notificante durante un período de cinco años a partir de la fecha de finalización de la actividad de subcontratación (artículo 33.4, en redacción a propuesta del Consejo).

— Encontrarse establecidos conforme a la legislación nacional y disponer de personalidad jurídica (epígrafe 1, redactado a propuesta del Consejo), lo que debe entenderse si perjuicio del reconocimiento de organismos establecidos en terceros Estados.

— Satisfacer los requisitos organizativos, de gestión de calidad, de recursos y de procesos que sean necesarios para el cumplimiento de sus tareas, así como los requisitos de ciberseguridad adecuados (epígrafe 2). La adición de los requisitos de ciberseguridad procede del Consejo.

— Disponer de una estructura organizativa, de asignación de responsabilidades, líneas jerárquicas y un funcionamiento que garantice la confianza en el desempeño y en los resultados de las actividades de evaluación de la conformidad que lleven a cabo (epígrafe 3).

— Ser independientes del proveedor de un sistema de IA de alto riesgo en relación con el cual realicen actividades de evaluación de la conformidad, y de cualquier otro operador que tenga un interés económico en el sistema de IA de alto riesgo que se evalúe, así como de cualquier competidor del proveedor; lo que no impedirá el uso de sistemas de IA evaluados que sean necesarios para el funcionamiento del organismo de evaluación de la conformidad o el uso de dichos sistemas con fines personales (epígrafe 4, inciso final añadido a propuesta del Consejo).

— No intervenir directamente en el diseño, el desarrollo, la comercialización o el uso de sistemas de IA de alto riesgo, ni representará a las partes que participen en dichas actividades, ni en ninguna actividad que pueda entrar en conflicto con su independencia de juicio o integridad en relación con las actividades de evaluación de la conformidad para las que se les notifique, lo que se aplicará, en particular, a los servicios de consultoría (nuevo epígrafe 5, añadido a propuesta del Parlamento).

— Estar organizados y operados de manera que se salvaguarde la independencia, objetividad e imparcialidad de sus actividades, para lo que documentarán y aplicarán

una estructura y procedimientos para salvaguardar la imparcialidad y promover y aplicar los principios de imparcialidad en todas sus actividades de organización, personal y evaluación (epígrafe 6).

— Disponer de procedimientos documentados que garanticen que su personal, comités, filiales, subcontratistas y cualquier organismo asociado o personal de organismos externos respeten la confidencialidad de la información que, de conformidad con el artículo 78 del RIA, obren en su poder durante la realización de las actividades de evaluación de la conformidad, excepto cuando la ley exija su divulgación. Por ello, se establece que el personal de los organismos notificados estará obligado a guardar secreto profesional con respecto a toda la información obtenida en el ejercicio de sus funciones, excepto en relación con las autoridades notificantes del Estado miembro en el que se desarrollen sus actividades (epígrafe 7, cuya única modificación ha sido, a propuesta del Consejo, la adición de la referencia al artículo 78 del RIA).

— Disponer de procedimientos para la realización de actividades que tengan debidamente en cuenta el tamaño de una empresa, el sector en el que opera, su estructura y el grado de complejidad del sistema de IA de que se trate (epígrafe 8).

— Suscribir un seguro de responsabilidad civil adecuado para sus actividades de evaluación de la conformidad, a menos que la responsabilidad sea asumida por el Estado miembro en el que estén establecidos de conformidad con la legislación nacional o que dicho Estado miembro sea directamente responsable de la evaluación de la conformidad (epígrafe 9, ligeramente modificado a propuesta del Consejo para especificar que el Estado concernido será el de establecimiento del organismo notificado).

— Ser capaces de llevar a cabo todas las tareas que les incumben en virtud del presente Reglamento con el máximo grado de integridad profesional y la competencia requerida en el ámbito específico, tanto si dichas tareas son realizadas por los propios organismos notificados como en su nombre y bajo su responsabilidad (epígrafe 10).

— Tener las competencias internas suficientes para poder evaluar eficazmente las tareas realizadas por partes externas en su nombre, para lo cual el organismo notificado dispondrá permanentemente de suficiente personal administrativo, técnico, jurídico y científico que posea experiencia y conocimientos relativos a los tipos pertinentes de sistemas de IA, datos y computación de datos, así como a los requisitos legalmente establecidos (epígrafe 11, en redacción a propuesta del Consejo).

— Participar en las actividades de coordinación a que se refiere el artículo 38 y participar en las organizaciones europeas de normalización, mediante su intervención directa o representación, o asegurándose de conocer y estar al día de las normas pertinentes (epígrafe 12).

— Cuando subcontrate tareas específicas relacionadas con la evaluación de la conformidad o recurra a una filial, asegurarse de que el subcontratista o la filial cumplan los requisitos establecidos en el artículo 31 e informar a la autoridad notificante en consecuencia (artículo 33.1).

— Asumir la plena responsabilidad de las tareas realizadas por los subcontratistas o filiales (artículo 33.2).

— Subcontratar o llevar a cabo actividades de evaluación de la conformidad mediante una filial sólo con el acuerdo del proveedor, y poner a disposición del público una lista de sus filiales (artículo 33.3).

— Mantener los documentos pertinentes relativos a la evaluación de las cualificaciones del subcontratista o de la filial y de los trabajos realizados por ellos a disposición de la autoridad notificante durante un período de cinco años a partir de la fecha de finalización de la actividad de subcontratación (artículo 33.4, en redacción a propuesta del Consejo).

Por su parte, el nuevo artículo 32, adicionado a propuesta del Consejo, establece que, cuando un organismo de evaluación de la conformidad demuestre su conformidad con los criterios establecidos en las normas armonizadas pertinentes o en partes de las mismas cuyas referencias se hayan publicado en el Diario Oficial de la Unión Europea, se presumirá que cumple los requisitos establecidos en el artículo 31 en la medida en que las normas armonizadas aplicables cubran dichos requisitos. Se trata de una norma dirigida a facilitar la acreditación de los requisitos por parte de los organismos, y que puede facilitar el acceso a esta actividad en igualdad de condiciones por parte de los candidatos, al tiempo que se incentiva la adopción de normas de previsiblemente elevada calidad.

2. OBLIGACIONES OPERACIONALES DE LOS ORGANISMOS NOTIFICADOS. COORDINACIÓN POR LA COMISIÓN

El artículo 34 del RIA, adicionado a propuesta del Consejo, detalla las obligaciones operacionales de los organismos notificados.

Más allá de la obviedad de que los organismos notificados verificarán la conformidad de los sistemas de IA de alto riesgo de conformidad con los procedimientos de evaluación de la conformidad a que se refiere el artículo 43 (epígrafe 1), destaca la previsión de que estos llevarán a cabo sus actividades evitando cargas innecesarias para los prestadores y teniendo debidamente en cuenta el tamaño de la empresa, el sector en el que opera, su estructura y el grado de complejidad del sistema de IA de alto riesgo de que se trate, pero respetando, no obstante, el grado de rigor y el nivel de protección requeridos para la conformidad del sistema de IA de alto riesgo con los requisitos del RIA. Uno de los objetivos políticos de mayor calado se ha elevado a la categoría de norma jurídica, al prever que se deberá prestar especial atención a la reducción al mínimo de las cargas administrativas y los costes de cumplimiento para las microempresas y las pequeñas empresas, tal como se definen en la Recomendación 2003/361/CE de la Comisión (epígrafe 2).

Finalmente, el epígrafe 3 del artículo 34 recoge la previsión originalmente contenida en el artículo 33, en cuya virtud los organismos notificados pondrán a disposición de la autoridad notificante y le presentarán, previa solicitud, toda la documentación pertinente, incluida la documentación de los proveedores, para que dicha autoridad pueda llevar a cabo sus actividades de evaluación, designación, notificación y seguimiento y para facilitar la evaluación.

Destaca, también, la función de coordinación que se atribuye a la Comisión Europea en relación con los organismos notificados en el artículo 38 del RIA. Y es que la Comisión velará por que, en lo que respecta a los sistemas de IA de alto riesgo, se establezcan y gestionen adecuadamente la coordinación y la cooperación

adecuadas entre los organismos notificados que participan en los procedimientos de evaluación de la conformidad, en forma de un grupo sectorial de organismos notificados (epígrafe 1, en redacción propuesta por el Consejo).

A tal efecto, se prevé la obligación de la autoridad notificante de velar por que los organismos notificados por ellos participen en los trabajos de dicho grupo, directamente o por medio de representantes designados (epígrafe 2).

Y no sorprende que, a propuesta del Consejo, se haya incorporado un nuevo epígrafe 3 al artículo 38, obligando a la Comisión disponer el intercambio de conocimientos y mejores prácticas entre las autoridades notificantes de los Estados miembros, previsión que resulta ciertamente conveniente.

3. EXPEDICIÓN Y VALIDEZ DE CERTIFICADOS

Como resultado satisfactorio de las actividades pertinentes de evaluación de la conformidad, el organismo notificado deberá expedir un certificado con los contenidos del anexo VII del RIA usando un lenguaje que pueda ser fácilmente comprensible para las autoridades relevantes del Estado en que se encuentre establecido el organismo (artículo 44.1, en redacción propuesta por el Consejo).

Respecto a su plazo de validez, tras los diálogos tripartitos, se ha establecido en un máximo de cinco años para los sistemas de IA del Anexo I, y de cuatro años para los sistemas de IA del Anexo III; validez que podrá ser extendida por periodos iguales, con base evaluaciones adicionales (artículo 44.2, en redacción propuesta por el Consejo).

Por supuesto, el resultado de las actividades de evaluación de la conformidad puede no resultar satisfactorio, de considerar el organismo notificado que el sistema de IA objeto de evaluación no cumple los requisitos legalmente establecidos, lo que implicará la negativa del organismo a expedir el correspondiente certificado, o a expedirlo con determinadas restricciones. Frente a dichas decisiones, el epígrafe 3 del artículo 44, segundo párrafo, del RIA, en redacción consensuada entre el Parlamento y el Consejo, garantiza la existencia de un procedimiento de recurso. Nótese, en relación con este artículo, que tanto la propuesta de la Comisión como el mandato del Parlamento exigían un interés legítimo para interponer dicha reclamación, exigencia que ha desaparecido en la redacción final.

Asimismo, cuando un organismo notificado compruebe que un sistema de IA ha dejado de cumplir los requisitos legalmente establecidos, suspenderá o retirará el certificado expedido o le impondrá, teniendo en cuenta el principio de proporcionalidad, a menos que el proveedor del sistema adopte medidas correctoras adecuadas en un plazo adecuado fijado por el organismo notificado, teniendo en cuenta el principio de proporcionalidad, decisión que deberá motivar (artículo 44.3, en redacción original de la Comisión, avalada por el Consejo).

Como hemos visto anteriormente, el organismo notificado puede verse afectado por cambios, lo que eventualmente afectará a la validez de los certificados expedidos.

En caso de cese en la actividad del organismo que expidió el certificado, el epígrafe 3 del artículo 36 del RIA prevé que los certificados podrán seguir siendo válidos durante un período temporal de nueve meses tras el cese de las actividades del organismo notificado, a condición de que otro organismo notificado haya confirmado

por escrito que asumirá responsabilidades respecto de los sistemas de IA cubiertos por dichos certificados, perdiendo su vigencia de forma inmediata, en caso contrario. En este caso, el nuevo organismo notificado deberá completar una evaluación completa de los sistemas de IA afectados al final de dicho período antes de expedir nuevos certificados para dichos sistemas.

En caso de suspensión o limitación de una designación, el nuevo epígrafe 8 del artículo 36 del RIA detalla las circunstancias en que los certificados que no hayan sido indebidamente expedidos podrán mantener su validez, existiendo dos posibilidades alternativas:

— Cuando la autoridad notificante haya confirmado, en el plazo de un mes a partir de la suspensión o limitación, que no existe riesgo para la salud, la seguridad o los derechos fundamentales en relación con los certificados afectados por la suspensión o limitación, y la autoridad notificante haya establecido un calendario y las medidas previstas para subsanar la suspensión o limitación.

— Cuando la autoridad notificante haya confirmado que no se expedirán, modificarán o volverán a expedir certificados relevantes a la suspensión durante el curso de la suspensión o limitación, e indique si el organismo notificado tiene la capacidad de seguir supervisando y seguir siendo responsable de los certificados existentes expedidos durante el período de la suspensión o limitación. En caso de que la autoridad responsable de los organismos notificados determine que el organismo notificado no tiene capacidad para respaldar los certificados existentes expedidos, el proveedor comunicará a las autoridades nacionales competentes del Estado miembro en el que el proveedor del sistema cubierto por el certificado tenga su domicilio social, en un plazo de tres meses a partir de la suspensión o limitación, una confirmación por escrito de que otro organismo notificado cualificado asume temporalmente las funciones del organismo notificado para supervisar y seguir siendo responsable de los certificados durante el período de suspensión o limitación.

Finalmente, en caso de retirada de una notificación, el nuevo epígrafe 9 del artículo 36 del RIA detalla las circunstancias en que los certificados que no hayan sido indebidamente expedidos podrán mantener su validez durante un período de nueve meses, exigiéndose dos condiciones acumulativas:

— Cuando la autoridad nacional competente del Estado miembro en el que el proveedor del sistema de IA cubierto por el certificado tenga su domicilio social haya confirmado que no existe ningún riesgo para la salud, la seguridad y los derechos fundamentales asociados a los sistemas en cuestión, y

— Otro organismo notificado haya confirmado por escrito que asumirá responsabilidades inmediatas en relación con dichos sistemas y que habrá completado la evaluación de los mismos en un plazo de doce meses a partir de la retirada de la designación.

4. OBLIGACIONES DE INFORMACIÓN DE LOS ORGANISMOS NOTIFICADOS

El artículo 45 del RIA establece obligaciones informativas que deben cumplir los organismos notificados, tanto frente a las autoridades notificantes (epígrafe 1) como

frente a otros organismos notificados (epígrafe 2), siguiendo el texto originalmente propuesto por la Comisión.

El epígrafe 3 del mismo artículo ha sido ligeramente modificado durante los diálogos tripartitos, mejorándose su redacción para referirse a los tipos de sistemas de IA, en lugar de a las tecnologías de IA, posiblemente en orden a garantizar mejor los secretos de empresa a que eventualmente puedan tener acceso los organismos notificados.

Finalmente, y a propuesta del Consejo, se ha incorporado un nuevo epígrafe 4, que sujeta el cumplimiento de las obligaciones de información precisamente al régimen de confidencialidad previsto en el artículo 78 del RIA, a cuyo análisis cabe remitirse.

IV. RECAPITULACIÓN

En general, se aprecia una importante similitud entre la regulación de los organismos notificados contenida en el RIA y las normas del Nuevo Modelo Legislativo, en especial la Decisión (UE) n.º 768/2008.

En lugar de remitir a la misma, posiblemente se ha decidido crear un régimen específico, aunque alineado con el ya existente, para cubrir las evaluaciones de la conformidad de sistemas de IA de alto riesgo, con intervención de organismos notificados, en dos situaciones: en el caso previsto en el anexo III.1 del RIA (determinados tratamientos biométricos) y en el caso de organismos notificados que no hayan sido previamente designados en el marco de las leyes de armonización conforme al Nuevo Modelo Legislativo (anexo I, sección A del RIA). En el caso de organismos ya notificados conforme al Nuevo Modelo Legislativo, posiblemente hubiera sido suficiente con el establecimiento de los requisitos adicionales que deberán aplicar, como en todo caso se ha hecho.

Sin embargo, resulta ciertamente llamativo que sólo en unos pocos casos de sistemas de IA de alto riesgo se exija realmente la intervención de un organismo notificado, dejando los restantes a la auto-evaluación de los proveedores, y sin control alguno en los restantes sistemas de IA. Será necesario esperar a ver los resultados de este enfoque antes de valorar el acierto de los co-legisladores en este modelo de (¿ausencia de?) control.

LAS OBLIGACIONES DE LOS PROVEEDORES E IMPLANTADORES DE SISTEMAS DE ALTO RIESGO

La evaluación de impacto de derechos fundamentales por quienes despliegan sistemas de inteligencia artificial en el Reglamento

EDUARD CHAVELI DONET

Abogado especialista en Derecho Digital
Head of Consulting Strategy en Govertis, Part of Telefónica Tech

I. INTRODUCCIÓN

Si el primer artículo del RIA al establecer su objetivo dispone que, además de «*mejorar el funcionamiento del mercado interior*» es «*promover la adopción de la inteligencia artificial centrada en el ser humano y digna de confianza*», y al mismo tiempo enfatiza que ello debe de ser «*garantizando al mismo tiempo un elevado nivel de protección de la salud, la seguridad y los derechos fundamentales consagrados en la Carta*», es lógico que las obligaciones que el mismo contempla deben de estar alineadas con dichos objetivos. Por ello la referencia a los derechos fundamentales es una constante en los considerandos y en el texto del mismo.

Como sabemos, los derechos humanos no son «algo nuevo» y con los siglos se ha producido una evolución en cuanto a su número, los sujetos beneficiarios, así como en cuanto a su alcance territorial; aunque su aplicación real diste de ser verdaderamente global. Pero, incluso en donde existe una «cultura» y sistemas tuitivos de los derechos humanos (como es el caso de Europa), la evolución industrial primero y la tecnológica después aportaron oportunidades y riesgos para los mismos que tuvieron que ser gestionadas. Del mismo modo, la Inteligencia Artificial (IA) es una revolución y, como tal, va a aportar oportunidades y riesgos que pueden afectar también a los derechos fundamentales y han de ser gestionados. El objetivo de este capítulo es precisamente realizar una aproximación conceptual y metodológica a una de las obligaciones que dispone el RIA para gestionar dichos riesgos: las Evaluaciones de impacto sobre los derechos fundamentales en ciertos sistemas de IA de alto riesgo, se denominan habitualmente Fundamental Rights Impact Assessments en inglés y aquí utilizaremos también el acrónimo en dicha lengua (FRIA). Estas tienen como objetivo que el implantador identifique los riesgos específicos para los derechos de las personas o grupos de personas que puedan verse afectados, e identifique las medidas que deben adoptarse en caso de que se materialicen estos riesgos.

Pero, como decíamos, aunque la eclosión de la IA se ha producido recientemente y ha supuesto la necesidad de su regulación, ya hace tiempo que los derechos humanos existen y hay precedentes de evaluaciones de impacto tanto sobre los derechos humanos (EIDH), así como evaluaciones de impacto social (EIS), éticas (EIE), así como sobre algún derecho concreto, como el conocido ejemplo de la protección de datos de carácter personal, que van a servir de base para abordar un análisis metodológico de las FRIA. Incluso ha habido ya metodologías y herramientas específicamente aplicados a los sistemas de IA que también analizaremos previamente a adentrarnos en el marco actual que contempla el RIA y como ha ido evolucionando en sus diversas versiones.

Asimismo, además de las FRIA que regula en su artículo 27 el RIA en esa aproximación al riesgo —como principio que lo inspira—, contempla también el análisis de riesgos como parte del sistema de gestión de IA (AARRIA) en su artículo 9, que se aborda en otro capítulo de este libro, lo que supone intersecciones y posibles confusiones entre ambos, que intentaremos deslindar.

II. LAS EVALUACIONES DE IMPACTO COMO HERRAMIENTA DE PONDERACIÓN DE DERECHOS FUNDAMENTALES

Pero, antes de adentrarnos en la parte adjetiva relativa a la metodología y analizar dichas intersecciones y diferencias, empezaremos hablando de las evaluaciones de impacto (EI) en general como herramientas de ponderación de derechos fundamentales; poniendo el acento en la parte objeto de dichas evaluaciones que son los derechos fundamentales, y también haciendo referencia a la tarea de ponderación, que es el objetivo de las mismas.

Ya desde un principio hemos de indicar que nos referiremos a ambos términos (derechos humanos y derechos fundamentales) de forma equivalente, pese a su distinción conceptual¹, además de que —como es sabido— la propia Constitución en su artículo 10.2 establece que *«las normas relativas a los derechos fundamentales y las libertades que la Constitución reconoce se interpretarán de conformidad a la*

1. Los Derechos Humanos tienen Carácter Universal y los Derechos Fundamentales, estando obviamente relacionados y coincidiendo en gran parte, dependen de cómo se aterrizan los derechos humanos en un ámbito específico a través de una norma: en el ámbito Europeo los derechos fundamentales son los que contempla la Carta y también las constituciones. En este sentido la Agencia de los Derechos Fundamentales de la Unión Europea en <https://fra.europa.eu/en/about-fundamental-rights/frequently-asked-questions#difference-human-fundamental-rights> acces— sed 10 de enero de 2021 («El término “derechos fundamentales” se utiliza en la Unión Europea (UE) para expresar el concepto de “derechos humanos” en un contexto interno específico de la UE. Tradicionalmente, el término “derechos fundamentales” se utiliza en un entorno constitucional, mientras que el término “derechos humanos” se utiliza en el derecho internacional. Los dos términos se refieren a una sustancia similar, como puede verse al comparar el contenido de en la Carta de los Derechos Fundamentales de la Unión Europea con el del Convenio Europeo de Derechos Humanos y la Carta Social Europea»).

De hecho, a modo de curiosidad, le pedí a una aplicación de IA generativa que realizase una comparación entre las similitudes y diferencias en cuanto a los derechos recogidos tanto entre la DUDH y la CDFUE, así como entre los que contempla CD-FUE y los que contempla la Constitución Española y la realizó con éxito indicando las grandes similitudes y también algunas diferencias.

Declaración Universal de los Derechos Humanos y los tratados y acuerdos internacionales sobre las mismas materias ratificados por España».

Una vez aclarado lo anterior y para poder aterrizar en qué consiste una Evaluación de impacto derechos fundamentales (o evaluación de impacto en derechos humanos, EIDH) hay que decir previamente que evaluar derechos significa ponderarlos. La ponderación de derechos se realiza en primera instancia por las propias normas al dotar de un rango de Derecho Fundamental o no a ciertos derechos. El hecho de que, por ejemplo, la Constitución otorgue distintos rangos a diversos derechos, no es cuestión baladí porque —como es sabido— de ello se derivarán una serie de consecuencias. Pero esa ponderación previa que realiza la Constitución no suele ser suficiente, especialmente cuando hablamos de derechos «en igualdad de condiciones»: por ejemplo, cuando hablamos de ponderación entre derechos fundamentales. En cada país, y a partir de la Constitución, el legislador y en ciertos supuestos excepcionales (aunque cada vez más frecuentes el ejecutivo) al aprobar disposiciones legales también realizan un ejercicio de ponderación. Asimismo —y como bien indica Pere Simón²— haciendo referencia al caso concreto del Derecho Administrativo —«en caso de no ser resueltos en primer lugar por el legislador democrático (Arrojo Jiménez, 2009: 27 y ss.), puede requerir también una ponderación que puede proceder a través del desarrollo de la actividad administrativa normativa o de la actividad administrativa no normativa— en forma de directrices, guías, cláusulas generales de apoderamiento, etc.»³. Y obviamente existe una labor de ponderación que realizan los tribunales, y no sólo el TEDH y en el caso de España el TC, sino todos los jueces y magistrados realizan dicha función como se deduce del art. 24 y 53 de la CE y más concretamente, entre otros, del art. 7 de la LOPJ.

Pero, además de dicha labor de ponderación que realizan el legislador, los tribunales y la administración se ha ido importando a Europa desde la tradición anglosajona la realización de evaluaciones de impacto *ex ante* por parte de los propios sujetos que son parte del proceso de toma de decisiones (administraciones y empresas) sobre los medios y fines que pueden afectar a los citados derechos. Se trata de una práctica derivada de que vivimos en una sociedad en la que cada vez existen mayor número de riesgos⁴ y una parte esencial de las evaluaciones de impacto (como después tendremos ocasión de profundizar) es, precisamente, la gestión de riesgos. Sin perjuicio de que volvamos con detalle a tratar el análisis de riesgos al adentrarnos más adelante en la metodología, sí que es importante indicar desde el principio que el análisis de riesgos es una práctica habitual desde hace años vinculada a sectores por exigencia del «negocio», como es el caso del sector financiero o asegurador. Se trata asimismo de una materia que

2. Simón Castellano, P., *La evaluación de impacto algorítmico en los derechos fundamentales*, Aranzadi, Cizur Menor, 2023, p. 48.

3. Como indica Simón Castellano, P., ob. Cit. (p. 48) «tal extremo ha sido aceptado por la dogmática jurídica continental y por la doctrina de la tradición jurídico civilista (Schmidt-Assmann, 2003: 2019 y ss. y Franzius, 2006: 108 y ss.)».

4. Como dice Mantelero, A., en ob. cit.p.13 «como consecuencia de la transformación de la sociedad moderna en una sociedad de riesgo, o al menos una sociedad en la que muchas actividades conllevan la exposición a riesgos y que se caracteriza por la aparición de nuevos riesgos».

históricamente se realizó sobre la base de estándares, tanto generales (como la ISO 31000:2018, de gestión de riesgos) así como otras normas ISO que han aterrizado la gestión de riesgos en diversos ámbitos TI, como por ejemplo la ISO/IEC 27001:2022 en materia de seguridad de la información, la ISO 22301:2020 sobre continuidad de negocio; o en protección de datos la ISO 27701:20021 sobre sistemas de gestión de privacidad, o la ISO/IEC 27018:2014 relativa al tratamiento de datos personales en cloud, entre otras. En todas ellas se hace referencia a la gestión de riesgos. Asimismo, el análisis de riesgos ya empezó a aterrizar en su día en disposiciones legales, como por ejemplo el Esquema Nacional de Seguridad (ENS)⁵ y está proliferando en otros ámbitos como el Compliance Penal, el Blanqueo de Capitales, la protección de datos y ahora en relación con la inteligencia artificial. De hecho, el RIA acuña un enfoque basado en riesgos —según su Considerando 26— *«para introducir un conjunto proporcionado y eficaz de normas vinculantes para los sistemas de IA»* adaptando *«el tipo y el contenido de tales normas a la intensidad y el alcance de los riesgos que pueden generar los sistemas de IA»*. Esto lo hace el RIA mediante la siguiente fórmula que algunos autores han criticado⁶: *«prohibir determinadas prácticas inaceptables de inteligencia artificial, establecer requisitos para los sistemas de IA de alto riesgo y obligaciones para los operadores correspondientes, y establecer obligaciones de transparencia para determinados sistemas de IA»*.

1. DIFERENCIAS ENTRE ANÁLISIS DE RIESGOS Y EVALUACIONES DE IMPACTO. EL EJEMPLO DE LA PROTECCIÓN DE DATOS

Para realizar algunas precisiones conceptuales necesarias sobre el concepto de análisis de riesgos (AARR) y el de evaluaciones de impacto (EI) nos serviremos del ejemplo de la protección de datos (con evidente relación con la Seguridad de la información), dado que —quizás— sea el ámbito que mayor madurez ha alcanzado en nuestro contexto en relación con la generalización de la realización de análisis de riesgos y evaluaciones de impacto (*ex art.35* del RGPD), habida cuenta de que se trata de una materia que «aplica» a todos los sectores de actividad. Nos parece importante porque: por un lado, muchos profesionales de los que ahora se acercan a la IA provienen de la protección de datos y también porque los sistemas de IA pueden y suelen conllevar un tratamiento de datos personales. No obstante, y a pesar de dicha aproximación comparativa, como en el siguiente apartado veremos, precisamente en el caso de los AARRIA y las FRIA no se pueden *aplicar mutatis mutandis* dichas consideraciones.

5. El Esquema Nacional de Seguridad en su versión inicial del RD 3/2010 modificado por el RD 951/2015 ya lo contemplaba, y obviamente ha mantenido el nuevo ENS aprobado por el Real Decreto 311/2022, de 3 de mayo.
6. Vid. Mantelero, A., en ob. cit. p.173 que indicaba *«Aunque esto es eficaz en términos de impacto político y aceptabilidad, es una forma débil de prevención de riesgos. La Propuesta hace una distinción bastante rígida entre el riesgo de alto nivel y el resto, no proporcionando ninguna metodología para evaluar el primero, y eximiendo en gran medida al segundo de cualquier mitigación (con la limitada excepción de las obligaciones de transparencia en ciertos casos)»*.

Las evaluaciones de impacto y los análisis de riesgos son dos cosas diferentes, pero íntimamente vinculadas., como dice la AEPD (y se desarrollan de forma detallada en la Guía⁷):

«El análisis y la gestión de riesgos son procedimientos que permiten a las organizaciones identificar y poder anticiparse a los posibles efectos adversos o no previstos que el tratamiento pueda tener para los derechos y libertades de las personas interesadas. Esta gestión debe permitir que la persona responsable tome las decisiones y acciones necesarias para conseguir que el tratamiento cumpla los requisitos del RGPD y la LOPDGDD, garantizando y pudiendo demostrar la protección de los derechos de las personas interesadas.

Por su parte, el RGPD establece que cuando sea probable que un tipo de tratamiento entrañe un alto riesgo, la persona responsable debe realizar una evaluación del impacto, proceso que permite a las organizaciones identificar los riesgos que un sistema, producto o servicio puede implicar para los derechos y libertades de las personas y, tras haber realizado ese análisis, afrontar y gestionar esos peligros antes de que se materialicen.

La gestión del riesgo y la EIPD son procesos que se encuentran estrechamente vinculados, ya que la segunda es una especificidad dentro de la primera. Así, la EIPD no puede existir sin formar parte de la gestión de riesgos, por lo que mientras que la gestión del riesgo es obligatoria para todo tratamiento, las obligaciones concretas que se establecen para la EIPD lo son exclusivamente para tratamientos de alto riesgo».

Asimismo, la protección de datos tiene una evidente relación con la Seguridad de la información, en tanto en cuanto los datos personales son un activo de la información y han de protegerse con las debidas medidas de Seguridad. Por ello y para visualizarlo con un ejemplo en el ámbito del sector público: el ENS, al que nos hemos referido, obliga a las organizaciones a realizar un análisis de riesgos respecto de la información que está dentro del alcance, para determinar las medidas a aplicar sobre la base de tener en cuenta la posible afectación a ella de cinco dimensiones de la seguridad: la confidencialidad, integridad, autenticidad, disponibilidad y trazabilidad.

En este caso por ejemplo el análisis de riesgos ayuda a proteger la seguridad de la información y no tiene en cuenta el riesgo específico que existe para el tratamiento de datos personales. Por ello el mismo ENS determina que *«cuando un sistema afecte a datos de carácter personal el responsable o el encargado del tratamiento, asesorado por el delegado de protección de datos, realizarán un análisis de riesgos conforme al artículo 24 del RGPD y, en los supuestos de su artículo 35, una evaluación de impacto en la protección de datos».*

Por tanto, y en conclusión: si hablamos de protección de datos de carácter personal el análisis de riesgos que hay que realizar de todos los tratamientos de datos de carácter personal puede tomar como base el que se haya realizado con otro marco (en este caso el ENS, al igual que puede tomarse en cuenta otro SGSI como el basado en ISO 27001, pero deberá de tenerse en cuenta adicionalmente

7. <https://www.aepd.es/prensa-y-comunicacion/notas-de-prensa/aepd-publica-nueva-guia-gestionar-riesgos-y-evaluaciones-impacto>

los riesgos para la protección adicional que en determinados casos y fruto del tratamiento de ciertos tipos de datos (por ejemplo categorías especiales de datos, haya que considerar.

Pero: ¿Podemos aplicar esta distinción del RGPD al caso del RIA? Veámoslo a continuación.

2. DIFERENCIAS ENTRE LOS RIESGOS A TRATAR DEL SISTEMA DE GESTIÓN DE RIESGOS DEL ARTÍCULO 9 VS EVALUACIÓN DE IMPACTO DE DERECHOS FUNDAMENTALES DEL ARTÍCULO 27 REGLAMENTO

Siendo la anterior la relación teórica entre los AARRPD y las EIPD que la AEPD ha recogido en su guía⁸ (haciéndose eco de artículos del RGPD y también de los WP del CEPD), la misma también describe no sólo su relación sino también las peculiaridades de la EIPD que la diferencian del AARR en protección de datos y quiero agradecer especialmente a Jordi Morera y María Loza que en un trabajo coral me han ayudado a trabajar una posición de «entendimiento común».

Por ello, y sólo a los efectos de que dicha distinción entre AARRPD y EIPD inspire el proceso comparativo entre los AARRIA que exige el artículo 9 del RIA y las FRIA que exige el artículo 27 del RIA, vamos a seguir (sólo dichos efectos de guion) dicha estructura que propone la AEPD para las EIPD:

Por un lado, las EIPD son exigibles «*cuando hay un alto riesgo para los derechos y libertades*». En el caso de los AARRIA y las FRIA la diferencia esencial es que en ambas se parte de una situación de «Alto Riesgo» (puesto que se está utilizando una IA de Alto Riesgo); pero en las FRIA además se exige sólo a los implantadores de ciertos procesos que, por ser públicos o por razón de la materia, que después veremos, se entiende que pueden suponer un riesgo «superior», que requiere de una supervisión adicional de las autoridades (por eso se incluye en documentación y en proceso de revisión de las Autoridades). Ello sin perjuicio de que los proveedores de sistemas de alto riesgo y que están sometidos al AARRIA del artículo 9 deben de registrar el sistema, luego esa supervisión es para todos los proveedores de alto riesgo (Art 51). Incluso los proveedores que consideren que no es de alto riesgo, también lo tienen que registrar.

A nivel de los sujetos obligados a realizarla, y como después desarrollaremos, al igual que la EIPD es una obligación específica del responsable del tratamiento, sin perjuicio de la asistencia del encargado del tratamiento, sobre el que después volveremos; en el caso de los AARRIA se trata de una obligación de los proveedores en su concepto amplio que contempla el RIA (*vid* el considerando 96), aunque también la puedan realizar los que realicen el despliegue o usuarios del sistema de IA. Pero la FRIA es una obligación del que realiza el despliegue, sin perjuicio de que pueda basarse en el AARRIA del proveedor.

En el caso de las EIPD se exige un análisis de la necesidad y proporcionalidad del tratamiento con relación a sus fines. El AARRIA no exige ese análisis de necesidad y proporcionalidad; y aunque tampoco lo exija el texto del RIA para las FRIA, como veremos si lo exigen ciertas metodologías que ya han existido como precedentes

8. <https://www.aepd.es/documento/gestion-riesgo-y-evaluacion-impacto-en-tratamientos-datos-personales.pdf>

y, en nuestro caso y como después profundizaremos, entendemos que debiera de realizarse, al menos de forma *soft*.

En el caso de las EIPD se exige su realización antes del inicio de las actividades de tratamiento, cosa que no se exige en los AARRPD. En el caso de los AARRIA no se dice nada, y aunque parezca lógico que antes de ponerse en marcha cualquier sistema de IA de alto riesgo se haga así, no se exige, a diferencia de las FRIA en que sí que lo dice expresamente; y quizá sea otra diferencia.

A diferencia de los AARRPD las EIPD exigen el asesoramiento del DPD, si este está nombrado. En cambio, el AARRIA no exige el asesoramiento *per se* de ningún rol, sin perjuicio de la conveniencia de la intervención de diversos roles a los que después nos referiremos. En el caso de las FRIA, exigirán el asesoramiento del DPD cuando haya un tratamiento de datos personales que exija una EIPD.

En el caso de los AARRPD nada se dice al respecto y en el caso de las EIPD se requiere recabar la opinión de los interesados, o sus representantes, cuando proceda, en el proceso de gestión del riesgo, justificando en su caso la no procedencia o la limitación en la comunicación de información. Por su parte, en el caso de los AARRIA el Considerando 96 a del RIA indica que, a la hora de identificar las medidas de gestión de riesgos más adecuadas, el proveedor deberá documentar y explicar las decisiones tomadas y, cuando proceda, implicar a expertos y partes interesadas externas. Y por su parte en el caso de las FRIA el considerando 96 dice que «cuando proceda, para recopilar la información pertinente necesaria para realizar la evaluación de impacto, los implantadores de sistemas de IA de alto riesgo, en particular cuando los sistemas de IA se utilicen en el sector público, podrían implicar a las partes interesadas pertinentes, incluidos los representantes de los grupos de personas que puedan verse afectados por el sistema de IA, expertos independientes y organizaciones de la sociedad civil en la realización de dichas evaluaciones de impacto y en el diseño de las medidas a adoptar en caso de materialización de los riesgos».

En el caso de las EIPD, a diferencia del AARRPD, su resultado se debe tener en cuenta para evaluar la viabilidad o inviabilidad del tratamiento desde el punto de vista de protección de datos. Por su parte en el caso de AARRIA y EIPDIA parece lógico que en ambos casos se debe de tener en cuenta para analizar si el sistema de IA se mantiene en funcionamiento o no, como podría pensarse en el caso de protección de datos; pero en el caso de las FRIA a diferencia del AARRIA se deja claro que debe de realizarse antes del primer uso.

A diferencia de los AARRPD, en el caso de las EIPD, en función del nivel de riesgo residual, obliga al responsable a realizar una Consulta Previa (art. 36 RGPD) a la Autoridad de Control. En el caso de los AARRIA no se dice nada; pero en cambio en el caso de las FRIA (no vinculada al riesgo residual sino en todo caso) debe de notificarse a la autoridad de vigilancia del mercado los resultados de la evaluación, presentando la plantilla cumplimentada a que se refiere el apartado 5 del artículo 27, excepto en ciertos supuesto, a los que posteriormente nos referiremos. Este es un punto importante, si sí o sí, tengo que notificar el resultado del FRIA, podría interpretarse que el FRIA junta tanto ciertas características de la EIPD como de la Consulta Previa del 36, ya que, al registrar el sistema con la Evaluación, las Autoridades (en base a los poderes generales que tienen, por ejemplo, el del art. 67), podrían ordenar la adopción de medidas de corrección sobre la IA para mitigar aún más los riesgos.

Otra diferencia entre los AARRPD y más aún de EIPD de privacidad es que en privacidad (en ambos casos) sí que hay que tener en cuenta precisamente las situaciones concretas del tratamiento (puede leerse de forma reiterada en la guía de la AEPD y se deduce del RGPD). Es decir el tipo de análisis. En el caso de los AARRIA se refiere a «riesgos conocidos y predecibles», tanto del uso «normal» como del mal uso. Y teniendo en cuenta que el obligado es el proveedor de la solución, pero no el implantador / usuario de la solución, se tratará de un análisis de riesgos «general»; es decir, la tipología de amenazas y taxonomía de las mismas serán sobre el producto en sí y sus usos previstos y malos usos previstos, pero no estarán aterrizados al caso de uso concreto de una empresa en cuestión. El análisis es más «personalizado» puesto que implica ya al implantador / usuario e importante el legislador lo que indica: «*shall perform an assessment of the impact on fundamental rights that the use of the system may produce. For that purpose, deployers shall perform an assessment consisting of... a description of the deployer's processes in which the high-risk AI system will be used in line with its intended purpose*». Por ello, la FRIA es un análisis totalmente mucho más enfocado al caso de uso concreto. Como he indicado en otros comentarios, se hace mucho más hincapié en describir los procesos y el uso concreto. (grupos de afectados, frecuencia, etc.).

Es decir, aunque nos haya servido a modo comparativo la referencia a los AARRPD y las EIPD, precisamente dicha comparación y también la lectura no sólo de los artículos 9 y 27 del RIA sino la lectura general del RIA nos lleva a entender que la comparativa AARRPD Y EIPD (los arts. 24 y 32 del RGPD en relación con el 35 RGPD) no es aplicable exactamente a los AARRIA y FRIA. Mientras que las EIPD del 35 del RGPD son claramente un plus de exigencia con respecto a la exigencia general (tanto para sector público como privado) respecto de los AARRPD; en cambio en el caso del RIA, parece dar a entender que el artículo 9 RIA pretende que toda solución de IA de Alto Riesgo tenga los riesgos documentados gestionados por el proveedor de la misma (como ocurre en otros productos) pero evidentemente no aterrizados al caso de uso concreto de una empresa concreta que los despliegue y conforme sus particularidades, sino a nivel de lo razonablemente previsiblemente y que vaya siendo actualizado conforme la monitorización del mercado. En cambio, el art. 27 al referirse a las FRIA sí que parece orientado a analizar los riesgos al caso de uso concreto, pero únicamente aplicable al caso de uso que implica o conecta con el ejercicio de funciones públicas y supuestos concretos que después se indican.

Por aterrizarlo a un ejemplo, suponiendo que proveedor A desarrollase un Chatbot (asumimos que sea de alto riesgo), este tendrá un análisis de riesgos de distintos escenarios de uso, pero si este Chatbot se implica por la Administración A, tendrá que hacer una FRIA sobre su caso de uso y la Administración B también (por ejemplo, la primera lo quiere implementar sobre ciudadanos y la segunda sobre empleados, variaría la circunstancia del art. 27.a 1. (c) (grupos de interesados), el resultado de esta de dos FRIA por dos implantadores / usuarios distintos podría ser distinto (aunque ya se deja entrever por el legislador la posibilidad de aplicar analogía casos similares ya validados).

En definitiva, las FRIA del art. 27 actúan como una salvaguarda para evitar abusos en ciertos supuestos, pero a su vez como mecanismo de seguridad jurídica y promover el uso de la IA en el sentido que, una vez validado un caso de uso, pueda ser ya tenido en cuenta para validar las FRIA sobre casos similares.

3. TIPOLOGÍAS DE EVALUACIONES DE IMPACTO

Una vez sentada la íntima relación entre los AARR y las EI es necesario conocer los diferentes modelos de EI que existen antes de aterrizar centrarnos en las FRIA que propone el RIA.

Mantelero⁹ se refiere a los ejemplos más cercanos que nos sirven de aproximación: La HRESIA o EISDH (Ede Impacto Social, Ético y de los Derechos Humanos), la PIA (Privacy Impact Assessment) o DPIA (Data Privacy Impact Assessment), la SIA (Social Impact Assessment) y la EtIA Ethical Impact Assessment); analiza las características, semejanzas y diferencias de cada modelo, sus ventajas y posibles inconvenientes que conllevan. No obstante, nos centraremos en las EIDH. Como los derechos fundamentales existen desde hace años y, de forma previa a la existencia de la IA y su reciente eclosión, los mismos han estado y están sometidos a riesgos en entornos diferentes a la IA. Por ello ha habido evaluaciones de impacto sobre diferentes temas sensibles: Por ejemplo, las EI en el medio ambiente fueron consideradas ya en la década de 1960, y actualmente son una exigencia legal en muchos países. En el caso de España por ejemplo las contempla la Ley 21/2013, de 9 de diciembre, de evaluación ambiental. Más allá del debate sobre la necesidad de que la protección del medio ambiente sea un derecho humano, lo cierto es que lo contempla Carta de Derechos Fundamentales de la Unión Europea (CDFUE) y además existe una evidente preocupación por el medio ambiente que se ha visto incrementada por el cambio climático y que también se ha proyectado en el propio RIA en diversos considerandos¹⁰ y artículos. No es casualidad que también la ISO 42001:2023 a la que después nos referiremos sobre IA haga referencia al medio ambiente y al cambio climático. No obstante, aunque ya profundizaremos más adelante, y a diferencia de otros derechos, en el caso del medio ambiente se trata de una evaluación de impacto sobre aspectos más objetivables (agua, suelo, aire etc.).

Históricamente las EIDH se han realizado de formas más habitual en relación con determinadas actividades «sensibles» a ciertos derechos, como excavaciones mineras a cielo abierto, operaciones petrolíferas o de gas, fábricas y otras actividades en las que —no pocas veces— se han desarrollado dichas actividades en países con un importante déficit en derechos humanos¹¹ y llevadas a cabo por las empresas que

9. Mantelero, A., *Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI*, Information Technology and Law Series, 2022.

10. Por ejemplo, dispone el RIA en su Considerando 27 «Por “bienestar social y ambiental” se entiende que los sistemas de IA se desarrollan y utilizan de manera sostenible y respetuosa con el medio ambiente, así como en beneficio de todos los seres humanos, al tiempo que se supervisan y evalúan los efectos a largo plazo en las personas, la sociedad y la democracia. La aplicación de estos principios debe traducirse, cuando sea posible, en el diseño y el uso de modelos de IA».

11. Un aspecto en el que la Guía de evaluación y gestión de impactos en los derechos humanos (EGIDH) del Foro Internacional de Líderes Empresariales (International Business Leaders Forum, IBLF) y la Corporación Financiera Internacional (International Finance Corporation, IFC), en asociación con la Oficina del Pacto Mundial de las Naciones Unidas, pone foco es el ámbito territorial:

- Una zona con gobernabilidad débil.
- Un país en estado precario y/o afectado por conflictos.
- Un área donde los compromisos en materia de derechos humanos son implementados de manera deficiente.

explotan dicha actividad y que pertenecen a países más desarrollados. Asimismo, y no en pocas ocasiones se ha acusado de que se trata de operaciones de intento de lavado de imagen ante noticias por denuncias.

Incluso algunos ejemplos cercanos de estos sectores podemos verlos en reconocidas multinacionales españolas, como el caso de Iberdrola o Repsol (aunque con alcance diferente: en el primer caso general y en el segundo afectando a una explotación concreta) y también con diferentes niveles de detalle de la información publicada, como se puede apreciar. En muchos casos las políticas aprobadas y las EIDH se han basado en los Principios Rectores de la ONU. Obviamente ha habido también EIDH en otros sectores privados, pero históricamente menos habituales.

Otro ejemplo habitual de EIDH lo constituyen las realizadas por ciertas ONG (como el caso de OXFAM o Cruz Roja, por ejemplo), que, sí que son habitualmente publicadas, pero que tampoco han estado exentas de críticas sobre posibles sesgos y de acusaciones de parcialidad.

El sector público tampoco es ajeno a las EIDH, con supuestos que han ido proliferando como el liderado por la Secretaría de Relaciones con las Cortes sobre la Administración General del Estado¹².

En otros casos ha habido aproximaciones que, más que con los sectores de actividad, tienen que ver con los sujetos afectados, como es el caso de los niños¹³ u otros grupos vulnerables. Y esto también se ha proyectado sobre la IA. No es casualidad que el RIA mencione precisamente a los niños por ejemplo en el Considerando 48¹⁴ y que el artículo 9 al hablar de gestión de riesgos¹⁵.

No somos ajenos a las tensiones derivadas de las diferentes visiones sobre los derechos humanos y de la discusión sobre el universalismo y el relativismo cultural en los mismos, pero los derechos humanos, aún con las consabidas diferencias, proporcionan un marco ampliamente aceptable, y aplicable como referencia en relación con las evaluaciones de impacto de IA.

Además de las EIDH también ha habido ejemplos de evaluaciones de impacto que integran los Derechos Humanos con aspectos sociales y éticos, pues no en balde

— Un área con riesgos e impactos medioambientales y/o sociales elevados.

— Un área habitada por comunidades locales vulnerables (por ejemplo, pueblos indígenas).

12. <https://www.abogacia.es/wp-content/uploads/2013/01/INFORME-DE-EVALUACION-DEL-PLAN-DE-DDHH.pdf>

13. Como es el caso de los niños en el que en 2012 UNICEF aprobó una guía para integrar los derechos del niño en las evaluaciones de impacto u otras.

14. El Considerando 48 del RIA se refiere a que los niños tienen derechos específicos consagrados en el artículo 24 de la Carta de la UE y en la Convención de las Naciones Unidas sobre los Derechos del Niño (desarrollados en la Observación General n.º 25 de la CNUDN por lo que se refiere al entorno digital), ambos de los cuales requieren que se tengan en cuenta las vulnerabilidades de los niños y que se les proporcione la protección y los cuidados necesarios para su bienestar.

15. Artículo 9.8. Al aplicar el sistema de gestión de riesgos descrito en los apartados 1 a 6, los proveedores tendrán en cuenta si, habida cuenta de su finalidad prevista, el sistema de IA de alto riesgo puede afectar negativamente a los menores de 18 años y, en su caso, a otros grupos vulnerables de personas.

los valores éticos y sociales son fundamentales para aterrizar los derechos humanos en cada contexto geográfico y cultural.

III. ANTECEDENTES DE EVALUACIONES DE DERECHOS FUNDAMENTALES DE SISTEMAS DE IA PREVIOS AL REGLAMENTO

Como hemos visto, anteriormente ya existían las evaluaciones de impacto tanto sobre derechos fundamentales, como sociales y éticas previas a la eclosión de la IA, pero ésta obligó a formular ya, antes de la aprobación del RIA, aproximaciones conceptuales y metodológicas, tanto doctrinales como prácticas sobre las Evaluaciones de Impacto Algorítmicas (EIA).

Creo justo destacar dos de las pocas, pero magníficas publicaciones que sobre la materia ha habido y de cita reiterada en este capítulo: de Mantelero¹⁶ y de Pere Simón¹⁷, sin perjuicio de que hay más, y de las que comentaremos algunas.

En el caso de Pere Simón realiza una clasificación de las posibles metodologías en función de diversos criterios:

— En función de las consecuencias: Los derechos e intereses legítimos potencialmente afectados.

— En función de la tecnología empleada.

— En función de la participación de terceras partes: proveedores, clientes, vendedores, auditores.

— Y en función de los principales riesgos, que agrupa en tres bloques: Riesgo de la desinformación, riesgo de discriminación (*disparate impact assessment*) riesgo de indefensión y segunda oportunidad.

Por su parte Mantelero en la obra citada, realiza una apuesta concreta por la HRESIA (Evaluación del Impacto Ético-Social en los Derechos Humanos), un modelo híbrido que tiene en cuenta «tanto el impacto ético como el social de una tecnología junto con las dimensiones legales y las de los derechos humanos» y que permite «combinar la universalidad de los derechos humanos con la dimensión local de los valores sociales». Sostiene que los informes tradicionales de EIDH suelen describir los riesgos encontrados y su posible impacto, pero sin una evaluación cuantitativa, y ofrecen recomendaciones sin clasificar el nivel de impacto, dejando a los responsables la tarea de definir un plan de acción adecuado.

En esta visión —continúa— «los valores éticos y sociales se contemplan a través de la lente de los derechos humanos y sirven para ir más allá de las limitaciones de la teoría jurídica o de la aplicación práctica a la hora de abordar eficazmente las cuestiones más urgentes relativas a las repercusiones sociales de la IA». De hecho, dichos valores éticos y sociales son tenidos en cuenta al interpretar los derechos humanos por parte de las autoridades y tribunales.¹⁰¹

Se trata —según Mantelero— no de una evaluación tecnológica, sino de una evaluación basada en los diferentes derechos y valores, que «puede responder a la exigencia de una protección más amplia de las personas en el contexto de la IA y responder mejor a la creciente demanda de una IA» y «es coherente con los estudios en el ámbito de la

16. Mantelero, A., ob. cit.

17. Ob. Cit. Simón Castellano, P.

*protección de datos colectivos*² que señalan la importancia de estas dimensiones no jurídicas en el contexto de las aplicaciones de uso intensivo de datos», lo que es coherente con el hecho de que los sistemas de IA muchas tienen en cuenta datos que afectan a grupos o colectivos.

Además de las aproximaciones doctrinales que ha habido en España hay que tener en cuenta que en diversos países y también diversas instituciones han elaborado metodologías y herramientas que Pere Simón aborda en el libro ya citado:

- En Canadá el Algorithmic Impact Assessment Tool. (RIAT).
- La herramienta RIA del Gobierno de Estados Unidos.
- En los Países Bajos el modelo FRRIA¹⁸.

— Y otros como el modelo PIO (Principios, Indicadores y Observables) de la OEIAC, o las propuestas del Ada Lovelace Institute y el Model Rules del European Law Institute.

No se trata aquí de profundizar en cada una de dichas herramientas, pero sí que realizamos unas consideraciones básicas sobre algunas, incluyendo las referencias oportunas para ampliar información y posteriormente algunas de sus características son tenidas en cuenta a la hora de analizar el modelo propuesto.

Respecto del RIAT Canadiense, decir que no sólo la herramienta está disponible *on line*¹⁹ sino también la explicación del modelo²⁰; y además de otros aspectos que se indican posteriormente en relación con la metodología de la misma, resaltaría lo relativo a las áreas que son objeto de análisis:

- Los derechos de las personas o comunidades.
- La salud de las personas o comunidades.
- Los intereses económicos de individuos, entidades o comunidades.
- Y la sostenibilidad continua del ecosistema.

Sin negar el valor de ser una herramienta pionera y la simplicidad de su modelo, lo cierto es que al entrar en ella se analizan algunos aspectos y derechos que se han indicado, pero hay algunos derechos que no se analizan y —asimismo— la profundidad del análisis de los riesgos de cada derecho me parece demasiado simple, lo que puede restarle validez²¹.

18. <https://www.government.nl/binaries/government/documenten/reports/2022/03/31/impact-assessment-fundamental-rights-and-algorithms/Fundamental+Rights+and+Algorithms+Impact+Assessment.pdf>

19. <https://open.canada.ca/aia-eia-js/?lang=en> última consulta el 12/03/2024.

20. <https://www.canada.ca/en/government/system/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html> última consulta el 12/03/2024.

21. Obviamente se trata de disposiciones legales y de una herramienta que son diferentes, pero queremos aquí poner de manifiesto que además los derechos que son objeto de protección también son diferentes. Por ejemplo hay derechos reconocidos en la CEDF o en la DUDH que la RIAT evalúa directa o indirectamente: la vida, la dignidad humana, la integridad psíquica y psíquica, el derecho al trabajo y las condiciones laborales justas, la protección de los datos personales o el derecho a la salud y la asistencia médica; pero hay otros que ni directa ni indirectamente la RIAT tiene

Por su parte EE. UU. es, también en IA, un país de absoluta referencia, sin perjuicio de las diferencias que tiene con Europa y que va a «llegar más tarde que Europa» a tener una regulación específica y completa de la IA. Ha habido diferentes iniciativas siendo la más notoria la reciente Orden Ejecutiva firmada por Joe Biden el 30 de octubre de 2023, que algunos han calificado como «una Ley de tiempo de Guerra», y que juntamente con otras preexistentes constituyen un germen de lo que será probablemente la futura regulación en E.E.U.U. Entre esas iniciativas preexistentes podemos citar la *Algorithmic Impact Assessment* (RIA) que es una herramienta on line de para ayudar a las empresas q gestionar los riesgos de la IA.

Por otro lado, el modelo FRRIA ya citado (*Fundamental Rights and Algorithm Impact Assessment*) de los Países Bajos es un manual puesto a disposición por parte del Ministerio del Interior y Relaciones para ayudar a las organizaciones en la toma de decisiones sobre el uso de sistemas de IA. La FRRIA distingue las siguientes fases:

— Parte 1. Why? Aquí se analiza el «Por qué» de la intención del algoritmo. Cuáles son los motivos y efectos y los valores subyacentes.

— Parte 2. What? (input) Aquí se pone foco en la forma de qué, del objeto, del algoritmo. Esta parte se divide en dos subpartes:

1. La parte 2A se refiere a la entrada del algoritmo: los datos que se utilizarán y las condiciones previas correspondientes.

2. La parte 2B se refiere al algoritmo en sí. Es decir: por ejemplo, qué tipo de algoritmo se utiliza y cuáles son las condiciones previas para un uso responsable del algoritmo.

— Parte 3. How? En esta parte se refiere a cómo se produce la implementación, uso y supervisión y los resultados.

— Parte 4. Fundamental rights Roadmap. Aquí se incorpora una «hoja de ruta de derechos fundamentales» con un doble objetivo:

1. Sirve como herramienta para identificar si el algoritmo a utilizar afectará derechos fundamentales;

2. Si es así, facilita una discusión estructurada sobre si existen oportunidades para prevenir o mitigar esta interferencia en el ejercicio de los derechos fundamentales, y si esta se considerara aceptable.

Hay otros ejemplos de metodologías, como por ejemplo «*Deepfakes, Phrenology, Surveillance, and More ; A Taxonomy of AI Privacy Risks*»²², que es una aproximación a eventos de IA integrados con la Taxonomía de eventos de Privacidad; u otras como PLOT4ai,²³ que es una herramienta que contiene una colección de 86 amenazas diferentes y una metodología de modelado de amenazas, pudiendo seleccionar deferentes catálogos o todos (Technique & Processes, Accessibility, Identifiability & Linkability, Security, Safety, Unawareness, Non-compliance y Ethics & Human Rights). Por ejemplo, para esta parte de ética y derechos humanos, se realizan

en cuenta, como por ejemplo el derecho a la educación, la Propiedad Intelectual, el derecho a la vivienda o la Libertad de Expresión y de Información.

22. <https://arxiv.org/pdf/2310.07879.pdf>

23. <https://plot4.ai>

17 preguntas. Se realizan preguntas y con base en las mismas se determinan las correspondientes amenazas, y todo ello se refleja en un informe.

Asimismo, y antes de llegar al objeto nuclear de este capítulo (las FRIA en el RIA) es necesario hacer referencia a las normas ISO que, como estándares no sólo han normalizado lo relativo a aspectos a los que nos hemos referido como metodología de AARR, por ejemplo, sino que específicamente en materia de IA. Hay diversas ISO que se refieren a la IA y que es conveniente tener en cuenta a la hora de realizar una aproximación metodología a EEIIDDFFIA:

— La ISO/IEC 42001:2023 *Information technology Artificial intelligence Management system*, que especifica los requisitos para establecer, implementar, mantener y mejorar de forma continua un sistema de gestión de inteligencia artificial (SGIA) en las organizaciones del que creo destacable el ANEXO D que hace referencia a algo que será habitual como la integración con otras normas ISO como la 27001, la 27701 o la 9001.

— ISO/IEC TR 24030:2021 *Information technology Artificial intelligence (AI)* que ofrece una recopilación de casos de uso de la inteligencia artificial (IA) en diversos ámbitos será reemplazada por ISO/IEC TR 24030, que está en proceso de publicación.

— ISO/IEC 22989:2022 *Information technology —Artificial intelligence— Artificial intelligence concepts and terminology*, que establece la terminología y describe los conceptos de la IA.

— La ISO/IEC 23894:2023, *guidance on risk management in AI*, que proporciona orientaciones para gestionar los riesgos relacionados con la Inteligencia Artificial (IA) en las organizaciones.

— Y la reciente ISO/IEC TR 5469:2024 *Artificial Intelligence Functional safety and AI systems* que describe las propiedades, los factores de riesgo, los métodos disponibles y los procesos relacionados con el uso de sistemas de IA para diseñar y desarrollar funciones relacionadas con la seguridad.

— Asimismo, hay que tener en cuenta la UNE CEN/CLC ISO/IEC/TR 24027:2023 que aborda los sesgos en relación con los sistemas de IA.

Todas ellas pueden: Por un lado, ayudar a disponer de músculo entorno a ciertos conceptos respecto de la IA; pero —sobre todo— a apoyarse en estándares de metodología de gestión de riesgos y de sistemas de gestión que tienen una visión y alcance global y pueden complementar el RIA, tanto para el AARRIA al que se refiere el artículo 9 RIA como para las FRIA que son objeto de este capítulo.

Además, la ISO/IEC 42002:2023 nos proporciona un concepto de Evaluación de impacto del sistema de IA como un «*proceso formal y documentado mediante el cual una organización que desarrolla proporciona o utiliza productos o servicios que utilizan inteligencia artificial identifica, evalúa y aborda las repercusiones sobre las personas, los grupos de personas, o ambos y las sociedades*».

IV. La evaluación de impacto sobre los derechos fundamentales en sistemas de inteligencia artificial de alto riesgo en el Reglamento

1. Evolución, tramitación y contenido final de los artículos del Reglamento implicados

Las FRIA han sido una introducción «reciente» en el proceso legislativo puesto que aparecieron por primera vez en la versión del Parlamento. Un primer cambio

que se aprecia en la versión final es que, a diferencia de la versión del Parlamento donde los obligados a realizarla eran los responsables del despliegue de los sistemas de alto riesgo, se ha circunscrito a determinados responsables del despliegue: los organismos de Derecho público u operadores privados que presten servicios públicos y los operadores que despliegan sistemas de alto riesgo contemplados en el anexo III, punto 5, letras b) y d) (a los que después haremos referencia) y que no entendemos acertada porque significa excluir de esta obligación a numerosos supuestos en el ámbito privado que, en nuestra opinión, debieran de estarlo:

— En cuanto al contenido de la evaluación, la estructura sigue siendo esencialmente el mismo, aunque se han realizado ciertos ajustes en la versión final:

— Por un lado, en la descripción, y además de la finalidad prevista del sistema de IA, se añade la necesidad de hacer referencia a los procesos del implantador en los que se utilizará el sistema. Se trata de una previsión lógica porque el entendimiento del proceso seguido en un sistema de IA es fundamental cuando hablamos de evaluar riesgos y medidas. Asimismo, se añade al período de tiempo en que utilizará la frecuencia, eliminándose la referencia al ámbito geográfico. La referencia al ámbito geográfico puede ser irrelevante en el marco de las evaluaciones que se realicen conforme al RIA porque, sea cual sea su alcance, los derechos que se aplican en el análisis son los derechos fundamentales contemplados en la CFDUE, que es lo que regula el RIA. Pero si se realiza una FRIA de un alcance mayor, tanto geográfico, como si se pretende añadir aspectos éticos y sociales más allá del alcance mínimo del RIA, la referencia y conocimiento del ámbito geográfico y por tanto de dicho contexto será necesaria.

— Por otro lado, y en cuanto a la periodicidad en que deberá de llevarse a cabo la FRIA, la versión final no discrepa mucho de la del parlamento pues ambas coinciden en que *«se aplicará al primer uso del sistema de IA de alto riesgo y que el implantador podrá, en casos similares, basarse en evaluaciones de impacto sobre los derechos fundamentales realizadas previamente o en evaluaciones de impacto existentes llevadas a cabo por el proveedor»*. También coinciden en que si, durante el uso del sistema de IA de alto riesgo, el implementador considera que ya no se cumplen los criterios enumerados en el apartado 1 y que llevaron a tener que realizarla, habrá que realizar nuevas acciones, pero a diferencia de la versión del parlamento que decía que se *«llevará a cabo una nueva evaluación de impacto en materia de derechos fundamentales»* la versión final dice que el implantador *«tomará las medidas necesarias para actualizar la información»*.

— Se simplifica la parte relativa a los riesgos, al llevar la referencia a los derechos fundamentales como objeto de análisis al inicio del artículo; y, asimismo y con ello, poniéndolos —como corresponde— en el epicentro del análisis. Y también modifica y clarifica que esos riesgos lo pueden ser para las categorías de personas físicas y grupos que puedan verse afectados por su utilización en el específico contexto y no únicamente, como mencionaba la versión del parlamento, para las *«personas marginadas o en los grupos vulnerables»*, que desde luego podrán ser un subconjunto de aquellos. Por último, en ese punto añade, en nuestra opinión de forma acertada y como después haremos referencia, que en la evaluación de impacto el responsable del despliegue debe de tener en cuenta la información facilitada por el proveedor con arreglo al artículo 13.

— En la versión final se elimina la referencia de la versión del parlamento a que en la misma se hará costar «*el impacto adverso razonablemente previsible del uso del sistema sobre el medio ambiente*». Se trata de una inclusión que —como también hemos indicado en este capítulo— aparece en diversas partes del RIA como refuerzo, aunque la protección del medio ambiente ya es en sí mismo un derecho fundamental que contempla la CEDF, por lo que no es necesaria.

— El último aspecto que destacar en relación con el contenido es que la versión del parlamento hacía referencia a que había que incluir un plan detallado sobre cómo se mitigarán los perjuicios y el impacto negativo sobre los derechos fundamentales identificados. Esta referencia ha sido eliminada del texto final, pero el análisis de riesgos de una evaluación de impacto lleva implícito ese plan de acciones, si bien la referencia única a la mitigación como medida de gestión del riesgo es reduccionista puesto que, como después se dirá, mitigar es efectivamente unas de las formas de tratar el riesgo, la más habitual, pero no la única.

— Ambos textos contemplan expresamente, dentro de las medidas a adoptar, tanto el sistema de gobernanza como la supervisión humana. No obstante, la versión final:

a) por un lado, ha añadido la referencia a los mecanismos de denuncia aspecto que entiendo sea una obligación legal, quizá una ampliación del ámbito subjetivo de la Directiva y Leyes estatales que la desarrollan por razón no sólo de las materias objeto de denuncia sino por el hecho en sí mismo del sistema de IA utilizado, pero no entiendo su ubicación en una evaluación de impacto, más allá de reforzar (o establecer su exigencia, si es el caso).

b) Y ha eliminado la referencia que hacía la versión del parlamento a la tramitación de reclamaciones y la reparación, cosa que me parece lógica puesto que siendo aspectos que contemplan el RIA y a los que se refieren otros capítulos de esta obra, no tiene sentido que se contengan en una evaluación de impacto.

— Por otro lado, la versión final ha eliminado la referencia de la versión del parlamento a que, «si no puede identificarse un plan detallado para mitigar los riesgos señalados en el curso de la evaluación, el implantador se abstendrá de poner en uso el sistema de IA de alto riesgo e informará de ello sin demora injustificada al proveedor y a la autoridad nacional de supervisión». En este sentido:

A) Por un lado, la referencia a que se abstendrá de ponerlo en uso realmente es una previsión lógica, pero que puede resultar redundante porque es obvio que si no hay un plan de tratamiento de los riesgos y además —cabría añadir— no se es capaz de bajar el umbral de riesgo hasta un riesgo aceptable que se haya definido no puede utilizarse el sistema.

B) Y, por otro lado, hay que distinguir:

— En relación con la previsión que existía de comunicación a la autoridad en estos supuestos, y que se ha eliminado, entendemos que el objeto era el mismo que contempla el RGPD con las consultas previas de las EIPD²⁴, aunque me temo que la

24. Como apunta el Considerando 94 del RGPD «*Debe consultarse a la autoridad de control antes de iniciar las actividades de tratamiento si una evaluación de impacto relativa a la protección de datos muestra que, en ausencia de garantías, medidas de seguridad y mecanismos destinados a mitigar los riesgos, el tratamiento entrañaría un alto riesgo para los derechos y*

experiencia del RGPD en este punto —creo— no ha constatado su utilidad práctica, al menos en España (revisar).

— Y en cuanto a la comunicación de dichos supuestos al proveedor, el único sentido que entiendo podría tener es porque se ha advertido que es un riesgo derivado, no del despliegue, sino del producto que provee el mismo. Y quizá por ello también decía la versión del parlamento en este artículo que *«las autoridades nacionales de supervisión, de conformidad con los artículos 65 y 67, tendrán en cuenta esta información cuando investiguen sistemas que presenten un riesgo a nivel nacional²⁵»*. Y quizá por ello, al referirse a los proveedores y no implantadores, y dado que para ellos dicha consulta se mantiene en los citados artículos, se ha eliminado de este artículo que afecta a los implantadores.

Asimismo la versión del parlamento contemplaba la obligación, que se ha eliminado, de las FRIA de que (con excepción de las pymes, que indicaba que lo podrían hacer voluntariamente así como en algún supuesto relativo a las autoridades públicas) el implantador *«lo notificara a la autoridad nacional de supervisión y a las partes interesadas pertinentes y, que “implicara a los representantes de las personas o grupos de personas que puedan verse afectados por el sistema de IA”»* y mencionaba algunos ejemplos como: *«los organismos de igualdad, los organismos de protección de los consumidores, los interlocutores sociales y los organismos de protección de datos, con vistas a recibir aportaciones para la evaluación de impacto»*. Asimismo, y como mencionamos en otro apartado de este capítulo se debe de implicar a las partes interesadas en la FRIA, pero no se prevé la obligación de publicar o comunicarles los resultados, lo que, como después veremos, puede ser una buena práctica.

— También la versión del parlamento obligaba a la cuando el responsable del despliegue fuese una autoridad pública o una empresa de las mencionadas en el artículo 51, apartado 1 bis, letra b (*«implementadores que sean empresas designadas como guardianes de acceso en virtud del Reglamento (UE) 2022/192»*), publicaran un resumen de los resultados de la evaluación de impacto. Esta obligación también ha desaparecido y al igual que sucede con la publicación y comunicación a las partes interesadas del resultado, y como después mencionaremos y —obviando partes que puedan ser sensibles o en modo resumen— puede ser una buena práctica también. En cambio, la versión final ha añadido una novedad posterior que no preveía la del parlamento: La obligación del responsable del despliegue de *«notificar a la autoridad de vigilancia del mercado los resultados de*

libertades de las personas físicas, y el responsable del tratamiento considera que el riesgo no puede mitigarse por medios razonables en cuanto a tecnología disponible y costes de aplicación» y en consecuencia el artículo 36.1. RGPD dispone que *«El responsable consultará a la autoridad de control antes de proceder al tratamiento cuando una evaluación de impacto relativa a la protección de los datos en virtud del artículo 35 muestre que el tratamiento entrañaría un alto riesgo si el responsable no toma medidas para mitigarlo»* para *«poder asesorarle»*.

25. El artículo 67.1 del RIA hace referencia a que cuando, habiendo realizado una evaluación con arreglo al artículo 65, previa consulta a la autoridad pública nacional pertinente a que se refiere el artículo 64, apartado 3, la autoridad compruebe que, aunque un sistema de IA de alto riesgo es conforme con el presente Reglamento, presenta un riesgo para la salud o la seguridad de las personas, los derechos fundamentales u otros aspectos de la protección del interés público, exigirá al operador pertinente que adopte todas las medidas adecuadas para garantizar que el sistema de IA en cuestión, cuando se comercialice o se ponga en servicio, deje de presentar ese riesgo sin demora indebida, en un plazo que podrá fijar.

la evaluación, presentando la plantilla cumplimentada a que se refiere el apartado 5 como parte de la notificación», con alguna excepción. En el caso contemplado en el artículo 47, apartado 1 (que se refiere a sistemas de IA relacionados con seguridad pública o de protección de la vida y la salud de las personas, protección del medio ambiente y protección de activos industriales e infraestructurales clave) Y en consonancia con esta novedad ha añadido un apartado que indica que **«la Oficina de AI elaborará un modelo de cuestionario, incluso mediante una herramienta automatizada, para facilitar a los usuarios el cumplimiento de las obligaciones del presente artículo de forma simplificada»**.

— Por último, y por lo que respecta a la relación, con las EIPD cuando el sistema de IA supusiese un tratamiento de datos personales, la versión del parlamento indicaba que el desplegador realizaría la EIPDDFFIA conjuntamente con la evaluación de impacto relativa a la protección de datos y que **«la evaluación de impacto relativa a la protección de datos se publicaría como adenda», de aquella**. La versión final ha mantenido el hecho de que se realicen conjuntamente pero no es baladí que realice varios cambios significativos:

— Por un lado, parte de que el supuesto para que la evaluación sea conjunta es si alguna de las obligaciones establecidas para las EIDDDFF ya se cumple mediante la EIPD, porque entendemos que está asumiendo la posible relación entre sistemas de IA y tratamiento de datos personales y posible necesidad de realizar una EIPD pero también, como después profundizaremos, que —habida cuenta del desarrollo y musculación de las EIPD— es posible que conjuntamente signifique obviamente coordinadas pero no necesariamente juntas.

— Y, quizá también por ello, elimina la referencia a que la EIPD se publicaría como adenda a la FRIA, tanto porque es posible que estando coordinadas sea un informe a parte como por el hecho de que se ha eliminado la obligación de publicación de las FRIA del texto final y no existe obligación por parte de la legislación sobre protección de datos de publicar las EIPD, sin perjuicio de que pueda considerarse buena práctica publicar partes de ella o extractos, gestionando los riesgos tanto para la seguridad de la información, del sistema de IA como de otros derechos e intereses legítimos de la organización, como pueda ser los secretos comerciales, por ejemplo.

2. ANÁLISIS DE LAS FRIA EN EL REGLAMENTO

A. Alcance subjetivo: ¿Quién es el obligado a llevarla a cabo? y quienes intervienen en la misma?

Como ya se ha comentado a lo largo de esta obra el RIA contempla diferentes operadores que existen en la cadena de un sistema de IA (el proveedor, el fabricante del producto, el implantador, el representante autorizado, el importador o el distribuidor) y con diferentes obligaciones.

En el caso de las evaluaciones de impacto el obligado a realizarla es el responsable del despliegue. El RIA matiza que esta obligación aplica a determinados responsables del despliegue concretos: los «organismos de Derecho público u operadores privados que presten servicios públicos y los operadores que desplieguen sistemas de alto riesgo a que se refieren las letras b) y c bis) del punto 5 del anexo III»).

Por tanto:

1. Por un lado deberán realizarla los organismos de Derecho Público (importante tener en cuenta las Leyes 39 y 40/2015) respecto de todos los sistemas de IA que sean de alto riesgo.

2. Por otro lado, los operadores privados que presten servicios públicos (de nuevo importante tener en cuenta las Leyes 39 y 40/2015) respecto de todos los sistemas de IA que se refieran a esos servicios públicos. De hecho, el Considerando 96 pone algunos ejemplos²⁶ pero que no pueden entenderse como un *numerus clausus*.

3. Y por otro lado (independientemente del carácter público o privado de dichas entidades) y por razón del objetivo de los sistemas, determinados operadores que desplieguen sistemas de alto riesgo a que se refieren las letras b) y c bis) del punto 5 del anexo III y que (coherente con el Considerando 96 son:

1. Por un lado a los «Sistemas de IA destinados a ser utilizados para evaluar la solvencia de personas físicas o establecer su puntuación crediticia, con la excepción de los sistemas de IA utilizados con el fin de detectar fraudes financieros». El considerando 96 pone como ejemplos las «entidades bancarias o aseguradoras».

2. y por otro lado a los «Sistemas de IA destinados a evaluar y clasificar llamadas de emergencia de personas físicas o a utilizarse para despachar o establecer prioridades en el despacho de servicios de primera respuesta de emergencia, incluidos los de policía, bomberos y ayuda médica, así como de sistemas de triaje de pacientes de atención sanitaria de emergencia».

Asimismo, el artículo 27.2. Al igual que el Considerando 86 dispone que «en casos similares el implantador podrá basarse en las evaluaciones de impacto sobre los derechos fundamentales realizadas previamente o en las evaluaciones de impacto existentes llevadas a cabo por el proveedor», por lo que es una clara alusión a que también los proveedores de dichos sistemas de IA citados en el alcance anterior deberán de llevarla a cabo, o quizá haya una confusión terminológica y deban de basarse en los AARRIA que ex. Artículo 9 RIA deben de llevar a cabo los proveedores. En todo caso lo que parece claro es que los proveedores deben de asistir al implantador, al igual que sucede en protección de datos entre responsable y encargado²⁷.

B. Equipo que debe de intervenir en la misma

A la hora de pensar en el equipo que la debe de abordar es fundamental implicar a todos los actores correspondientes y como bien dice Cannataci²⁸ hacer referencia a

26. El Considerando 96 añade «Algunos servicios importantes para las personas que son de carácter público también pueden ser prestados por entidades privadas. Los operadores privados que prestan estos servicios de carácter público están vinculados a tareas de interés público, como en el ámbito de la educación, la asistencia sanitaria, los servicios sociales, la vivienda o la administración de justicia».

27. De la misma forma que sucede en protección de datos, donde la GUIA AEPD sobre AARR y EIPD citada y después de dejar sentada que la obligación de realizarla es del responsable del tratamiento, también menciona la obligación de los encargados de asistirle (en consonancia con lo dispuesto en el RGPD (considerando 95) «El encargado del tratamiento debe asistir al responsable cuando sea necesario y a petición suya, a fin de asegurar que se cumplen las obligaciones que se derivan de la realización de las evaluaciones de impacto relativas a la protección de datos y de la consulta previa a la autoridad de control».

28. Vid. Cannataci en su prólogo a la obra de Mantelero, A. ob. Cit. «Desde hace más de cuarenta años, hemos ido abandonando gradualmente el enfoque monodisciplinar en la reso-

un enfoque holístico se ha convertido en una especie de cliché, pero quizá el contexto que lo requiera más que otro sea precisamente la IA.

Es verdad que son muy diversos los perfiles que —habida cuenta— del alcance y posible impacto de la IA pueden verse afectados, pero vamos a intentar realizar una aproximación pragmática y funcional.

Para ello quizá lo primero que hay que entender es que habrá roles que serán necesarios siempre; y —en cambio— otros serán contingentes en función del sistema de IA al que se refiera. Por otro lado, que habrá algunos perfiles estarán especializados en la materia; y otros transversales que aportarán su visión desde su ámbito de competencia. Y, por último, que también la intensidad de la intervención de estos puede ser dispar: en algunos casos intervendrán a lo largo de todo el proceso; y en otros casos en momentos puntuales. Un buen ejemplo de propuesta con esta visión aterrizada lo podemos ver en el FRRIA²⁹ de la que, destacaría que, además de los diversos perfiles que cita, contempla que el responsable de proyecto (cosa obvia), y al responsable del área de conocimiento a la que se refiera el algoritmo (lo que podemos entender referido al área propietaria del mismo si tiene un propósito concreto) también implica al *legal advisor* implica en todas las fases.

Lo anterior nos lleva a dos reflexiones: Por lo que respecta al área propietaria del algoritmo y su participación, que se puede producir un conflicto de interés (también en otros roles) por el evidente interés en que «salga adelante» el sistema que (probablemente) ha solicitado y «le interesa» al área propietaria, por lo que una buena práctica para gestionar dicho conflicto, además de establecer normas de gestión de conflicto de intereses, es garantizar la transparencia y documentar el proceso de toma de decisiones. Asimismo, otro aspecto que llama la atención en dicha aproximación es la participación en todo el proceso del asesor legal, lo que nos da una visión de la importancia de la participación de perfiles jurídicos en estos supuestos en donde hablamos de una evaluación de impacto que pone el foco en los derechos humanos.

En mi opinión creo que también un científico de datos debiera de participar en todo el proceso puesto que, tanto para analizar como para poder gestionar riesgos que afectan a derechos pero que pasan por el uso de una tecnología que es moldeable, hay que conocer sus potencialidades para ser «balceada y configurada»

lución de problemas para adoptar un enfoque multidisciplinar, a menudo acompañado de un enfoque interdisciplinar. La perspectiva obtenida en la intersección de varias disciplinas puede ser también profundamente más precisa y más práctica/pragmática que la que se ve limitada por los conocimientos y las prácticas de una sola disciplina. De hecho, la propia noción de HRESIA implica tener en cuenta la perspectiva de otras disciplinas ajenas al Derecho de los Derechos Humanos, la ética y el impacto social. Las ciencias de la computación, las tecnologías aplicadas, la economía y la psicología social son sólo algunas de las otras disciplinas que me vienen a la mente y que deben estar profunda y constantemente involucradas en la forma en que la sociedad debe pensar en la IA. Hablar de “un enfoque holístico” se ha convertido en una especie de cliché, pero es difícil pensar en un contexto que lo requiera más que la IA... y ése es básicamente el núcleo del mensaje de la obra actual de Mantelero».

29. El FRRIA menciona diversos perfiles en función de las fases, Los perfiles que mencionan son: Interest Group, Management, Citizen panel, CISO o CIO, Communications specialist, Data scientist, Data controller or data source owner, Data protection officer, HR staff member, Domain Expert, Legal Advisor, Algorithm developer, Commissioning client, Project leader, Strategic ethics consultant y Other project team members.

en función del riesgo. Desde luego la intervención de asesores éticos es un rol que, heredado de aproximaciones éticas hacia la IA ha ido calando y algunas empresas ya han nombrado asesores o comités éticos y aprobado directrices éticas adicionales y normalmente en consonancia con otros principios internacionales. Qué duda cabe de la importancia de considerar la ética, pero no es una aproximación de la exigencia del RIA, Y esto dependerá del enfoque y consecuente alcance que se acuerde para la EI, si se centra sólo los derechos humanos o abraza también los aspectos éticos.

Otro aspecto para considerar es si nos encontramos con un sistema de IA para uso interno (donde deberemos de considerar los roles y áreas internas, sin perjuicio obviamente de poder contar con asesores externos y siempre deberemos de considerar a las partes interesadas, como indica el Considerando 64³⁰ a semejanza de lo que hace el RGPD³¹, y que —en materia de IA y en términos de la ISO /IEC 42001:2023 y en consonancia con la ISO/IEC 22989:2022— los define como la *«persona u organización que puede afectar, verse afectada o percibirse afectada por una decisión o actividad»*.

Asimismo, y para completar la foto de los intervinientes hay que considerar si se trata de un sistema de IA como un producto para clientes. En estos casos, además de las obligaciones que dispone el RIA en el caso de productos y por lo que respecta aquí al perfil de los implicados en la EI DFFIA decir que habrá que implicar a los perfiles necesarios en función del producto y destinatarios. Un ejemplo gráfico es el famoso caso de Hello Barbie³², aunque poco tenga que ver con los supuestos de exigencia de realización que hemos indicado, pero que si nos sirve para ilustrar por ejemplo que

-
30. El Considerando 64 a del RIA indica que, a la hora de identificar las medidas de gestión de riesgos más adecuadas, el proveedor deberá documentar y explicar las decisiones tomadas y, cuando proceda, implicar a expertos y partes interesadas externas.
31. Artículo 35.9. RGPD: *«Cuando proceda, el responsable recabará la opinión de los interesados o de sus representantes en relación con el tratamiento previsto, sin perjuicio de la protección de intereses públicos o comerciales o de la seguridad de las operaciones de tratamiento»*.
32. Mantelero, Alessandro, en ob. cit p.61 cita el ejemplo real de Hello Barbie. Era una muñeca interactiva producida por Mattel para el mercado anglosajón, equipada con sistemas de reconocimiento de voz y funciones de aprendizaje basadas en la IA, que funcionaba como un dispositivo IoT. La muñeca podía interactuar con los usuarios, pero no con otros dispositivos IoT. El objetivo del diseño era proporcionar una conversación bidireccional entre el muñeco y los niños que juegan con él, incluyendo capacidades que hacen que el muñeco pueda aprender de esta interacción, por ejemplo, adaptando las respuestas al historial de juego del niño y recordando conversaciones anteriores para sugerir nuevos juegos y temas. La muñeca ya no es comercializada por Mattel debido a varias preocupaciones sobre la seguridad del sistema y del dispositivo. Por citar sólo uno de sus riesgos asociado al tema de perfiles que estamos tratando aquí: En preguntas frecuentes sobre Hello Barbie («P: ¿Puede Hello Barbie decir el nombre de un niño? No. Hello Barbie no pregunta el nombre de un niño y no está programada para responder con el nombre de un niño, por lo que no podrá recitar el nombre de un niño»). Pero Mantelero cita una respuesta en el diálogo con la muñeca: «Barbie: A veces me pongo un poco nerviosa cuando le digo a la gente mi segundo nombre. ¡Pero estoy muy contenta de habértelo dicho! Cuál es tu segundo nombre?». Con este ejemplo se pone de manifiesto que, en este caso hace falta un equipo que trabaje frases y diálogos, y para ello posiblemente haya que acudir —añado— a psicólogos especialistas en educación infantil o perfiles similares. Para ilustrarlo con dicho ejemplo cita lo siguiente:

si se quiere poner en marcha un servicio de IA en el ámbito de la Educación infantil es posible que haya que implicar a psicólogos y/o psicopedagogos, por ejemplo.

C. ¿Cuándo se realiza?

El apartado 1 del artículo 27 del RIA dispone que esta debe de realizarse «antes de desplegar un sistema de IA de alto riesgo» y el apartado dos añade que la misma «aplica al primer uso del sistema de IA de alto riesgo». En algún modelo, como el de la RIAT Canadiense se contempla un ulterior check de la evaluación antes de la puesta en producción,³³ y quizá, esto aunque parezca lógico, podría haberse apuntalado en el texto legal, puesto que en definitiva la FRIA define unos controles que hemos de verificar antes de ponerlo en marcha.

Asimismo, el artículo 27 en su apartado dos añade: «Si, durante el uso del sistema de IA de alto riesgo, el implantador considera que alguno de los factores enumerados en el apartado 1 cambian o ya no están actualizados, el implantador tomará las medidas necesarias para actualizar la información».

A los efectos de tener en cuenta qué supone factores, el Considerando 96 y el artículo 27 2 clarifican que los factores que debe de tenerse en cuenta si han cambiado son los que indica el apartado 1 del mismo artículo, es decir:

1. Los usos o finalidades previstas.
2. El período de tiempo y la frecuencia con que se prevé utilizar.
3. Las categorías de personas físicas y grupos que puedan verse afectados por su uso.
4. Los riesgos específicos de daños que puedan afectar a las categorías de personas o grupos de personas.
5. Las medidas de supervisión humana.
6. Las medidas que deben adoptarse en caso de materialización de estos riesgos, incluidas sus disposiciones en materia de gobernanza interna y mecanismos de denuncia. En el mismo sentido, la ISO 42001:2023 cuando hace referencia a los AARR y también a las Evaluaciones de Impacto del sistema, dice que deben de realizarse «a intervalos planificados o cuando se propongan o produzcan cambios significativos».

Las EIDFFFIA deben de formar parte, a su vez, de un Sistema de Gestión de IA, que al final puede estar basado en un estándar como la ISO 420012: 2023 o no, y a su vez es posible que esté integrado en otro sistema de gestión, pero que como todos ellos se basa en un PDCA, ciclo de Deming o proceso de mejora continua, como también indica el RIA³⁴, lo que —por definición— significa mejora continua y —por tanto— actualización.

33. El RIA Canadiense dice: «El RIA debe completarse al comienzo de la fase de diseño de un proyecto. Los resultados del RIA guiarán los requisitos de mitigación y consulta que se cumplirán durante la implementación del sistema de decisión automatizado según la directiva.

El RIA debe completarse por segunda vez, antes de la producción del sistema, para validar que los resultados reflejen con precisión el sistema que se construyó.»

34. El Considerando 42 a RIA dispone: «El sistema de gestión de riesgos debe consistir en un proceso continuo e iterativo que se planifique y ejecute a lo largo de todo el ciclo de vida de un sistema de IA de alto riesgo. Este proceso debe tener por objeto identificar y mitigar los riesgos

D. ¿Sobre qué se realiza? Alcance sustantivo. Una preEIA

Cuando vamos a realizar un AARR al igual que una EI debemos tener claro en primer lugar sobre qué se realiza. Por ejemplo, cuando hablamos de protección de datos todos tenemos claro que hablamos de tratamientos (sin perjuicio de los debates sobre la mayor o menor granularidad de este concepto); cuando hablamos por ejemplo de un AARR en el marco del ENS de sistemas de información. Y siguiendo caso de la protección de datos no todos los tratamientos deben de tener que disponer de una EIPD por lo que hay que realizar lo que coloquialmente se conoce como una PreEIPD o PrePIA y que analiza la necesidad de realizar la EIPD o no.

Lo mismo hay que hacer cuando hablamos de Sistemas de IA hablamos de Evaluaciones de impacto de sobre los derechos fundamentales de los sistemas de IA de alto riesgo en las entidades a las que les es de aplicación, sin perjuicio de que otras no obligadas puedan voluntariamente decidir realizarla.

En este caso, al evaluar si se debe de realizar o no hay dos partes en el objeto de la evaluación:

- a) Que se trate de sistemas de IA de alto riesgo.
- b) Que esos sistemas de IA se refieran a las entidades y/o servicios concretos dentro del alcance subjetivo al que nos hemos referido.
- c) Y —por último— obviamente, que afecten o puedan afectar a Derechos fundamentales. Lo que es un sistema de IA, cuales son de alto riesgo y quienes son las entidades obligadas ha sido abordado en otros capítulos de este libro, pero sí que merece más atención el acotar que nos centremos en qué significa que puedan afectar a derechos fundamentales. Por un lado, es importante indicar que, sin perjuicio de otras aproximaciones que se puedan realizar «a mayores» añadiendo aspectos éticos o sociales la base mínima objeto de la evaluación deben de ser los derechos fundamentales; y, por tanto, la fuente de requisitos debe de provenir de la CEDF; si se quiere, coherente con los derechos fundamentales que contempla la CE, de casi absoluta coincidencia como hemos mencionado. Asimismo, es importante aclarar que, en el ámbito europeo, a diferencia de la DUDH, y gracias al reconocimiento específico por parte de la CEDF, la inclusión de la protección del medio ambiente queda clara, aunque en el caso de España no estaría incluido como derecho fundamental, ya que la mención a la protección del medio ambiente se realiza en el capítulo III de la Constitución, y hay un debate sobre ello³⁵, quizá por esa falta de homogeneidad a nivel Europeo, el RIA «zanja el debate» sobre su inclusión. Pero es

pertinentes de los sistemas de inteligencia artificial para la salud, la seguridad y los derechos fundamentales. El sistema de gestión de riesgos debe revisarse y actualizarse periódicamente para garantizar su eficacia permanente, así como la justificación y documentación de todas las decisiones y acciones significativas adoptadas con arreglo al presente Reglamento».

35. Si que podría sostenerse una posible discrepancia del alcance de los derechos en CE y en contexto Europeo, profundizando en sí, en el caso Español, deberemos extender tales análisis a todo el Capítulo II del Título I o solo a la sección 1ª, y/o en su caso marchar contra la Carta de los Derechos Fundamentales de la UE + CEDH. Si se asume que el medio ambiente queda incluido, entonces, en la parte del artículo 9 «gestión de riesgos» donde habla de medio ambiente/salud y luego DDF, quizás debería enfocarse a integrar AARR de medio ambiente más «regulados» sea en el contexto de las exigencias sectoriales concretas/administrativas/ISO, etc.

que además existe una íntima relación entre la posible afectación de la IA al medio ambiente y viceversa³⁶, que viene constatada por las diversas referencias al medio ambiente en el RIA y en algunas de las ISO que abordan la IA.

Esta necesidad de acotar el alcance de los posibles derechos afectados por el sistema de IA nos obligará a realizar una PreEIA, de forma similar a como en protección de datos venimos realizando prePIAs.

En esta fase quizá no sólo podamos acotar los derechos afectados, sino incluso plantearnos si la relevancia o peso desmedido de alguno de ellos en el proyecto, aconseja realizar una evaluación de impacto específica sobre esa materia, separada de la general FRIA. No obstante, y en mi visión, debería de intentar tenderse a realizar FRIA integradas; sin perjuicio de la posible excepción —quizá más habitual— de las relativas a protección de datos y seguridad de la información³⁷, que ya tienen no sólo un desarrollo y roles maduros y que, aunque después comentaremos deben de realizarse «conjuntamente» y coordinadas, pero no necesariamente ello supone que unidas «del todo».

Otro aspecto para considerar, de cara a poder hacer factible el análisis de dichos derechos, es agrupado por áreas lo que también, como después veremos, se puede a su vez enlazar con los colectivos o grupos impactados³⁸.

Qué duda cabe que cuando uno realiza una EI en la que puede haber tantos derechos afectados debe de realizar un análisis de diversas fuentes de requisitos y sus consiguientes controles para cada uno de los derechos, lo que sitúa a quienes la realizan a una tarea titánica, puesto que hablar de derechos supone conocer no sólo su descripción sino su aterrizaje en legislación de desarrollo, directrices y resoluciones de las autoridades administrativas y judiciales etc., incluso con la ayuda de un equipo multidisciplinar como hemos mencionado.

Aunque el RIA ha mantenido una visión centrada en los derechos fundamentales sin incluir como obligatorios los aspectos éticos y sociales, aspecto que ha sido criticado por parte de la doctrina³⁹, lo cierto es que estos pueden y será habitual que se

36. Como dice el Relator Especial de las Naciones Unidas sobre los derechos humanos y el medio ambiente «Todos los seres humanos dependen del medio ambiente en el que viven. Un entorno seguro, limpio, saludable y sostenible es indispensable para el pleno disfrute de una amplia gama de derechos humanos, entre otros el derecho a la vida, la salud, la alimentación, el agua y los saneamientos.

En ausencia de un medio ambiente saludable, somos incapaces de realizar nuestras aspiraciones. Y quizá ni siquiera logremos acceder a los criterios mínimos de dignidad humana». <https://www.ohchr.org/es/special-procedures/sr-environment/about-human-rights-and-environment> Último acceso el 12/03/2023.

37. De hecho, la ISO 42001:2023 las pone como ejemplo.

38. Por ejemplo Telefónica en su EIDH que se puede consultar en <https://www.telefonica.com/es/sala-comunicacion/reportes/el-proceso-de-debida-diligencia-de-telefonica-en-ddhh-y-medioambiente/amp/> los agrupa en las siguientes 5 áreas: Ética y Gobernanza, cadena de valor, operaciones, recursos humanos y productos y servicios).

39. Mantelero, A. en ob. cit p. 173. ya decía «... tras varios años de debate sobre la dimensión ética de la IA, la visión predominante parece ser la de delegar las cuestiones éticas a otras iniciativas no integradas en la evaluación jurídica. De la misma manera que centrarse exclusivamente en la ética era crítico, esta falta de integración entre los impactos legales y sociales de

añadan, especialmente en organizaciones que —más allá de los derechos humanos— ya están abordando dichas cuestiones.

A todo lo anterior se deberá de añadir, caso habitual en grandes corporaciones, la normativa interna en dichas materias que no podrá restar, pero si incrementar requisitos y sus correspondientes controles.

V. PASOS A REALIZAR EN UNA FRIA

Una FRIA es, en sí mismo, un proceso que tiene diversas fases dispuestas en un PDCA y que a su vez se pueden integrar en el plan de acción en un sistema de gestión de inteligencia artificial; y este, a su vez, podrá y será habitual que se integre en otros sistemas de gestión, como hemos adelantado. No obstante, aquí analizaremos el PDCA en sí mismo que constituye la propia EIPDFFIA. En el fondo las diversas Metodologías de Evaluaciones de Impacto en las que ha habido «tradición» (por ejemplo, las de Derechos Humanos o las EIPD) tienen una estructura similar, ampliamente aceptada; sin perjuicio de matices propios que pueda haber: bien por razón de quien la propone; o bien por razón de la materia evaluada.

También es importante precisar que la realización de una FRIA, al igual que sucede con las EIDH, podría realizarse de forma integrada o no⁴⁰ con otras EI, lo que, al igual que en aquella, plantea también ventajas e inconvenientes.

Asimismo, hay que considerar que puede haber diferentes ejemplos de Sistemas de IA: desde ejemplos dedicados a propósitos concretos (supuestos de IA que den soporte en procesos o áreas concretas) como otros relacionados con sistemas más complejos (como las *Smart Cities*). Ello puede suponer que, aunque se utilice la misma metodología, y en función de la dimensión del sistema de IA, haya que realizar ajustes o «adiciones». Por ejemplo, en los supuestos en que se combinan muchos sistemas de IA es posible que haya que realizar una Evaluación de Impacto adicional, para disponer de una visión global.

Iremos incardinando, dentro de cada una de las fases propuestas, el contenido exigido por el RIA; dado que, para disponer del mismo, deberá extraerse de ellas.

Como hemos adelantado, y dado que la FRIA en sí misma es un PDCA, la primera fase es la de Plan. La fase de Plan, a su vez se compone de diferentes fases que desembocarán en el Informe de FRIA donde, entre otras cuestiones y como veremos, se definirán las correspondientes acciones de tratamiento del riesgo.

Fase 1: Análisis preliminar sobre la necesidad de realizar una EIDDF y concreción de los Sistemas y derechos afectados (determinación inicial del alcance)

En primer lugar, dispondremos de un inventario de los sistemas dentro del alcance. En cuanto a los sistemas que deben de ser objeto de FRIA son los de alto riesgo, y en este libro ya se ha abordado esta cuestión.

la IA es problemática. Un modelo de evaluación integrado, como el HRESIA, podría superar esta limitación en consonancia con el modelo basado en el riesgo propuesto».

40. Vid. El apartado A.6.8. que se titula «¿Deberían ser las EIDH independientes o integradas?». p. 27 del https://www.humanrights.dk/sites/humanrights.dk/files/media/document/HRIA%20Toolbox_INTRO_Spanish.PDF última consulta realizada el 13/03/2024.

Por lo que respecta a los derechos fundamentales en los que debe centrarse la evaluación, también hemos indicado en este capítulo a cuáles se refiere. La realización de este análisis se puede realizar sobre la base del conocimiento del Sistema de IA, del RIA y a nivel operativo con un *check list* sin mucho trabajo de campo ni implicación de las partes interesadas.

Fase 2: Contexto, planificación y detalle dentro del alcance

Antes de iniciar una FRIA debemos de disponer de la información necesaria. En diversas metodologías esto se aterriza de forma diferente⁴¹.

A nuestro entender los grandes aspectos a conocer son:

1. Determinar el equipo que la va a llevar a cabo, lo que tiene que ver con determinar de entre los posibles intervinientes a los que nos hemos referido en este capítulo, cuales se requieren en este caso, pues algunos perfiles serán siempre necesarios y otros dependerá del caso.

2. Determinar el detalle dentro del alcance de la FRIA, profundizar en ciertos aspectos a partir del alcance previo establecido) lo que implica:

a) Conocer el tipo de proyecto o actividades de la organización a las que afecta el sistema de IA.

b) Disponer una comprensión suficiente sobre el Sistema de IA, particularmente de la información que maneja en relación con los posibles derechos afectados, dado que los siguientes aspectos deberán de reflejarse en el oportuno informe:

i) una descripción de los procesos del implantador en los que se utilizará el sistema de IA de alto riesgo de acuerdo con su finalidad prevista;

ii) una descripción del período de tiempo y la frecuencia con que se prevé utilizar cada sistema de IA de alto riesgo

iii) las categorías de personas físicas y grupos que puedan verse afectados por su uso en el contexto específico;

c) Disponer de un conocimiento correcto del contexto.

d) Identificar las partes interesadas relevantes: partes interesadas en sentido estricto, así como titulares de los derechos, titulares de los deberes, otras partes relevantes.

3. Determinar la metodología a emplear y las fuentes de requisitos a utilizar.

41. Por ejemplo, en el caso del RIA canadiense se dice que antes de empezar es útil tener información sobre: Prácticas institucionales de gestión de servicios de tecnología de la información (ITSM) la decisión administrativa que informará o tomará el sistema de decisiones automatizado, el contexto en el que se utilizará el sistema y la forma en que el sistema ayudará o reemplazará el juicio de un tomador de decisiones humano los clientes sujetos a la decisión, incluida la evidencia de cualquier vulnerabilidad (por ejemplo, socioeconómica, demográfica, geográfica); el algoritmo, incluidos los parámetros y técnicas de procesamiento de datos, y la salida; los datos de entrada utilizados por el sistema, incluidos detalles sobre el tipo, la fuente, el método de recopilación y la clasificación de seguridad. Las medidas de garantía de calidad planificadas o existentes; medidas de transparencia planificadas para comunicar información sobre la iniciativa a los clientes y al público. Partes interesadas internas y externas a ser consultadas y registro de recomendaciones o decisiones tomadas por el sistema, y cualquier registro o explicación generada por el sistema para dicho registro.

Asimismo, tal como ya se ha dicho, esta obligación de realizar la FRIA la tiene el implantador, pero este podrá basarse en las llevadas a cabo por el proveedor (ex. artículo 27 2. RIA, como ya hemos mencionado) por lo que esta será también parte del conocimiento del sistema de IA.

Por último, y aunque no lo exija el RIA y dado que vamos a considerar la posibilidad de realizar una EIPD conjuntamente con una FRIA, debería de disponerse al menos de una descripción de los datos personales tratados⁴².

Fase 3: Necesidad, proporcionalidad y calidad de los datos

Como es conocido las evaluaciones de impacto en protección de datos incorporan el análisis sobre la necesidad y proporcionalidad del tratamiento. Este aspecto se ha convertido recientemente en y polémico dado que la AEPD en su nueva Guía Sobre tratamientos de control de presencia mediante datos biométricos⁴³ cambia el criterio seguido hasta la fecha y al «tensar el sentido» de que los datos personales sólo deben tratarse si la finalidad del tratamiento no pudiera lograrse razonablemente por otros medios, como indican los compañeros Patricio Monreal y Maria loza «*será prácticamente imposible superar el juicio de necesidad del tratamiento (salvo en supuestos muy concretos y residuales), pues siempre existirán otros medios menos intrusivos e igualmente eficaces*». Y añaden, en relación con dichos tratamientos que impliquen IA, que la propia Guía mencionada indica que «*deberán tener en cuenta las prohibiciones, limitaciones y exigencias establecidas en la normativa de inteligencia artificial, pero, ¿superará el juicio de necesidad en algún momento? Si acudimos al documento de la AEPD[4] «Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción» en este punto se indica que la utilización de soluciones basadas en IA puede conllevar un alto nivel de riesgo por lo que «debería valorarse si el objeto del tratamiento no puede ser conseguido utilizando otro tipo de solución que alcance la misma funcionalidad, con un margen de rendimiento aceptable y un nivel de riesgo menor*».

Por tanto: obviamente los sistemas de IA de alto riesgo que traten datos personales deberán de conllevar al menos este análisis sobre necesidad y proporcionalidad en relación con el tratamiento de datos personales⁴⁴, pero el RIA no realiza una se refiere a este aspecto. No obstante, en mi opinión, al igual que se ha realizado por parte de algunas metodologías⁴⁵, sí que requiere que haya un «momento» en el que se

42. Lo que la EIPD de la AEPD en su guía (citar) denomina «Describir el ciclo de vida de los datos».

43. <https://www.aepd.es/documento/guia-control-presencia-biometrico.pdf>

44. Por exigencia del artículo 35.7.b del RGPD.

45. El RIA pone algunos ejemplos de posibles motivos para introducir la automatización:

1. Acumulación existente de trabajo o casos.
2. Mejorar la calidad general de las decisiones.
3. Menores costes de transacción de un programa existente.
4. El sistema está realizando tareas que los humanos no podrían realizar en un periodo de tiempo razonable.
5. Utilice enfoques innovadores.

6. otros que se puedan especificar También el FRRRIA en su Parte 1 trata sobre el «Por qué» de la intención de desarrollar, comprar, ajustar y/o uso de un algoritmo (en adelante abreviado: el uso de un algoritmo). Cuáles son las razones, ¿Los motivos subyacentes y los efectos previstos del uso del algoritmo? ¿Cuáles son los ¿Valores subyacentes que dirigen el despliegue del algoritmo? Estas preguntas generales deben primero debe discutirse en un proceso de toma de decisiones sobre el uso de

analicen la necesidad y proporcionalidad, valorando los aspectos que se han tenido en cuenta para implantar el sistema de IA, considerando aspectos como: porqué precisamente esa IA y no otra; las consecuencias que tendría el no implantarlo; y, al menos de forma previa y sin entrar en el análisis de riesgos que será el objeto de trabajo en profundidad, que al menos se ha realizado una aproximación previa sobre los beneficios y sacrificios que ello supone. Obviamente ese balance o ponderación tiene grandes diferencias entre el ámbito público y el privado, por ello ese análisis en los supuestos de FRIA que siempre son exigibles asociadas a servicios públicos, tienen mayor sentido aún. Un ejemplo en el ámbito público lo pone la FRRIA, ya citada, con un ejemplo: «*Supongamos, por ejemplo, que un algoritmo es una herramienta eminentemente adecuada y necesaria para mejorar la eficiencia de la toma de decisiones, pero existe un riesgo real de que la herramienta refuerce los patrones discriminatorios. En ese caso, ¿se considera razonable seguir implementando la herramienta? No es posible formular criterios estrictos y objetivos para determinar el peso y el equilibrio los diversos derechos, intereses, objetivos y valores públicos. Generalmente podemos decir, sin embargo, que cuanto más grave sea la infracción esperada de los derechos fundamentales, más graves serán los objetivos sociales pesar en comparación*».

Planteo fruto de las reflexiones con Jordi Morera dos posibles debates que interseccionan con privacidad:

1. Si se está haciendo un análisis de un sistema de Alto Riesgo y se ha notificado a la autoridad tras la FRIA: ¿Hay que entender que automáticamente se supera dicha ponderación en la EIPD del tratamiento objeto de dicho sistema de IA? Quizá la respuesta deba de ser afirmativa puesto que no tendría sentido que una autoridad administrativa (la AEPD), llevara a cabo un pronunciamiento contrario a el despliegue de un sistema de IA, que, aunque sea de Alto Riesgo, ha sido reconocido como tal en el RIA y que además ha sido notificado (y validado) por la Autoridad, so riesgo de conflicto entre distintas autoridades. De aceptar lo contrario se estaría atribuyendo a la AEPD la función de «legislador negativo», pues si indica que un tratamiento no es proporcional por razón de un sistema de IA que el RIA no ha prohibido y que regula con una serie de garantías (los de alto riesgo) se estaría sobrepasando en sus atribuciones y corrigiendo al propio legislador.

algoritmos, antes de obtener a preguntas sobre, por ejemplo, las condiciones previas o el posible impacto en los aspectos fundamentales derechos. Las respuestas dadas a las preguntas de esta parte son relevantes para responder a las preguntas más preguntas específicas en las siguientes partes.

Y después en la Parte 4 al analizar los derechos fundamentales dice: *Necesidad y subsidiariedad*. Para alcanzar los objetivos políticos se puede utilizar una amplia gama de herramientas y medios, entre ellos: algoritmos. Incluso si se elige una herramienta específica, a menudo se puede utilizar de varias maneras. Además, a veces puede resultar posible suavizar los efectos perjudiciales de un determinado instrumento mediante medidas compensatorias o mitigadoras. La elección que se puede hacer entre varias herramientas es fundamental en la cuestión de la necesidad y subsidiariedad de elegir un algoritmo específico.

Equilibrio de intereses/proporcionalidad.

Incluso si un algoritmo parece ser una herramienta adecuada y necesaria para lograr los objetivos formulados; aún así es necesario dar siempre un último paso. Este paso tiene que ver con el peso relativo del derecho fundamental en juego, en comparación con el peso relativo de los objetivos sociales y los valores públicos.

2. No obstante lo anterior, hay otro punto que deberá hacer reflexionar a los DPD: si se utiliza un sistema de IA de Alto Riesgo, donde se tratan datos de carácter personal, automáticamente debemos asumir que se requiere de una EIPD, pues a fin de cuentas, el legislador del RIA ha determinado que dicho sistema genera un alto riesgo para los DDFF, y —por tanto— su valoración es extrapolable a los datos personales que se utilicen para el mismo. Es decir, cuando se trata de un sistema de IA de alto riesgo ello puede entenderse que equivale automáticamente a tratamiento de alto riesgo y consecuente a la realización de una EIPD obligatoria.

Fase 4. Gestión de riesgos

Ya hemos hecho referencia, en varios apartados de este capítulo, a la aproximación de riesgo del RIA y además es objeto de otro capítulo de este libro, por lo que no nos extenderemos en este apartado. La gestión de riesgos es la parte central de cualquier EI y, por tanto, también de las FRIA. Por ello, y aun teniendo en cuenta las diferencias que hemos comentado sobre el alcance y sentido del artículo 9 RIA respecto de la gestión de riesgos y la parte de riesgos que menciona el artículo 27; en esencia la gestión de riesgos tiene una columna vertebral ampliamente consolidada. Para ello, además de tener en cuenta las especificidades del RIA, y tener en cuenta el RIA, lo mejor es inspirarse en criterios aceptados como estándares globales. En este caso de la ISO 31000:2009 Gestión del Riesgo — Principios y Directrices, de la que el resto de las normas ISO se inspiran y adecúan a entornos concretos, además de las ISO específicas en IA a las que nos hemos referido también, particularmente la ISO/IEC 23894:2023, *guidance on risk management in AI*.

Las fases que seguir, con matices (el RIA agrupa identificación y análisis), coinciden, y podemos decir que son:

A) Identificación de los riesgos

Se trata de identificar los riesgos conocidos y razonablemente previsible del sistema de IA puede tener sobre los derechos fundamentales que se hayan definido en el alcance, cuando el mismo se utiliza de acuerdo con su finalidad prevista.

B) Análisis de los posibles escenarios de riesgo

A continuación, hay que estimar la vulnerabilidad del sistema de IA para dichos derechos en dos sentidos: la probabilidad y el impacto.

Aunque el artículo 7 del RIA no se refiere a los AARR ni a las FRIA, sino que menciona criterios a tener en cuenta por la Comisión para evaluar la modificación de los sistemas considerados de alto riesgo en el anexo III, quizá es posible considerar también dichos elementos de cara a la evaluación la probabilidad y el impacto.

a) Por un lado la probabilidad que —según la citada ISO— es la posibilidad de que un evento suceda. Puede definirse medirse o determinarse objetiva o subjetivamente, cualitativa o cuantitativamente, y describirse utilizando términos generales o matemáticos (como una probabilidad matemática o una frecuencia en un período de tiempo determinado). Esta es una posible tabla para calcular la probabilidad de que un sistema de IA afecte cada factor de riesgo identificado para cada derecho humano en el alcance, aunque puede haber otras válidas:

NOMBRE	DESCRIPCIÓN
(MUY BAJA) >= 1 vez cada 100 años	Se produce al menos una vez cada 100 años
(BAJA) >= 1 vez cada 10 años	Se produce al menos una vez cada 10 años
(MEDIA) >= 1 vez al año	Se produce al menos una vez al año
(ALTA) >= 10 veces al año	Se produce al menos 10 veces al año
(MUY ALTA) >= 100 veces al año	Se produce al menos 100 veces al año

b) Por otro lado el impacto o consecuencias de que se materialicen los escenarios del riesgo definidos puede ser cierto o incierto y puede tener efectos directos o indirectos sobre los objetivos. Asimismo, y según la citada ISO puede tener efectos positivos o negativos.

Asimismo, las consecuencias se pueden expresar de manera cualitativa o cuantitativa. Al igual que en el caso de la probabilidad hay diversas escalas que se pueden utilizar, pero una posible forma de representar la posible severidad de las consecuencias que, para los derechos fundamentales analizados, tendría el hecho de que se produjese un evento que afectase al sistema de IA serían la siguientes:

IMPACTO	DESCRIPCIÓN
DESPRECIABLE	Los titulares del derecho no se verán prácticamente afectados o encontrarán alguna pequeña inconveniencia
LIMITADO	Los titulares del derecho podrán encontrar inconveniencias no significativas
SIGNIFICATIVO	Los titulares del derecho encontrarán consecuencias significativas que deberían poder superar sin dificultades serias
MÁXIMO	Los titulares del derecho encontrarán consecuencias significativas o incluso irreversibles, que no podrán llegar a superarse.

C) Estimación y evaluación de los escenarios de riesgo o amenazas

Como dice el RIA (art. 9.2.b.) haciendo a los AARR en general y es aplicable a esta fase de la FRIA, el siguiente paso es estimar y evaluar los riesgos «*que pueden surgir cuando el sistema de IA de alto riesgo se utiliza de acuerdo con su finalidad prevista y en condiciones de uso indebido razonablemente previsibles*». La referencia en la letra c) a «*la evaluación de otros riesgos que puedan surgir, basada en el análisis de los datos recogidos en el sistema de seguimiento postcomercialización contemplado en el artículo 61*» y que obliga a los proveedores a establecer y documentar un sistema de seguimiento postcomercialización de manera proporcionada a la naturaleza de las tecnologías de inteligencia artificial y a los riesgos del sistema de IA de alto riesgo no impacta en

la evaluación inicial sino que es un apuntalamiento de que la metodología requiere de iteraciones.

La fórmula utilizada para la estimación del riesgo es $\text{RIESGO} = \text{PROBABILIDAD} \times \text{IMPACTO}$ (consecuencia). De esta manera, se genera una matriz de riesgo, que, de forma coherente con los umbrales definidos (insistimos que pueden ser otros válidos), podría ser la siguiente:

	Impacto			
Probabilidad	Despreciable	Limitado	Significativo	Máximo
Muy alta	Medio	Alto	Muy alto	Muy alto
Alta	Medio	Medio	Alto	Muy alto
Media	Bajo	Medio	Medio	Alto
Baja	Bajo	Bajo	Medio	Medio
Muy baja	Muy bajo	Bajo	Bajo	Medio

El resultado obtenido de la evaluación es el riesgo inicial.

D) Gestión o tratamiento del riesgo

Ante los riesgos iniciales detectados, deberá adoptarse alguna de las siguientes estrategias o formas de tratar el riesgo, que el RIA en su artículo 9.4 también especifica, de forma concordante con la teoría de gestión de riesgos mencionada:

1. Una opción es mitigar o reducir el riesgo inicial implantando controles que reduzcan el riesgo por debajo del umbral definido como aceptable. Para ello se pueden adoptar dos opciones:

- a) Reducir el impacto causado por un escenario del riesgo.
- b) Reducir la probabilidad de que un escenario del riesgo se materialice.

2. Otra opción es evitar o eliminar el riesgo anulando, excluyendo o sustituyendo el elemento o funcionalidad del diseño. Esta opción no siempre es viable dado que a veces puede provocar la pérdida de alguna funcionalidad esencial.

3. La última opción teórica, según la «teoría general de gestión de riesgos» es ignorar o asumir el riesgo. Es decir, no hacer nada para tratarlo. Teóricamente esto sería posible en tres escenarios:

1. Cuando el impacto o consecuencia sea aceptable.
2. Cuando el riesgo sea aceptable.
3. Y cuando el coste de las medidas a adoptar sea desproporcionado en comparación al impacto y riesgo. Pero: ¿cabe asumir riesgos aceptables en derechos fundamentales?

Hemos de partir de que el RIA ya ha realizado una labor previa de gestión normativa del riesgo prohibiendo ciertos sistemas de IA. Además —como es sabido— el nivel de «riesgo cero» no existe.

Según Mantelero, a diferencia de la noción de riesgo aceptable que «*procede de la normativa sobre seguridad de los productos*⁴⁶ en el ámbito de los derechos fundamentales el principal factor de riesgo es la proporcionalidad y supone la ausencia de riesgo o de los riesgos mínimos» y concluye que «*si aceptamos esta interpretación, la aceptabilidad es incompatible con el alto riesgo de impactos adversos de la IA sobre los derechos fundamentales y cualquier evaluación de impacto basada en una cuantificación de los niveles de riesgo jugará un papel crucial en la gestión del riesgo*». Es cierto que si la ecuación de riesgo se compone de dos factores (probabilidad e impacto) y el impacto en derechos fundamentales consideramos que es siempre alto, mantener dicha postura citada nos llevaría a negar de plano el uso de la metodología de gestión de riesgos que propugna el RIA. Pero el propio artículo 9 4. del RIA anula esa visión y valida —por tanto— la existencia de riesgos aceptables al indicar que las medidas de gestión de riesgos «*serán tales que el riesgo residual pertinente asociado a cada peligro, así como el riesgo residual global de los sistemas de IA de alto riesgo, se consideren aceptables*». También conviene recordar que ya venimos realizando AARR y EIPD legalmente exigidas basadas en un derecho fundamental como es la protección de datos de carácter personal y utilizando la metodología de gestión de riesgos. La AEPD en su guía ha dicho que «*se podrían considerar como niveles de riesgo residual asumibles aquellos de valor bajo y medio que exigirán esfuerzos de gestión proporcionales a lo largo del ciclo de vida del tratamiento*» y cita las Directrices WP248 con ejemplos de riesgos no aceptables⁴⁷, pero obviamente hay muchos otros riesgos aceptables. En otros casos, como el derecho al medio ambiente, esto se puede ver de forma más clara con un ejemplo, como el de una construcción que cause mucho ruido y en la que se establezca un umbral de ruido aceptable que tenga en cuenta la legislación (suele haber regulación local) u otros criterios como estudios de impacto ambiental cuando sea necesario, o incluso consultas a la comunidad. Pero podrían adoptarse medidas como máquinas menos ruidosas, barreras anti-ruido, limitar las horas del mismo etc. para bajar el riesgo inicial hasta el umbral de ruido aceptable. Pero, pongamos un ejemplo de IA: Imaginemos un sistema de IA aplicable a procesos de selección de personas en el que analizamos el riesgo de que

46. Mantelero, A., en ob. cit. p 172 «*La letra b) del artículo 2 de la Directiva 2001/95/CE, relativa a la seguridad general de los productos, define un producto seguro como aquel que no presenta ningún riesgo o sólo los riesgos mínimos compatibles con el uso del producto, considerados aceptables*». De hecho en este sentido el Considerando 27 del RIA dispone que «*Los sistemas de IA de alto riesgo solo deben introducirse en el mercado de la Unión, ponerse en servicio o utilizarse si cumplen determinados requisitos obligatorios. Estos requisitos deben garantizar que los sistemas de IA de alto riesgo disponibles en la Unión o cuyos resultados se utilicen de otro modo en la Unión no planteen riesgos inaceptables para los intereses públicos importantes de la Unión reconocidos y protegidos por el Derecho de la Unión*».

47. «*Un ejemplo de riesgo residual elevado inaceptable incluye casos en los que los interesados pueden encontrarse con consecuencias importantes, o incluso irreversibles, de las que no puedan recuperarse (p. ej.: un acceso ilegítimo a datos que suponga una amenaza para la vida de los interesados, un despido, un peligro financiero) o cuando parezca obvio que existirá un riesgo (p. ej.: por no poder reducir el número de personas que acceden a los datos debido a sus modos de intercambio, uso o distribución, o cuando no se corrige una vulnerabilidad conocida).*»

incumpla el derecho a la igualdad. El sistema en fase de aprendizaje se entrena con datos de currículums para seleccionar los más idóneos para seguir el proceso. Un posible riesgo identificado para la igualdad es el sesgo del algoritmo (imaginemos que toma más datos de hombres que superan el proceso en el entrenamiento) que podría llevar a discriminar a las mujeres en la fase de inferencia. Una posible forma de determinar el umbral aceptable es considerar un porcentaje de falsos positivos y negativos como máximo aceptable. Y caso de que así fuese adoptar medidas que podrán ir desde la reevaluación de los datos de entrenamiento o la modificación del algoritmo para que balancee o compense ese «riesgo», antes de llegar a la medida extrema de no utilizar dicho sistema de IA.

En definitiva, asumiendo *lege data* la utilización de la metodología de gestión de riesgos que ha establecido el RIA y asumiendo también que no pueden tolerarse impactos altos en derechos fundamentales, no cabrían riesgos aceptables de nivel alto.

Pues bien, atendiendo a la naturaleza del riesgo, deberán adoptarse salvaguardas o controles, que pueden incorporar medidas para bajar dicho riesgo inicial hasta el umbral de riesgo aceptable, medidas que podrán ser de diferente índole, pero sin duda y a diferencia de los controles exigidos al proveedor según el análisis de riesgos del artículo 9, donde hay un peso más importante de los controles técnicos, en el caso de los controles derivados de la FRIA (tendrán más peso los controles relativos a la Gobernanza y cumplimiento legal que le corresponden al implantador.

El artículo 27 1. hace referencia a varias medidas específicas que deberán de tenerse en cuenta: Las medidas de supervisión humana, disposiciones en materia de gobernanza interna y mecanismos de denuncia.

En todo caso él dice el considerando 64 que las medidas a adoptar *«deben de tener en cuenta el estado de la técnica, ser proporcionadas y eficaces»*.

Como hemos indicado y como referencia de posibles medidas, además de las que indica el RIA podemos considerar las referidas en las ISO citadas.

Como resultado de tratar los riesgos se obtiene el riesgo residual, definido como el nivel de riesgo resultante en el tratamiento una vez se hayan aplicado medidas de control para mitigar y/o reducir su nivel de exposición con relación al conjunto de factores de riesgo identificados. A diferencia del riesgo inherente, el riesgo residual contempla las medidas de control definidas sobre el sistema de IA. Por tanto y como dice el artículo 9.4. RIA: *«Las medidas de gestión de riesgos ... serán tales que el riesgo residual pertinente asociado a cada peligro, así como el riesgo residual global de los sistemas de IA de alto riesgo, se consideren aceptables»*.

En definitiva, la conclusión a la que deberá de llegar la FRIA es si dado unos riesgos iniciales, aplicando las medidas o controles oportunos seremos capaces de que el riesgo residual sea inferior al riesgo aceptable.

Dichas medidas deberán implantarse en la fase de DO, y en las fases de *check y act* proceder a su revisión y mejora continua.

Pero antes de la fase de DO el RIA añade que no sólo deben de implantarse, sino que deben de someterse a pruebas, de los que se habla en otros capítulos de este libro.

Otro aspecto importante de la FRIA como parte del sistema de gestión de riesgos es que, como dice el artículo 9 1. «*se establecerá, aplicará, documentará y mantendrá*». Es decir: la documentación y mantenimiento es fundamental.

F) Representación visual, y gestión

Como hemos indicado nos encontramos con una EI que puede considerar en su alcance, en función del sistema de IA, diversos derechos afectados, lo que, multiplicado por los posibles factores de riesgo identificados y evaluados, puede conllevar numerosos riesgos a tratar y consecuentemente muchos controles a aplicar. Ello nos lleva a varios aspectos críticos a abordar:

1. Uno de ellos es la representación clara de los mismos. No se trata únicamente de generar un informe que contenga el contenido exigido por el artículo 27, sino que dicho informe debe de ser comprensible y cuando hablamos de tantos datos no siempre es así. La verdad es que muchas EIDH existentes han tenido mucha literatura, pero los resultados no se ven de forma gráfica, otras sí. Para ello los gráficos pueden ser muy útiles. Una propuesta podría ser el gráfico radial⁴⁸, pero hay otros modelos como por ejemplo podría utilizarse un mapa de calor típico para cada derecho, y aspecto o dimensión analizada (ejemplo seguridad u otras) en el que se identifiquen los y otro a modo de resumen en el que se visualicen los riesgos para cada uno de ellos; y adicionalmente uno consolidado (especialmente idóneo cuando hablamos de muchos derechos afectados) en el que se tenga una visión global⁴⁹.

2. Otro aspecto tiene que ver con la necesidad de utilizar herramientas (se han puesto de moda las herramientas GRC) que permiten no sólo realizar el AARR sino integrarlo en el PDCA permitiendo dar soporte a la visión dinámica consecuente con el plan de mejora que requiere, integrándolo en medidas que requieran otros marcos normativos y en otros sistemas de gestión.

G) Comunicación a la autoridad

Como dice el artículo 27 3 y una vez realizada la FRIA «*el responsable del despliegue notificará a la autoridad de vigilancia del mercado los resultados de la evaluación, presentando la plantilla cumplimentada a que se refiere el apartado 5 como parte de la notificación*».

Que exista una plantilla que la autoridad ponga a disposición de los responsables de reportar es obvio que condiciona la información y el formato del informe, para poder cumplir con esta obligación; pero —en nuestra opinión— no podemos confundir la obligación de reportar a la autoridad mediante un cuestionario con la propia evaluación, el informe en sí y la necesidad de gestión que el mismo requerirá.

En definitiva, la EIDDDFFIA debe de contemplar la información y formato del cuestionario para comunicar a la autoridad, pero previsiblemente más información y en un sistema que permita su iteración y mejora continua.

Esta obligación la tienen todos los responsables de sistemas de IA de alto riesgo excepto aquellos contemplado el artículo 47, apartado 1 del RIA que también están

48. Según Mantelero, A. en ob. cit. p.59. «*el gráfico radial es, por tanto, la mejor herramienta para representar el resultado de la EIDH, mostrando gráficamente los cambios después de introducir las medidas de mitigación*».

49. Dicha visión estaría alineada con la posibilidad a la que se refiere el RIA en varios artículos sobre disponer de una visión del impacto global.

exentos del procedimiento de evaluación de la conformidad y que se refiere a autoridades de vigilancia del mercado que introduzcan en el mercado o pongan a disposición sistemas de IA en la UE por motivos excepcionales de seguridad pública o de protección de la vida y la salud de las personas, protección del medio ambiente y protección de activos industriales e infraestructurales clave.

H) *¿Comunicación a las partes interesadas? ¿Publicación?*

Como hemos mencionado el RIA recoge la necesidad de notificar a la autoridad de vigilancia del mercado los resultados de la evaluación mediante una plantilla. Pero la pregunta es ¿debe de publicarse?. El RIA no obliga a ello. Obviamente es necesario garantizar que las decisiones de los sistemas de IA sean justas e imparciales. Podemos considerar que ello queda garantizado con el hecho de que se comunique a la autoridad; ahora bien, si se pretende además conseguir la confianza y en el caso del sector público además la participación ciudadana, quizá deba de considerarse como buena práctica (dada la falta de exigencia legal) la publicación, que podría contribuir a la mejora de los sistemas en aspectos como por ejemplo la reducción de sesgos. De hecho, es «curioso» que el RIA mencione la participación de las partes interesadas, al igual que a los expertos a la hora de identificar las medidas de gestión de riesgos más adecuadas, pero que no obligue a comunicarles el resultado.

Como es posible que haya información que bien por cuestiones de confidencialidad del negocio no quieran publicarse cabría publicar una versión resumida o suprimiendo dichos aspectos.

VI. EVALUACIONES DE IMPACTO EN DERECHOS FUNDAMENTALES Y EVALUACIONES DE IMPACTO EN PROTECCIÓN DE DATOS

No es objeto de este apartado analizar las múltiples intersecciones que se pueden producir entre los sistemas de IA y el tratamiento de datos de carácter personal que se abordan en otro capítulo de esta obra. Ni siquiera es objeto de este apartado desarrollar la metodología de las específicas EIPD sobre las que, ya que existe abundante literatura y una «tradición asentada en los últimos años», y a la que en parte se ha hecho referencia en este capítulo. El objetivo de este apartado es hacer referencia a la previsión del artículo 27 4. RIA relativa específicamente a que, si alguna de las obligaciones establecidas en el citado artículo ya se cumple mediante la evaluación de impacto relativa a la protección de datos realizada conforme al RGPD, la evaluación de impacto relativa a los derechos fundamentales se realizará juntamente con la misma.

Como ya hemos avanzado a lo largo de este capítulo y es sabido, la protección de datos y también las EIPD son una de las tipologías de EI que mayor despliegue han tenido en los últimos años en Europa. La AEPD no sólo ha elaborado varias guías sobre la materia y ha dictado informes y diversas resoluciones sancionadoras —algunas polémicas— sobre la materia.

Hablar de evaluaciones conjuntas entre las EIPD y las FRIA pasa primero por determinar que: Por un lado, nos encontramos con un sistema de IA; y que, asimismo, trata o va a permitir tratar datos personales. Como dice el documento publicado por la AEPD «*Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una*

introducción»⁵⁰, «Si un componente IA realiza el tratamiento de datos personales, elabora perfiles sobre una persona física o si toma decisiones sobre la misma, tendrá que someterse al RGPD. En caso contrario, no será necesario».

Ejemplos de sistemas de IA que tratan datos personales son numerosos, como por ejemplo los de reconocimiento facial u otros tipos relacionados con el tratamiento de datos biométricos, los de selección de personal, los de márketing y un largo etcétera; pero también hay muchos otros, como por ejemplo los relacionados con procesos de calidad industriales, que no tratan datos de carácter personal.

Asimismo, no siempre es fácil determinar si en alguna etapa del ciclo de vida de un sistema de IA se tratarán o no datos de carácter personal.

Como hemos avanzado, en la metodología propuesta, existe una fase previa o Preevaluación de impacto para delimitar el alcance de los derechos afectados. En dicha fase se determinará (*prima facie*, y sin perjuicio de que si posteriormente se detectase que hay un tratamiento de datos o no se «rectifique» la postura adoptada) si hay un tratamiento de datos y también si existe la obligación de realizar una EIPD para lo cual —y como es sabido— ya existen no solo unos criterios que establece el RGPD y que han sido objeto de interpretación y desarrollo⁵¹. Una cuestión que cabría plantearse quizá aquí es si «automáticamente» podemos asumir que si hay un tratamiento de datos personales (sean o no uno de los supuestos obligados *per se* a realizar EIPD) y además se trata de una IA de Alto Riesgo, si ello equivale a una EIPD necesaria.

Es fundamental que para realizar dicho análisis y tomar la decisión se implique, cuando se haya nombrado, al DPD. En caso de que este no haya sido nombrado por no ser preceptivo o no haberlo nombrado voluntariamente, quien asuma el asesoramiento legal en protección de datos y el CISO deberían ser escuchados al respecto, como sugiere la citada guía de la AEPD sobre EIPD, añadiendo que dichas «sugerencias han de registrarse documentalmente, así como las decisiones tomadas a partir de ellas». Incluso en los supuestos de no existencia de obligación de realizarla el responsable puede tomar la decisión de efectuarla por diversos motivos que cita la citada guía, como, por ejemplo: «en aras de una mayor diligencia a la hora de implementar la responsabilidad proactiva», o para «mejorar la calidad de sus productos y servicios, fomentar la cultura de protección de datos en su organización o bien como simple mecanismo para garantizar la confianza de sus clientes».

Pero, para continuar con el análisis, suponiendo que nos encontremos ante un sistema de IA obligado a realizar una FRÍA y que existe un tratamiento de datos personales que esté obligado a realizar una EIPD, o en sendos supuestos que se decida voluntariamente realizarlas, lo que dice el citado artículo 27 4. RIA es que se la FRÍA se realizará conjuntamente con la EIPD. Que se realicen conjuntamente es lógico ya que la privacidad, al igual que la seguridad desde el diseño, y en general el cumplimiento desde la concepción de cualquier proceso o sistema, es un proceso que debe tender a ser integral e integrado. Ahora bien, la protección de datos ha sido una materia que —como ya hemos indicado— tiene especificidades importantes, alguna de las cuales —como la necesidad y proporcionalidad a la que nos hemos

50. <https://www.aepd.es/documento/adecuacion-rgpd-ia.pdf>

51. Véase la nota 9 de este mismo capítulo.

referido— tienen un sentido y calado diferente (según la propia interpretación que le han dado las autoridades de protección de datos), y asimismo el catálogo de escenario de riesgos y controles correspondientes es muy detallado, que ya vienen integrándose en un sistema de gestión de protección de datos, muchas veces ya integrado en un SGSI, así como otras especificidades. Por ello y en mi opinión, una cosa es que se realicen idealmente de forma conjunta y otra que es posible y, a veces será recomendable que, sin perjuicio de la coordinación de ambas evaluaciones, la de protección de datos devenga en un propio informe *ad hoc* (con las correspondientes «llamadas» o referencias desde el informe de la «FRIA»). De hecho, el propio RIA en el Anexo B de la sección VIII exige que en todo caso haya sendos resúmenes distintos en el caso del registro de los sistemas de alto riesgo:

A) Por un lado un resumen de las conclusiones de la evaluación de impacto sobre los derechos fundamentales realizada de conformidad con el artículo 27;

B) y por otro lado un resumen de la evaluación de impacto relativa a la protección de datos realizada de conformidad con el artículo 35 del Reglamento (UE) 2016/679 o el artículo 27 de la Directiva (UE) 2016/680, tal como se especifica en el artículo 27, apartado 6, del presente Reglamento, cuando proceda.

Otro aspecto para considerar en la ejecución de estos escenarios conjuntos es que en la realización de la EIPD deberá de preguntarse siempre al proveedor qué medidas ha implementado y basarse en la información del AARRIA.

Obviamente cuando hablamos de sistemas maduros e integrados, una cosa será el informe y otra cosa la gestión de los riesgos identificados que, en tanto en cuanto se integren un sistema de mayor calado, conseguirán que la gestión sea integral, consiguiendo mayor eficacia.

VII. RECAPITULACIÓN Y CONCLUSIONES

Hemos enmarcado las evaluaciones de impacto (EI) como herramientas de ponderación de derechos fundamentales «importadas» desde la tradición anglosajona y que son realizadas por los propios sujetos que son parte del proceso de toma decisiones, sin perjuicio de las que realiza el legislador, los tribunales y la administración en determinados supuestos.

A continuación, hemos tomado como ejemplo la protección de datos para distinguir conceptualmente entre Análisis de riesgos en protección de datos (AARRPD) y evaluaciones de impacto (EIPD), aunque después al trasladar este tema conceptual al ámbito del RIA hemos concluido que dichas diferencias no son extrapolables exactamente en este ámbito ya que —como también se ha comentado en el capítulo que trata Pere Simón— la gestión de riesgos del artículo 9 RIA viene referida a todos los sistemas de IA de alto riesgo y es una obligación que han de cumplir los proveedores; en cambio la evaluación de impacto que contempla el artículo 27 (EIDDDFF) está orientada a supuestos concretos y los obligados a realizarla son determinados responsables del despliegue de dichos específico sistemas de alto riesgo.

Se han analizado diversas tipologías de evaluaciones de impacto como la HRESIA o EISDH (evaluación de impacto social, ético y de los Derechos Humanos), la PIA (*Privacy Impact Assessment*) o DPIA (*Data Privacy Impact Assessment*), la SIA (*Social*

Impact Assessment) y la *EtIA Ethical Impact Assessment*) así como los antecedentes de evaluaciones de impacto de derechos fundamentales de sistemas de IA previos al RIA, con referencias a metodologías utilizadas en diferentes países y propuestas por diferentes organizaciones.

A continuación hemos aterrizado a las FRIA en el texto del RIA y analizado su evolución, tramitación y contenido final respecto del nuevo artículo 27, que ha sido una «introducción reciente» puesto que apareció en la versión del parlamento para posteriormente centrarnos en cómo ha quedado su regulación actual y realizar una aproximación metodológica, de la que podemos destacar los siguientes aspectos:

1. Que los obligados a realizar las FRIA son únicamente determinados responsables del despliegue concretos.

2. Que es importante que el equipo que intervenga sea multidisciplinar y que hay que considerar que habrá roles que serán siempre necesarios y otros que serán contingentes; así como perfiles especializados en IA y otros que aportaran su visión desde su campo de competencia. Asimismo, que en su realización es la importante implicar a las partes interesadas y posibles expertos.

3. Que la FRIA debe de realizarse tanto antes de desplegar el sistema de IA de alto riesgo concreto, así como que deberá de actualizarla ante posibles cambios en factores que tengan que ver con el sistema de IA (usos o finalidades, período de tiempo y frecuencia de uso previsto o categorías de personas físicas o grupos afectados), riesgos que se identificaron o medidas que se adoptaron para gestionar dichos riesgos.

4. Que hay que acotar el alcance sustantivo, antes de realizarla, mediante una preFRIA en la que se tendrán en cuenta los siguientes requisitos:

a) Que se trate de sistemas de alto riesgo.

b) Que esos sistemas se refieran a las entidades y/o Servicios concretos dentro del alcance subjetivo que indica el RIA

c) Y —por último— que afecte a derechos fundamentales, centrando a cuáles.

5. Que, teniendo en cuenta dicha fase previa de acotar el alcance, podríamos establecer una propuesta de fases de la propia FRIA, que serían:

1. Fase 1. Análisis preliminar sobre la necesidad de realizar la FRIA y concretar los Sistemas y derechos afectados.

2. Fase 2. Conocer el contexto (lo que va más allá de la fase de preFRIA, ya que spondrá profundizar en diferentes aspectos como los procesos del implantador, el período y frecuencia previsto de uso o las categorías de personas físicas y grupos afectados en el contexto específico) y planificarla, lo que conllevará también aterrizar el correspondiente cronograma con hitos y el equipo implicado en cada fase en relación con una metodología a utilizar.

3. Fase 3. Aunque no lo contempla específicamente el RIA, en nuestra opinión un análisis sobre la necesidad y proporcionalidad de despliegue del sistema de IA debería de realizarse, no sólo cuando se exija en relación con el tratamiento de datos personales, según el RGPD.

4. Fase 4. La fase de gestión de riesgos que incluye la identificación de los riesgos, análisis de posibles escenarios de riesgos, la estimación y evaluación de los mismos y la posterior gestión mediante la propuesta de las medidas oportunas. En este punto

hemos prestado especial atención a la asunción de riesgo aceptable y asumido lege data que el RIA asume que no existe el riesgo 0, que pueden existir riesgos tolerables, pero —esto ya esa una opinión personal— entendemos que dado que estamos hablando de que uno de los factores del riesgo es el impacto, y de que hablamos de impacto en derechos fundamentales, el umbral de riesgo aceptable debiera de ser más bajo y no debieran de aceptarse riesgos altos.

6. Que toda la información recabada y el análisis realizado llevará a una conclusión sobre si se entiende que, dados unos riesgos iniciales determinados, aplicando las medidas o controles oportunos, la organización será capaz de que el riesgo residual sea inferior al riesgo aceptable.

7. Que toda la anterior información se hará constar en el correspondiente informe, pero:

a) Que es fundamental que dicho informe sea claro, comprensible y una de las mejores formas de hacerlo es que sea gráfico.

b) Que las medidas contempladas en dicho informe tienen que integrarse en un ciclo de gestión y mejora continua, lo que aconseja el uso de herramientas que permitan dicha iteración y mejora continua.

c) Que una cosa es el informe y otra una plantilla a presentar a la autoridad. Obviamente esta plantilla condiciona la información mínima y el formato del informe, pero en mi opinión no podemos confundir ni tienen porqué coincidir (salvo que dicha plantilla sea muy completa y exigente) la información a reportar con la información a recabar y también a gestionar, que puede ser mayor, pero no menor que la de la plantilla a comunicar.

d) Que el RIA no habla de la publicación de este ni de la comunicación a las partes interesadas, pero entendemos que (salvando aspectos que por motivos como la confidencialidad de «negocio» u otros que sea legítimo obviar) puede ser una buena práctica la publicación del mismo y/o la comunicación —al menos— a las partes interesadas.

8. No ha sido objeto de este capítulo (ya que ha sido abordado en otros) analizar todas las múltiples posibles intersecciones que se pueden producir entre los sistemas de IA y el tratamiento de datos de carácter personal, que se abordan en otro capítulo de este libro, pero si analizar la exigencia establecida en el artículo 27 del RIA de que si alguna de las obligaciones establecidas en el citado artículo relativa a las FRIA ya se cumple mediante la evaluación de impacto relativa a la protección de datos conforme al RGPD, la evaluación de FRIA se realizará juntamente con la misma, y hemos concluido que una cosa es que se realicen —idealmente— de forma conjunta; y otra que es posible, y a veces será recomendable que, sin perjuicio de la coordinación de sendas evaluaciones, la de protección de datos tenga su informe *ad hoc*, y de hecho el propio RIA exige sendos resúmenes distintos para ambas: la EIPD sobre sistemas de IA de alto riesgo y la FRIA de alto riesgo.

Los sistemas de gestión de riesgos como obligación específica para los sistemas de inteligencia artificial de alto riesgo en el artículo 9 del Reglamento

PERE SIMÓN CASTELLANO

Profesor Titular de Derecho Constitucional
Universidad Internacional de la Rioja - UNIR1

I. QUÉ ES UN SISTEMA DE GESTIÓN DE RIESGOS. AUTONOMÍA CONCEPTUAL RESPECTO DE FIGURAS AFINES

En el presente estudio nos vamos a ocupar de la regulación de los sistemas de gestión de riesgos, que el artículo 9 del RIA establece como una obligación jurídica o un requisito mínimo indispensable de cualquier sistema de IA que sea clasificado como de riesgo alto.

Un sistema de gestión de riesgos es un conjunto de procesos, políticas, procedimientos y herramientas diseñadas para identificar, evaluar, mitigar y monitorear los riesgos que enfrenta una organización en el logro de sus objetivos. Estos sistemas ayudan a las organizaciones a comprender y gestionar los riesgos de manera efectiva, con el fin de minimizar pérdidas, maximizar oportunidades y garantizar la continuidad del negocio.

Obviamente, existen muchos tipos de sistemas de gestión, en función del ámbito sectorial concreto en el que aplican y los riesgos que pretenden minimizar. Así, los sistemas de gestión de riesgos abordan riesgos de diferentes tipos, incluyendo riesgos financieros, operativos, medioambientales, relativos a la calidad de los procesos, tributarios, de seguridad de la información, estratégicos, legales, de privacidad, de cumplimiento genérico o penal (prevención de riesgos penales) y de reputación. Estos riesgos pueden surgir de diversas fuentes, como la volatilidad del mercado, cambios en el entorno legal y regulatorio, fallas en los procesos internos, desastres naturales, ciberataques, entre otros.

Un sistema de gestión de riesgos generalmente sigue un proceso cíclico que incluye las siguientes etapas:

1. El presente trabajo se realiza en el marco del Proyecto «Derechos y garantías públicas frente a las decisiones automatizadas y el sesgo y discriminación algorítmicas» 2023-2025 (PID2022-136439OB-I00) financiado por MCIN/AEI/10.13039/501100011033/ FEDER, UE.

1. IDENTIFICACIÓN DE RIESGOS (O FASE DE APRECIACIÓN)

Herramientas y técnicas	Proceso de apreciación del riesgo				
	Identificación del riesgo	Análisis del riesgo			Evaluación del riesgo
		Consecuencia	Probabilidad	Nivel de riesgo	
Tormenta de ideas	MA ¹⁾	NA ²⁾	NA	NA	NA
Entrevistas estructuradas o semiestructuradas	MA	NA	NA	NA	NA
Delphi	MA	NA	NA	NA	NA
Listas de verificación	MA	NA	NA	NA	NA
Análisis preliminar de peligros	MA	NA	NA	NA	NA
Estudios de peligros y de operatividad (HAZOP)	MA	MA	A ³⁾	A	A
Análisis de peligros y puntos de control críticos (HACCP)	MA	MA	NA	NA	MA
Apreciación de riesgos ambientales	MA	MA	MA	MA	MA
Estructura «y sí...» (SWIFT)	MA	MA	MA	MA	MA
Análisis de escenario	MA	MA	A	A	A
Análisis del impacto económico	A	MA	A	A	A
Análisis de la cauMA primordial	NA	MA	MA	MA	MA
Análisis de los modos de fallo y de los efectos	MA	MA	MA	MA	MA
Análisis del árbol de fallos	A	NA	MA	A	A
Análisis del árbol de sucesos	A	MA	A	A	NA
Análisis de cauMA-consecuencia	A	MA	MA	A	A
Análisis de cauMA-y-efecto	MA	MA	NA	NA	NA
Análisis de capas de protección (LOPA)	A	MA	A	A	NA
Diagrama de decisiones	NA	MA	MA	A	A
Análisis de fiabilidad humana	MA	MA	MA	MA	A
Análisis de pajarita	NA	A	MA	MA	A
Mantenimiento centrado en la fiabilidad	MA	MA	MA	MA	MA
Análisis del circuito de fuga	A	NA	NA	NA	NA
Análisis Markov	A	MA	NA	NA	NA
Simulación Monte-Carlo	NA	NA	NA	NA	MA
Estadísticas Bayesianas y redes Bayes	NA	MA	NA	NA	MA
Curvas FN	A	MA	MA	A	MA
Índices de riesgo	A	MA	MA	A	MA
Matriz de consecuencia/probabilidad	MA	MA	MA	MA	A
Análisis de costes/beneficios	A	MA	A	A	A
Análisis de decisión multi-criterios (MCDA)	A	MA	A	MA	A

1) Muy aplicable.
 2) No aplicable.
 3) Aplicable.

Figura 1. Detalle del grado de aplicabilidad de las herramientas utilizadas para la apreciación del riesgo. Fuente UNE ISO 31010:2010

La fase de apreciación o identificación consiste en identificar y comprender los riesgos potenciales que enfrenta la organización en el logro de sus objetivos. Para ello existen muchas técnicas que se comparten, a continuación, en la tabla 1 que deriva de la ISO 31010:2010, de técnicas de apreciación del riesgo. Probablemente se trata de las técnicas más utilizadas, todas ellas explicadas con especial nivel de detalle en la norma internacional ISO 31010:2010, muchas de ellas ejemplificadas con figuras o diagramas. Lógicamente aquí sólo hemos citado algunas de las técnicas que incorpora el citado estándar internacional. Buena parte de las técnicas y herramientas descritas se encuentran en la Figura 1, con detalle de su grado de aplicación y eficacia (1 - Muy aplicable; 2 - No aplicable; 3 - Aplicable) en las distintas fases de la gestión del riesgo.

2. EVALUACIÓN DE RIESGOS DE DERECHOS Y SU DIFERENCIACIÓN DEL ANÁLISIS Y GESTIÓN DE RIESGOS DEL PROVEEDOR

En esta fase se trata de analizar y evaluar la probabilidad de ocurrencia y el impacto de los riesgos identificados, para priorizarlos según su importancia. Evaluar un riesgo implica considerar todos los posibles escenarios en los cuales el riesgo se haría efectivo. La evaluación de riesgos consiste en valorar el impacto de la exposición a la amenaza, junto a la probabilidad de que esta se materialice. El impacto, por su parte, se determina en base a los posibles daños que se pueden producir si la amenaza se materializa, por ejemplo, un impacto sería despreciable si no tuviera consecuencias sobre los bienes jurídicos protegidos o, por el contrario, un impacto sería significativo si el daño ocasionado sobre estos fuese crítico. Según la probabilidad y el impacto, asociados a las amenazas, es posible determinar el nivel de riesgo inherente.

La valoración del riesgo está inextricablemente vinculada a la matriz de riesgos que se construye en función del método que se emplea. Por lo que se refiere a los métodos para cuantificar, los ejemplos y modelos de matrices y de mapas de riesgo, nos remitimos a obras específicas sobre sistemas de cumplimiento normativo².

Las matrices de riesgos más utilizadas son las de 3x3 y 5x5 y normalmente cuentan con los factores de probabilidad e impacto o gravedad, si bien se pueden utilizar distintas matrices y fórmulas que también apliquen otros elementos o criterios como la función, la sustitución, la profundidad, el grado de externalización, el nivel de agresión y la vulnerabilidad.

La fase de evaluación de riesgos en el seno de un sistema de gestión de riesgos debe ser necesariamente diferenciada de las evaluaciones de impacto en determinados ámbitos específicos o también denominadas evaluaciones de riesgos *ad hoc*, como podrían ser la evaluación de impacto en protección de datos y evaluación de impacto algorítmico en derechos fundamentales. La última, será objeto de estudio en esta misma obra, *infra*, en el capítulo que rubrica Eduard Chaveli, en relación

2. Véanse los trabajos de Simón Castellano, P. «Responsabilidad penal de las personas jurídicas, mapa de riesgos y cumplimientos en la empresa», en Simón Castellano, P. y Abadías Selma, A. (coordinadores), *Mapa de riesgos penales y prevención del delito en la empresa*, Wolters Kluwer — Bosch, 2020, pp. 31-76 y Salvador Lafuente, A. «Mapa de riesgos: identificación y análisis de riesgos y controles», en Simón Castellano, P. y Abadías Selma, A. (coordinadores), *Mapa de riesgos penales y prevención del delito en la empresa*, Wolters Kluwer — Bosch, 2020, pp. 78-119.

con el contenido del artículo 27 del RIA, vinculado con las evaluaciones de impacto algorítmico en los derechos fundamentales.

Resulta aquí especialmente útil realizar una analogía con los sistemas de gestión de seguridad de la información (en adelante, SGSI), especialmente útiles para garantizar la privacidad y la protección de los datos personales, activos singulares y esenciales (información y datos) de las empresas y administraciones públicas. En este sentido, existen una serie de estándares.

La AEPD ha elaborado distintas guías para la realización de análisis de riesgos en el seno de sistemas de gestión de riesgos de seguridad de la información y de protección de datos, con el objetivo de establecer una hoja de ruta para afrontar los riesgos del tratamiento de datos personales mediante el establecimiento de medidas de seguridad y controles que garanticen los derechos y libertades de los individuos en el ámbito de la privacidad y de la protección de datos. Es en este ámbito en el que encontramos la guía práctica de análisis de riesgos en los tratamientos de datos personales sujetos al RGPD, cuyo enfoque es un mix de buena parte de los principios y directrices de las metodologías ISO y de los sistemas de gestión.

Un buen ejemplo es la matriz de riesgos y la fórmula que se propone para calcular el riesgo inherente y el riesgo residual, que comparten las dos guías y que se ilustra muy bien en la Figura 2. La principal ventaja es que se trata de una fórmula muy sencilla; el principal inconveniente es que pierde nivel de detalle del riesgo en relación con otras fórmulas, que pueden incluir cinco o más niveles de probabilidad e impacto.



Figura 2. Matriz de riesgo. Fuente: Guía práctica AEPD

Para lo que aquí interesa, se estructura el análisis en tres fases basadas en el principio de responsabilidad proactiva, la comunicación, la revisión y la mejora continua. Esas tres fases son la identificación, el análisis y el tratamiento de los riesgos. Lo que se ilustra en la Figura 3.

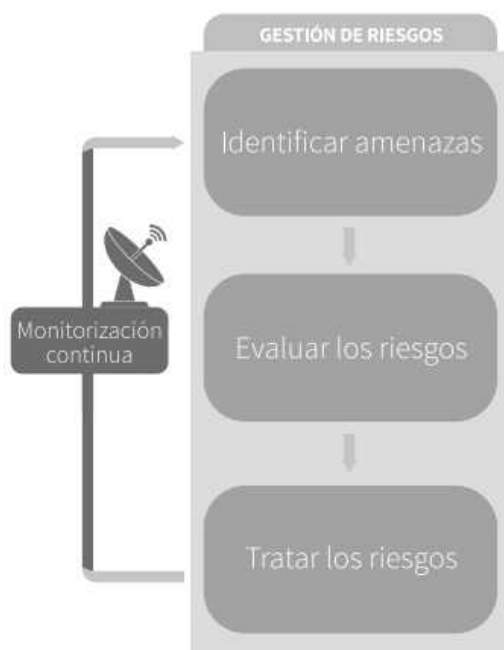


Figura 3. Tres fases guiadas por el principio de monitorización o mejora continua.
Fuente: Guía práctica AEPD

Otro buen ejemplo de lo comentado en relación con los SGSI lo encontramos en los estándares internacionales para la implementación y gestión de un sistema de gestión. La UNE-EN ISO/IEC 27001:2023, que es la norma europea que a su vez adopta la norma internacional sobre requisitos de los sistemas de gestión de los sistemas de información Norma Internacional ISO/IEC 27001:2022, y que se complementa con la ISO/IEC 27005, que proporciona directrices para la gestión de riesgos de seguridad de la información.

La citada norma internacional proporciona directrices sobre la aplicación de un enfoque de gestión de riesgos orientado a procesos para ayudar en la aplicación de manera satisfactoria y al cumplimiento de los requisitos de gestión de riesgos de seguridad de la Norma ISO/IEC 27001.

Veamos un gráfico con el detalle de las relaciones entre las normas ISO/IEC de la familia de los SGSI.

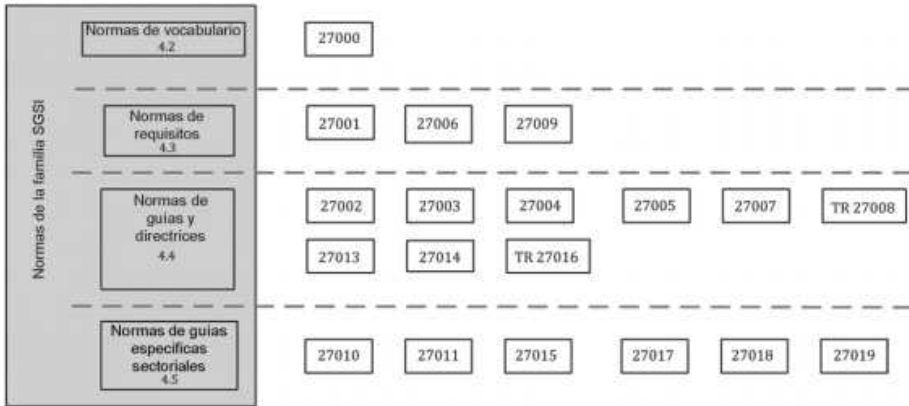


Figura 4. Normas de la familia SGSI. Fuente: UNE-EN ISO/IEC 27000:2019

La norma ISO/IEC 27005 contiene diferentes recomendaciones y directrices generales para la gestión de riesgo en los SGSI. En ella se define el riesgo como una amenaza que explota la vulnerabilidad de un activo pudiendo causar daños y se relaciona el riesgo con el uso, propiedad, operación, distribución y la adopción de las tecnologías de la información de la empresa. El estándar internacional utiliza un proceso estructurado, sistemático y riguroso de análisis de riesgos para la creación del plan de tratamiento de riesgos. A través de este sistema de gestión se identifican los activos de información que se deben proteger, entre ellos protección de datos personales, y se valoran los riesgos desde una perspectiva de debilidades o vulnerabilidades y amenazas a las que están expuestos proponiéndose controles de para tratar el riesgo reduciéndolo, aceptándolo, transfiriéndolo o incluso eliminándolo.

La norma internacional es muy completa e incluye anexos específicos de matriz de riesgos, para definir el alcance y límites del sistema de seguridad, para identificar y valorar los activos en función de su impacto, para cuantificar la probabilidad y el impacto del riesgo, así como propone métodos para asesorar en relación con las vulnerabilidades, la amenazas tradicionales y definición de riesgo aceptable y criterios para su modificación.

En la figura 5, en inglés, se detalla el paso del riesgo inherente al riesgo residual, en el tratamiento de riesgos aceptables como consecuencia de un asesoramiento satisfactorio.

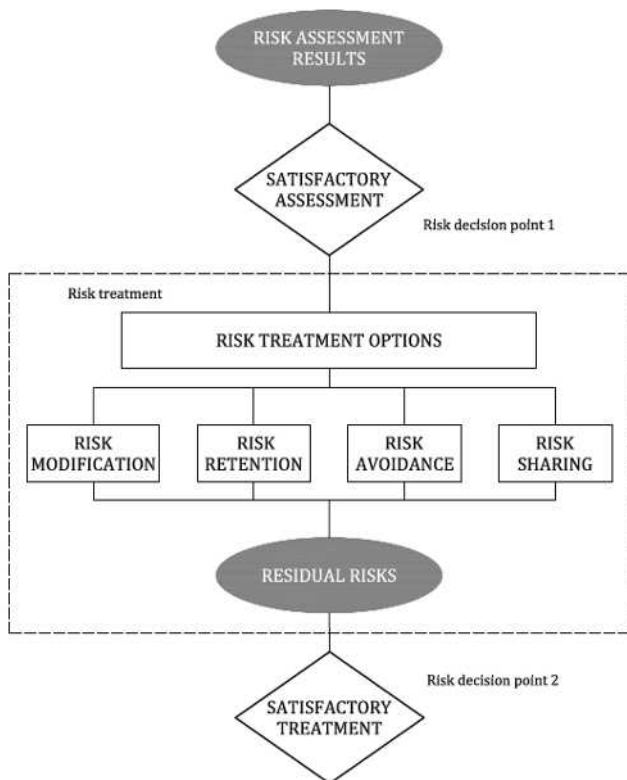


Figura 5. De la evaluación al tratamiento aceptable. Fuente: UNE-EN ISO/IEC 27005:2018

Sea como fuere, la norma o estándar internacional más reciente, la citada UNE-EN ISO/IEC 27001:2023 nos dice que la organización debe desarrollar y aplicar un proceso para evaluar los riesgos de seguridad de la información. Este proceso debe cumplir con los siguientes requisitos: la organización debe definir y mantener criterios sobre los riesgos de seguridad de la información, incluyendo los criterios de aceptación de riesgos y los criterios para llevar a cabo las evaluaciones de riesgos.

Asimismo, se debe asegurar que las sucesivas evaluaciones de riesgos de seguridad de la información generen resultados consistentes, válidos y comparables.

En cuanto a la identificación de los riesgos de seguridad de la información, esto se logra mediante la realización del proceso de evaluación de riesgos, para identificar los riesgos asociados con la pérdida de confidencialidad, integridad y disponibilidad de la información dentro del alcance del sistema de gestión de la seguridad de la información. También es importante identificar a los propietarios de los riesgos.

Para analizar los riesgos de seguridad de la información, es necesario evaluar las posibles consecuencias que surgirían si los riesgos identificados llegaran a materializarse, realizar una valoración realista de la probabilidad de ocurrencia de los riesgos identificados y determinar los niveles de riesgo.

Por último, se debe evaluar los riesgos de seguridad de la información comparando los resultados del análisis de riesgos con los criterios de riesgo establecidos y priorizando el tratamiento de los riesgos analizados. Además, la organización debe conservar información documentada sobre el proceso de evaluación de riesgos de seguridad de la información.

Como observamos, la fase de evaluación de riesgos propia de un sistema de gestión de riesgos nada tiene que ver con lo que es y supone una evaluación de impacto en ese ámbito concreto sectorial, cuyo equivalente sería la evaluación de impacto en protección de datos.

Sobre esta materia, para seguir con la analogía, encontramos también una guía específica de la AEPD para realizar una evaluación de impacto, cuya metodología es distinta a las evaluaciones de riesgos genéricas de los SCSL. La evaluación de impacto (en adelante, EIPD) es, ante todo, una herramienta de desarrollo de la privacidad desde el diseño en el seno de las organizaciones, de igual modo que lo son el diseño y arquitectura del registro de actividades de tratamiento y las demás evaluaciones de riesgos. La EIPD, al igual que cualquier otra evaluación, debe hacerse para cada actividad de tratamiento, sin perjuicio que se puedan extraer indicadores globales o agrupados por procesos de negocio o departamentos.

La gran diferencia entre la EIPD y una evaluación de riesgo al uso es, fundamentalmente, el enfoque desde los derechos de los interesados y el empleo de los principios de la protección de datos como marcos conceptuales para el análisis del riesgo que entraña el tratamiento³. Además que la evaluación gira en torno a un tratamiento concreto de datos personales y a su ciclo de vida (de los datos y del tratamiento) La EIPD se centra en identificar las amenazas sobre los derechos y libertades del interesado, en un contexto de tratamiento de datos personales, con lo que la EIPD, en síntesis, no constituye un análisis funcional de un sistema de información en el que se evalúan los riesgos tecnológicos, así como tampoco es una auditoría de seguridad de la información o de cumplimiento normativo, en general.

Fruto de ese enfoque los escenarios de riesgo con los que trabajaremos en una EIPD serán la discriminación, usurpación de la identidad, fraude, pérdida financiera, daño reputacional, pérdida de la confidencialidad de datos sujetos a secreto profesional, reversión no autorizada de la seudonimización, pérdida de control sobre los datos personales, revelación de origen racial o étnico del interesado, revelación de opinión política, creencia religiosa o filosófica o militancia sindical, revelación de detalles sobre la salud o historial sexual del interesado, revelación de condenas penales o infracciones administrativas del sujeto, entre otras. Si nos fijamos, las temáticas y el enfoque son más amplios que en una evaluación de riesgos genérica de un sistema de gestión de riesgos, a pesar de que ese estudio se realiza de forma pormenorizada sobre un único tratamiento de datos personales o sobre una única operación de procesamiento de datos, teniendo en cuenta todo el ciclo de vida de estos.

Así las cosas, como advertimos anteriormente, en el caso de la EIPD, el objeto del análisis son los riesgos respecto a la privacidad en las actividades de tratamiento

3. Al respecto, véase Simón Castellano, P. «El ejercicio de las funciones del delegado de protección de datos en la supervisión y gestión de procesos críticos», en Simón Castellano, P. y Bacaria Martrus, J. (coordinadores), *Las funciones del delegado de protección de datos en los distintos sectores de actividad*, Wolters Kluwer — Bosch, 2020, pp. 27-74.

de datos personales y los sistemas de información implicados en la organización. A tal efecto, se ha desarrollado un estándar o norma internacional específica, la ISO/IEC 29134:2017, que incorpora las fases en las que tiene desarrollarse una EIPD y la estructura que debe seguir el informe resultante del proceso.

Más allá del artículo 35 del RGPD, y de la mención escueta del artículo 28.1 de la LOPDGDD, las directrices, metodologías y controles para realizar una EIPD pueden encontrarse en normas internacionales, según se indica en la siguiente figura.

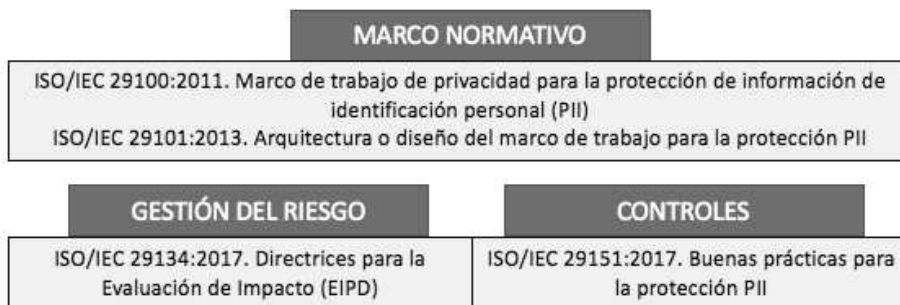


Figura 6. Marco de trabajo, gestión de riesgos y controles de la EIPD. Fuente: elaboración propia

La EIPD se puede realizar siguiendo distintos métodos y herramientas. Ni la norma europea, ni la española, establecen preferencias u obligatoriedad por una determinada metodología o sistema. En cualquier caso, tenemos a nuestra disposición el marco normativo descrito en la Figura 6, que incluye la ISO/IEC 29134:2017, específica para evaluaciones de impacto en protección de datos. También disponemos de la guía práctica de la AEPD para las evaluaciones e impacto en la protección de datos sujeta al RGPD, que ya hemos citado anteriormente, aunque tiene un nivel de detalle muy inferior o menor que el marco normativo, de gestión del riesgo y controles previstos en las normas y estándares internacionales citados en la ilustración.

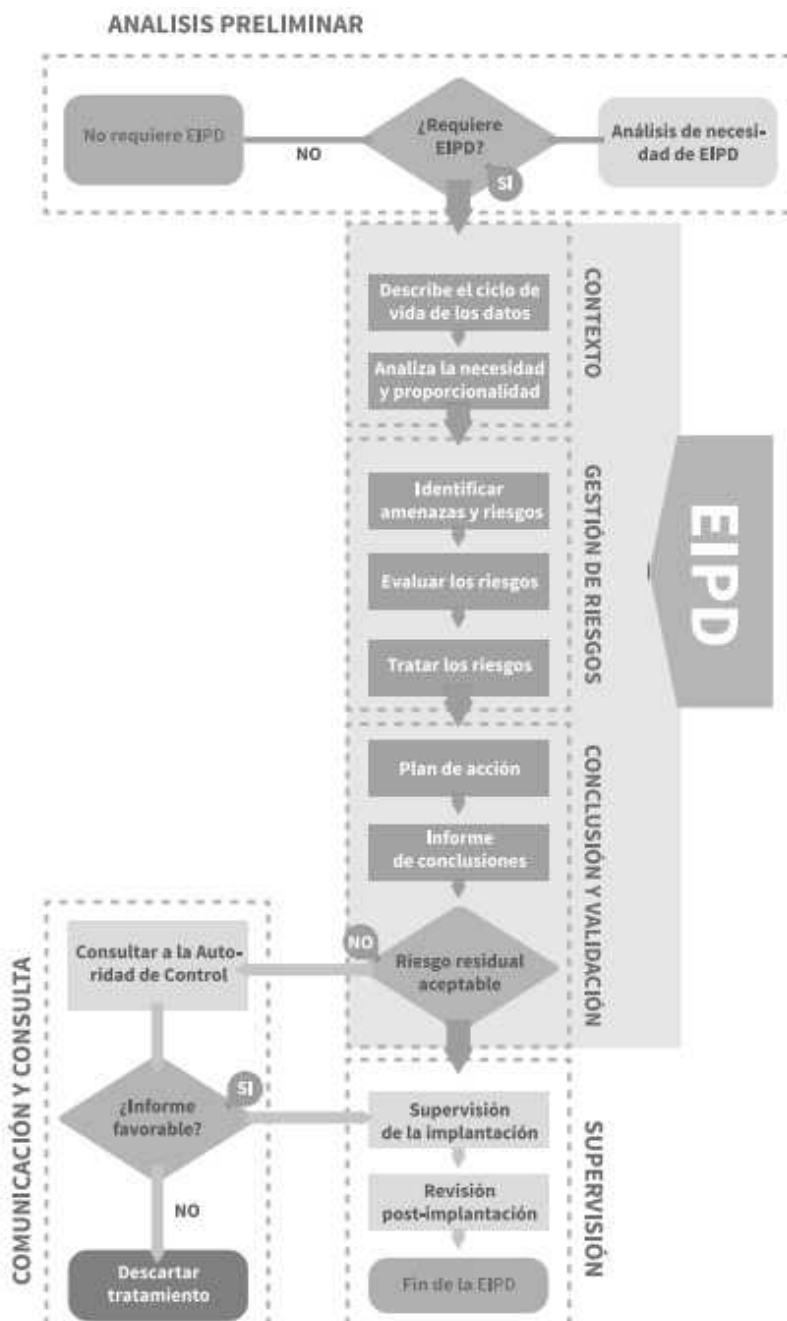


Figura 7. Metodología EIPD. Fuente: Guía práctica EIPD de la AEPD

Con todo, existe una diferencia más importante aún entre las obligaciones previstas en el artículo 9 y el artículo 27 del RIA, que no tienen tanto que ver en el qué o en el contenido de un sistema de gestión de riesgos versus el objeto de una evaluación de impacto algorítmico en los derechos fundamentales, sino también en quién es el sujeto obligado a mantener el sistema o realizar la evaluación. La primera, el artículo 9, se refiere a un requisito mínimo, y como se verá a continuación en el apartado II.2 mediante un cuadro resumen de obligaciones en relación con sujetos obligados, proyecta efectos mayoritariamente sobre los proveedores de los sistemas de IA de riesgo alto. En cambio, el artículo 27 del RIA se centra y se refiere, en exclusiva, a los usuarios (en fase de tramitación se utilizaba este vocablo) o «responsables del despliegue» (terminología final en la última versión conocida el 13 de marzo de 2024 con la aprobación de las enmiendas definitivas por el Parlamento europeo), que son las empresas o administraciones públicas que deciden utilizar, emplear o implementar un sistema de IA de alto riesgo, con independencia de las obligaciones de fabricante o proveedor, de los importadores y de los distribuidores de los mismos. Así las cosas, la obligación específica de realizar una evaluación de impacto algorítmico en los derechos fundamentales recaen en exclusiva sobre los usuarios o responsables del despliegue de la herramienta tecnológica clasificada como de alto riesgo.

3. MITIGACIÓN DE RIESGOS, MONITOREO Y REVISIÓN

En esta fase el principal objetivo es desarrollar e implementar estrategias y medidas para reducir la probabilidad de ocurrencia o el impacto de los riesgos, así como prepararse para responder efectivamente en caso de que ocurran. También incluye la supervisión continua de los riesgos y de las medidas de mitigación implementadas, así como a la revisión periódica del sistema de gestión de riesgos para asegurar su efectividad y relevancia.

La última fase del proceso de gestión de riesgos es el tratamiento para minimizar o mitigar sus efectos. El objetivo de tratar los riesgos es disminuir su nivel de exposición con medidas de control que permitan reducir la probabilidad y el impacto, gravedad o severidad de que estos se materialicen. El riesgo inherente se puede tratar con el objetivo de reducir o mitigar el mismo, en función de la medida que se adopte, hasta situar el riesgo residual en un nivel que se considere razonable. El riesgo residual será el resultado de reducir el nivel de riesgo inherente en función de la eficacia de los controles activos que se calcula, entre otros, teniendo en cuenta el porcentaje de vulnerabilidad de estos. La vulnerabilidad se puede calcular de distintas formas, tal y como se explica en estudios exhaustivos o específicos sobre los mapas de riesgos⁴ (penales).

Sea como fuere, para lo que aquí interesa y a modo introductorio, resulta fundamental comprender qué consecuencias prácticas tiene el principio de mejora continua, que exige la revisión constante o periódica de los sistemas y programas de cumplimiento.

4. Véase Simón Castellano, P. y Abadías Selma, A. (coordinadores), *Mapa de riesgos penales y prevención del delito en la empresa*, Wolters Kluwer — Bosch, 2020.

La mejora continua se alcanza básicamente con la participación e impulso de forma periódica del órgano de cumplimiento normativo, unipersonal o colegiado, en dos ámbitos fundamentales de cualquier sistema de gestión de riesgos: (1) la gestión de los registros e inventarios de la empresa y (2) la comunicación, consulta, seguimiento y revisión de las evaluaciones de riesgo previas.

Los registros e inventarios permiten relacionar y conectar el contexto de la organización con los riesgos y también con los controles que derivan de las evaluaciones de riesgo, lo que permite tener una visión actualizada a las necesidades de la organización. Se trata de un instrumento funcional, operativo y accionable, que debe ser consultado y revisado por el oficial de cumplimiento de forma recurrente, o por el responsable del sistema de gestión. De hecho, debe informarse al órgano de cumplimiento o al oficial de cumplimiento (o responsable, puesto que en función del tipo de sistema los nombres de los responsables y/o los cargos pueden variar) de cualquier cambio o modificación en los citados inventarios.

La mejora continua también se alcanza con el encargo de auditorías internas o externas, y con la gestión y resolución de incidencias concretas o mediante la depuración de responsabilidades a través de canales específicos, como los de denuncia, que necesariamente deben contar con medidas para proteger al alertador o denunciante⁵. El proceso de mejora continua exige así definir un plan de auditorías y de revisiones periódicas en base a las actividades y procesos de negocio de la organización, así como en función de los resultados de las evaluaciones de riesgos.

Un sistema de gestión de riesgos bien establecido y ejecutado ayuda a las organizaciones a tomar decisiones informadas, mejorar su resiliencia frente a los riesgos y crear valor a largo plazo, aunque evidentemente, sólo en los ámbitos en los que despliega o proyecta efectos: medioambiental, calidad, financiero, penal o cumplimiento normativo, seguridad de la información, protección de datos o seguridad y resiliencia de un sistema de IA, entre muchas otras posibilidades.

¿Qué puede en especial aportar las metodologías y los sistemas de gestión de riesgos en el ámbito de la aplicación de un sistema de IA de alto riesgo? Los enfoques de gestión de riesgos aplicados a lo largo del ciclo de vida del sistema de IA pueden identificar, evaluar, priorizar y resolver situaciones que podrían afectar adversamente el comportamiento y los resultados de un sistema.

Se pueden identificar distintas fases para gestionar los riesgos de la IA asegurando el respeto por los derechos humanos y los valores democráticos, y todo ello sin confundir, como decíamos anteriormente, este proceso de gobierno y gestión interna con lo que supone realizar propiamente una evaluación de impacto en derechos fundamentales. Los sistemas de gestión de riesgos pueden estar basados en el marco de gestión de riesgos de IA del NIST, anteriormente citado, el marco de gestión de riesgos de la familia ISO 31000, que también hemos detallado *ut supra*, y la guía de

5. Véanse sobre la materia los trabajos de León Alapont, J. *Canales de denuncia e investigaciones internas en el marco del compliance penal corporativo*, Tirant lo Blanch, 2023; Simón Castellano, P. «La inmunidad penal como recompensa a los denunciantes. Allende un nuevo factor subjetivo-formal de punibilidad», *Revista electrónica de ciencia penal y criminología*, n.º 24, 2022.

diligencia debida de la Organización para la Cooperación y el Desarrollo Económicos (en adelante, la OCDE)⁶.

Esas distintas fases podrían clasificarse, siguiendo la citada guía de la OCDE, en las siguientes: (1) definición del alcance, el contexto y los criterios, incluyendo los principios de inteligencia artificial relevantes, los interesados y los actores para cada fase del ciclo de vida del sistema de inteligencia artificial y para el ciclo de vida en sí mismo; (2) fase de evaluación de los riesgos para una inteligencia artificial confiable mediante la identificación y el análisis de problemas a niveles individuales, agregados y sociales, y evaluando la probabilidad y el nivel de daño (por ejemplo, los pequeños riesgos pueden acumularse y convertirse en un riesgo mayor); (3) tratamiento de los riesgos para cesar, prevenir o mitigar los impactos adversos, en proporción con la probabilidad y el alcance de cada uno; (4) gobernar el proceso de gestión de riesgos mediante la incorporación y el cultivo de una cultura de gestión de riesgos en las organizaciones; supervisando y revisando el proceso de manera continua; y documentando, comunicando y consultando sobre el proceso y sus resultados.

La única forma de alcanzar entonces una IA confiable y, también, responsable, es que los actores implicados se aprovechen de los procesos, indicadores, estándares, esquemas de certificación, auditorías y otros mecanismos que permiten hacer seguimiento y garantizar esos procesos y componentes en cada fase del ciclo de vida del sistema de IA. Esto debería ser un proceso iterativo donde los hallazgos y resultados de una etapa de gestión de riesgos alimenten a las demás, alcanzando una suerte de escenario de mejora continua. Y es en este sentido en el que es fácil identificar las diferencias entre un sistema de gestión de riesgos y una evaluación de impacto algorítmico en derechos fundamentales, que tiene una visión exclusiva en el riesgo y sus derivadas, un enfoque mayor en cuanto a la potencial afectación (colectivos afectados, derechos y principios afectados, duración en el tiempo, proporcionalidad, etc.) pero mucho más limitado por lo que se refiere a un tratamiento, procesamiento o empleo concreto de la tecnología (IA) en cuestión.

Los riesgos de la inteligencia artificial pueden evaluarse en diferentes niveles, incluyendo a nivel de gobernanza y proceso, centrándose en riesgos relacionados con principios basados en valores (por ejemplo, la responsabilidad), y a nivel técnico, centrándose en riesgos técnicos (por ejemplo, robustez y rendimiento), y sub-riesgos subyacentes (por ejemplo, precisión estadística).

Un paso hacia garantizar la responsabilidad en la IA es vincular los principios, derechos y riesgos con atributos procedimentales y técnicos específicos. Si bien algunos marcos (modelos de sistemas de gestión de riesgos en entornos IA) existentes proporcionan a los actores de la IA una guía sustancial, como la taxonomía de confiabilidad de la inteligencia artificial en Newman⁷ (2023) o la taxonomía de las

6. Véase OECD (2018), *OECD Due Diligence Guidance for Responsible Business Conduct*, disponible en <http://mneguidelines.oecd.org/OECD-Due-Diligence-Guidance-for-Responsible-BusinessConduct.pdf> (fecha de última consulta: 9 de marzo de 2024).

7. Newman, J., «A Taxonomy of Trustworthiness for Artificial Intelligence: Connecting Properties of Trustworthiness with Risk Management and the Lifecycle», *UC Berkeley*, 2023, disponible en https://cltc.berkeley.edu/wpcontent/uploads/2023/01/Taxonomy_of_AI_Trustworthiness.pdf (fecha de última consulta: 10 de marzo de 2024).

garantías jurídicas de la IA en Simón⁸ (2023), convertir los principios basados en valores en requisitos y atributos técnicos específicos es un campo en evolución, y que en cualquier caso no puede resultar exhaustivo en la medida en la que no existe un modelo de gestión ideal, sino tantos modelos como empresas y administraciones con procesos de negocio, contextos, tratamientos de datos, naturaleza y alcance de la IA y su uso singular o único.

II. EVOLUCIÓN DEL SIGNIFICADO, CONTENIDO Y DESTINATARIOS DE LA OBLIGACIÓN DE CONTAR CON UN SISTEMA DE GESTIÓN DE RIESGOS (ARTÍCULO 9 REGLAMENTO)

La normativa objeto de análisis en este capítulo y, más concretamente, la obligación jurídica de diseñar, implementar y monitorizar un sistema de gestión de riesgos, se recoge en el capítulo III del RIA, que lleva por título «Sistemas de IA de alto riesgo», y que contiene las reglas de clasificación de los sistemas de IA de alto riesgo en su sección primera, mientras que la sección segunda, donde se encuentra ubicada la obligación del artículo 9 del RIA, establece los requisitos mínimos preceptivos de los sistemas de IA de alto riesgo.

Estos requisitos son, a su vez, una derivada de las directrices éticas para una inteligencia artificial fiable que fueron elaboradas por el grupo independiente de expertos de alto nivel sobre inteligencia artificial que se creó por la Comisión Europea, en junio de 2018⁹. Se considera la adaptabilidad en relación con las soluciones técnicas requeridas para alcanzar la conformidad con los requisitos mencionados, los cuales pueden derivar de normativas o especificaciones técnicas, o bien ser desarrollados conforme a los conocimientos científicos o sectoriales concretos.

En este sentido, se concede un amplio margen de discreción al proveedor del sistema de inteligencia artificial para determinar cómo satisfacer los requisitos, tomando en consideración el estado actual de la tecnología y los avances científicos y tecnológicos. Así las cosas, estamos ante requisitos mínimos obligatorios que se pueden alcanzar de distintos modos: en el ámbito del artículo 9 del RIA, se puede optar por modelos diseñados de forma exclusiva o única en el marco del contexto, alcance y naturaleza de la organización y de la IA a utilizar; o seguir los requisitos de ciertos estándares internacionales tales como la ISO/IEC 42001:2023, la ISO/IEC TR 24030:2021 y la ISO/IEC TR 5469:2024 (que son brevemente resumidas, entre otras normas que ayudan a la interpretación de estos tres estándares —por ejemplo, la norma 24027:2023 que aborda los sesgos o la norma 22989:2022 que aborda los conceptos y la terminología—, en el capítulo que rubrica, más adelante, Eduard Chaveli), o los modelos NIST 800 218 PW.I.I.; NIST 800 218RV.1.I.; o el modelo de gestión de riesgos de IA de la OCDE denominado *High-level AI risk-management interoperability framework*.

La ISO 42001 tiene como objetivo ayudar a las organizaciones a desempeñar responsablemente su papel con respecto a los sistemas de IA (por ejemplo, utilizar,

8. Simón Castellano, P. «Taxonomía de las garantías jurídicas en el empleo de los sistemas de inteligencia artificial», *Revista de Derecho Político*, n.º 117, 2023, pp. 153-196.

9. Unión Europea, *Directrices Éticas para una IA fiable. Grupo de expertos de alto nivel sobre inteligencia artificial*, Comisión Europea, Bruselas 2019.

desarrollar, monitorear o proporcionar productos o servicios que utilicen IA). La IA plantea consideraciones específicas como el uso de IA para la toma de decisiones automáticas o automatizadas, a veces de manera no transparente y no explicativa, puede requerir un manejo específico más allá del manejo de los sistemas de IT clásicos; el uso de análisis de datos, percepción y aprendizaje automático, en lugar de la lógica codificada por humanos para diseñar sistemas, lo que aumenta las oportunidades de aplicación de los sistemas de IA y cambia la forma en que se desarrollan, justifican y despliegan dichos sistemas; los sistemas de IA que realizan un aprendizaje continuo y cambian su comportamiento durante el uso, lo que requiere una consideración especial, en sintonía con la mejora continua y el carácter iterativo de los sistemas de gestión, para garantizar que su uso responsable continúe ante un comportamiento cambiante.

La ISO 42001 proporciona requisitos para establecer, implementar, mantener y mejorar continuamente un sistema de gestión de IA en el contexto de una organización. Se espera que las organizaciones centren su aplicación de requisitos en características que son únicas para la IA. Ciertas características de la IA, como la capacidad de aprender y mejorar continuamente o la falta de transparencia o explicabilidad, pueden justificar diferentes salvaguardias si plantean preocupaciones adicionales en comparación con cómo se realizaría tradicionalmente la tarea.

La adopción de un sistema de gestión de IA para ampliar las estructuras de gestión existentes es una decisión estratégica para una organización. Las necesidades y objetivos de la organización, los procesos, el tamaño y la estructura, así como las expectativas de las diversas partes interesadas, influyen en el establecimiento e implementación del sistema de gestión de IA. Otro conjunto de factores que influyen en el establecimiento e implementación del sistema de gestión de IA son los muchos casos de uso para la IA y la necesidad de encontrar el equilibrio adecuado entre los mecanismos de gobernanza y la innovación. Las organizaciones pueden optar por aplicar estos requisitos utilizando un enfoque basado en el riesgo para garantizar que se aplique el nivel adecuado de control para los casos de uso, servicios o productos de IA particulares dentro del alcance de la organización. Se espera que todos estos factores influyentes cambien y se revisen de vez en cuando.

El sistema de gestión de IA debe integrarse con los procesos de la organización y la estructura de gestión general. Los problemas específicos relacionados con la IA deben considerarse en el diseño de procesos, sistemas de información y controles. El modelo que propone la ISO 42001 fija una serie de pautas para la implementación de controles aplicables para respaldar dichos procesos, y evita orientaciones específicas sobre procesos de gestión. La organización puede combinar marcos generalmente aceptados, otros estándares internacionales y su propia experiencia para implementar procesos cruciales como la gestión de riesgos, la gestión del ciclo de vida y la gestión de la calidad de los datos que sean apropiados para los casos de uso, productos o servicios de IA específicos dentro del alcance.

La propia ISO 42001 indica que una organización que cumple con los requisitos de su estándar es una organización que puede generar evidencia de su responsabilidad y rendición de cuentas con respecto a su papel en relación con los sistemas de IA; siendo un modelo que aplica una estructura armonizada (números de cláusula idénticos, títulos de cláusulas, texto y términos comunes y definiciones centrales)

desarrollada para mejorar la alineación entre los estándares de sistemas de gestión, o lo que es lo mismo, que la hace compatible con otros estándares internacionales de sistemas de gestión de riesgos de IA. El sistema de gestión de IA proporciona requisitos específicos para gestionar los problemas y riesgos derivados del uso de IA en una organización. Este enfoque común facilita la implementación y la coherencia con otros estándares de sistemas de gestión, por ejemplo, relacionados con la calidad, seguridad, seguridad y privacidad.

Por su parte, otro buen ejemplo lo encontramos, como decíamos anteriormente, con el modelo de gestión de riesgos de IA de la OCDE denominado *High-level AI risk-management interoperability framework*. En la figura 8 vemos la estructura que la OCDE propone para el diseño y los componentes mínimos (principios, ciclo de vida del sistema de IA y fases en la gestión del riesgo) de un sistema de gestión de riesgos en el contexto de uso de un sistema basado en IA; en la figura 9, en cambio, podemos comprobar los componentes a través de una vista funcional en la que se pone de relieve la importancia de la comunicación y consulta, de las evidencias documentales y de los procesos de monitorización y revisión para alcanzar la mejora continua a través de todo el ciclo de vida útil del sistema de IA.

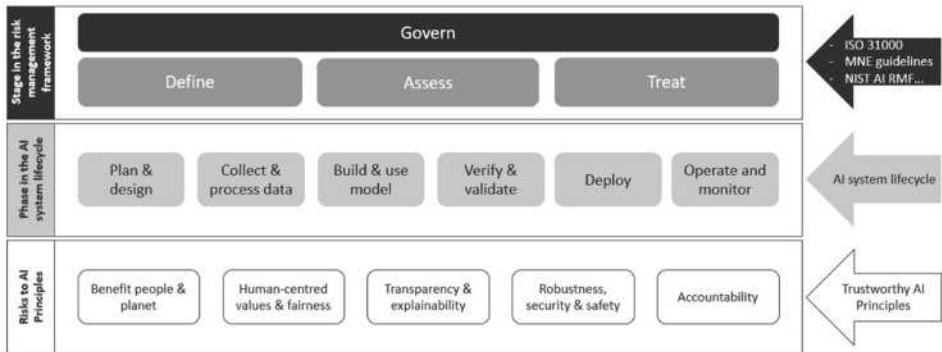


Figura 8. Estructura de un sistema de gestión de riesgos e interoperabilidad para un sistema de IA de alto riesgo. Fuente: Informe OCDE intitolado «Advancing accountability in AI» disponible en <https://doi.org/10.1787/2448f04b-en>

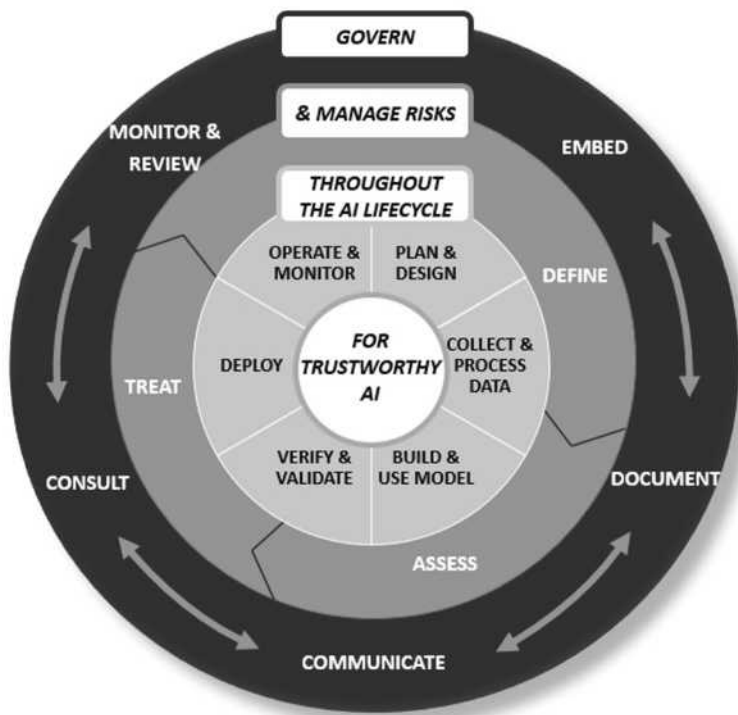


Figura 9. Vista funcional de un sistema de gestión de riesgos e interoperabilidad para un sistema de IA de alto riesgo. Fuente: Informe OCDE intitulado «Advancing accountability in AI» disponible en <https://doi.org/10.1787/2448f04b-en>

Gobernar el proceso de gestión de riesgos es clave para lograr una inteligencia artificial confiable. La gobernanza es una actividad transversal que consta de dos elementos principales. El primer elemento se refiere a la gobernanza del propio proceso de gestión de riesgos e incluye la supervisión y revisión, documentación, comunicación y consulta sobre el proceso y sus resultados. El segundo elemento de gobernanza asegura la efectividad del proceso de gestión de riesgos al incorporarlo en la cultura y los procesos de gobernanza más amplios de las organizaciones.

Sea como fuere, los requisitos mínimos del capítulo III del RIA para los sistemas de riesgo alto, dentro de los que se encuentra ubicado en la cabecera la significativa y preceptiva mención a los sistemas de gestión de riesgos, se trata de un conjunto de obligaciones horizontales que se imponen a los proveedores de sistemas de IA de alto riesgo¹⁰, aunque el RIA en su capítulo III también fija una serie de requisitos

10. Sobre esta cuestión véase Simón Castellano, P. *Justicia cautelar e inteligencia artificial: la alternativa a los atávicos heurísticos judiciales*, J. M. Bosch, 2021; Cotino Hueso, L., «Los usos de la inteligencia artificial en el sector público, su variable impacto y categorización jurídica», *Revista Canaria de Administración Pública*, n.º 1, 2023, pp. 211-242;

mínimos u obligaciones para los usuarios (entiéndase usuario en el sentido del RIA, es decir, cualquier empresa u organización que utiliza o emplee sistemas de IA; no equiparando en ninguna caso a los «usuarios» con los usuarios finales o destinatarios fruto de la aplicación de los sistemas de IA) y otros agentes o actores como pueden ser los importadores, distribuidores y representantes autorizados¹¹. Con ello se pretende reforzar la eficacia de los derechos y recursos existentes mediante el establecimiento de requisitos y obligaciones específicos, en particular en materia de transparencia, documentación técnica y registro de los sistemas de IA. Deben aplicarse requisitos a los sistemas de IA de alto riesgo en lo que respecta a la gestión de riesgos, la calidad y pertinencia de los conjuntos de datos utilizados, la documentación técnica y el mantenimiento de registros, la transparencia y el suministro de información a los usuarios, la supervisión humana, y la solidez, precisión y ciberseguridad. Estos requisitos son necesarios para mitigar eficazmente los riesgos que potencialmente el uso de la IA proyecta para bienes jurídicos tan diversos como la privacidad, la seguridad, la salud, la igualdad o la libertad de información, entre muchos otros.

1. QUÉ ES UN SISTEMA DE GESTIÓN DE RIESGOS SEGÚN EL REGLAMENTO. EL CONTENIDO DE LA OBLIGACIÓN

El legislador europeo ha realizado un esfuerzo ingente para concretar la obligación o el requisito específico contenido en el artículo 9 del RIA, y lo ha hecho a través de los considerandos, cuyas principales ideas vamos a tratar de aunar y sistematizar en las próximas líneas. Un sistema de gestión de riesgo de la IA es un conjunto de medidas diseñadas para identificar, evaluar y mitigar los riesgos asociados a los sistemas de IA considerados de alto riesgo, que se colocan en el mercado o que se ponen en servicio. El objetivo es asegurar un alto nivel de confiabilidad y responsabilidad de los sistemas de alto riesgo, aplicando ciertos requisitos mínimos y preceptivos, teniendo en cuenta el propósito previsto y el contexto de uso del sistema de IA. Las medidas adoptadas por los proveedores para cumplir con los requisitos obligatorios del RIA deben tener en cuenta el estado generalmente reconocido del arte en inteligencia artificial, ser proporcionales y eficaces para cumplir con los objetivos de este (véase Considerando 42 del RIA).

Siguiendo el enfoque del nuevo marco legislativo y la estrategia digital de la Unión, según se establece en el aviso de la Comisión titulada «Guía azul» sobre la aplicación de las normas de la UE sobre productos de 2022 (C/2022/3637), la regla general es que varios fragmentos de la legislación de la Unión Europea pueden tener que tenerse en cuenta para un producto, ya que la puesta a disposición o la puesta en servicio solo puede tener lugar cuando el producto cumple con toda la legislación de armonización de la UE aplicable. Los peligros de los sistemas de IA cubiertos por los requisitos del RIA afectan a aspectos diferentes que las disposiciones de armonización

Presno Linera, M. A. *Derechos fundamentales e inteligencia artificial*, Marcial Pons, 2022; Presno Linera, M. A., «La propuesta de “Ley de Inteligencia Artificial” europea», *Revista de las Cortes Generales*, n.º 116, 2023, pp. 81-133.

11. Al respecto recomendamos el trabajo de Ramón Fernández, F., «Inteligencia artificial y transparencia en relación con la regulación de los servicios y mercados digitales», en Cobas Cobiella, M. E. y Guillén Catalán, R. (directoras), *Equidad y transparencia en la prestación de servicios*, Dykinson, Madrid, 2023, pp. 147-169.

de la Unión existentes. Esto requiere una aplicación simultánea y complementaria de las diversas normas legislativas.

Para garantizar la coherencia y evitar cargas administrativas o costos innecesarios, los proveedores de un producto que contenga uno o más sistemas de IA de alto riesgo, deberían tener flexibilidad en las decisiones operativas sobre cómo garantizar el cumplimiento de un producto que contenga uno o más sistemas de IA con todos los requisitos aplicables de la legislación de armonización de la Unión de la mejor manera posible. Esta flexibilidad no debería de ninguna manera socavar la obligación del proveedor de cumplir con todos los requisitos aplicables, entre los que se encuentra, de forma significativa, el requisito de contar con un sistema operativo de gestión de riesgos.

Detalla el Considerando 42a) del RIA que el sistema de gestión de riesgos debe consistir en un proceso continuo e iterativo que se planifique y ejecute a lo largo de todo el ciclo de vida de un sistema de IA de alto riesgo. La idea de la mejora continua y de la tecnología en movimiento exige que se diseñe un sistema de gestión que, en cualquier caso, debe evolucionar a la medida que cambia el contexto y alcance tecnológico, sus usos concretos dentro de la organización y los efectos que proyecta. Este proceso iterativo, no estático, debe tener como objetivo identificar, evaluar y mitigar los riesgos relevantes de los sistemas de inteligencia artificial para la salud y la seguridad (responsabilidad por daños materiales o productos defectuosos) y, también, los derechos fundamentales de las personas, si bien de forma genérica y no pormenorizada porque ese es el objeto de otra obligación, la prevista *ex art.* 27 RIA, con la evaluación de impacto algorítmico en derechos fundamentales.

El sistema de gestión de riesgos debe ser revisado y actualizado regularmente para garantizar su efectividad continua, así como la justificación y documentación de cualquier decisión significativa y acciones tomadas en virtud del RIA (de nuevo, seguimos el Considerando 42a). Este proceso debe asegurar que el proveedor identifique los riesgos o impactos adversos e implemente medidas de mitigación para los riesgos conocidos y razonablemente previsibles de los sistemas de inteligencia artificial para la salud, la seguridad y los derechos fundamentales, teniendo en cuenta su propósito previsto y el uso razonablemente previsible, incluidos los posibles riesgos derivados de la interacción entre el sistema de IA y el entorno en el que opera.

En la fase de tratamiento, el sistema de gestión de riesgos debe adoptar las medidas de gestión de riesgos más apropiadas a la luz del estado del arte en inteligencia artificial. Al identificar las medidas de gestión de riesgos más apropiadas, el proveedor debe documentar y explicar las decisiones tomadas y, cuando sea relevante, involucrar a expertos y partes interesadas externas. Al identificar el uso razonablemente previsible de los sistemas de IA de alto riesgo, el proveedor debe cubrir usos de los sistemas de IA que, aunque no estén directamente cubiertos por el propósito previsto y especificados en las instrucciones de uso, puedan ser razonablemente esperados como resultado de comportamientos humanos fácilmente previsibles en el contexto de las características específicas y el uso del sistema de IA particular. Cualquier circunstancia conocida o previsible relacionada con el uso del sistema de IA de alto riesgo de acuerdo con su propósito previsto o en condiciones de uso razonablemente previsibles, que puedan dar lugar a riesgos para la salud, la

seguridad o los derechos fundamentales, debe incluirse en las instrucciones de uso proporcionadas por el proveedor.

Una parte de lo que integra el sistema de gestión exige también transparencia. La finalidad es asegurar que el usuario esté al tanto y tenga en cuenta esos riesgos previsibles al utilizar el sistema de IA de alto riesgo. Identificar e implementar medidas de mitigación de riesgos para usos previsibles bajo el RIA no debería requerir medidas de formación adicionales específicas para el sistema de IA de alto riesgo por parte del proveedor para abordarlos. Sin embargo, el RIA anima a los proveedores a considerar tales medidas de formación adicionales para mitigar usos razonablemente previsibles según sea necesario y apropiado.

En la redacción final del artículo 9 del RIA se establece una delimitación clara de lo que se entiende por «sistema de gestión de riesgos» en relación con los sistemas de IA de alto riesgo. Se indica que el sistema de gestión debe ser entendido como proceso iterativo continuo, planificado y ejecutado durante todo el ciclo de vida de sistema tecnológico basado en IA y que requerirá revisiones y actualizaciones sistemáticas periódicas. Como mínimo, el apartado 2 del artículo 9 del RIA exige que conste de las siguientes fases o etapas:

«a) la determinación y el análisis de los riesgos conocidos y previsibles que el sistema de IA de alto riesgo pueda conllevar para la salud, la seguridad o los derechos fundamentales cuando el sistema de IA de alto riesgo se utilice de conformidad con su finalidad prevista;

b) la estimación y la evaluación de los riesgos que podrían surgir cuando el sistema de IA de alto riesgo se utilice de conformidad con su finalidad prevista y cuando se le dé un uso indebido razonablemente previsible;

c) la evaluación de otros riesgos que podrían surgir, a partir del análisis de los datos recogidos con el sistema de vigilancia poscomercialización a que se refiere el artículo 72;

d) la adopción de medidas adecuadas y específicas de gestión de riesgos diseñadas para hacer frente a los riesgos detectados con arreglo a la letra a)».

En el apartado 3 del artículo 9 del RIA se indica que los riesgos a que se refiere el presente artículo son únicamente aquellos que pueden reducirse o eliminarse razonablemente mediante el desarrollo o el diseño del sistema de IA de alto riesgo o el suministro de información técnica adecuada. Esta mención vincula directamente al sistema de gestión con la fase de desarrollo o diseño del sistema (técnicos u organizativos, tanto en la fase de elaboración de código como ante potenciales errores de sesgo en la fase de extracción de datos y entrenamiento) y la fase de suministro de información técnica adecuada (transparencia y derechos de los destinatarios finales de saber qué garantías y razones lógicas operan tras la herramienta de IA). También podría interpretarse ese apartado en el sentido de diferenciar los riesgos con los que de forma pormenorizada y exhaustiva deben valorarse en el seno de la obligación del artículo 27 del RIA, de realizar una evaluación de impacto algorítmico en derechos fundamentales.

En el artículo 9.9 del RIA se hace una mención especial a la necesidad de tutela y protección de los menores de edad, así como de otros grupos vulnerables (en la práctica un concepto jurídico indeterminado que la oficina europea de IA y la AESIA española tendrán que matizar y delimitar), en el sentido que en la implementación o

implantación de cualquier sistema de gestión de riesgos en el ámbito de los sistemas de IA de riesgo alto los proveedores deberán prestar especial atención en el caso que este pueda proyectar efectos negativos sobre las personas menores de dieciocho años y, también, a otros grupos de personas vulnerables, estableciendo medidas o controles *ad hoc* para mitigar esos riesgos.

Muy interesante y relevante es, a efectos prácticos, el contenido del artículo 9.10 del RIA, que indica «en el caso de los proveedores de sistemas de IA de alto riesgo que estén sujetos a requisitos relativos a procesos internos de gestión de riesgos con arreglo a otras disposiciones pertinentes del Derecho de la Unión, los aspectos previstos en los apartados 1 a 9 podrán formar parte de los procedimientos de gestión de riesgos establecidos con arreglo a dicho Derecho, o combinarse con ellos». Esto significa que en caso de coexistencia con otros sistemas de gestión preceptivos por el Derecho de la Unión, los sistemas de gestión podrán coordinarse e incluso ser integrados, ya sea por sistemas de gestión sectoriales (más difícil de imaginar, pero por ejemplo podrían efectuarse integraciones en los sistemas de gestión de riesgos medioambientales y de sostenibilidad o en los sistemas de gestión ética y socialmente responsable) o por sistemas específicos de seguridad de la información (los que exigen ENISA y ENS o los específicos para infraestructuras críticas).

Por su parte, los apartados 4 a 8 del artículo 9 del RIA se centran en las medidas de gestión de riesgos aplicables a los sistemas de IA de alto riesgo. De un lado, efectúa la consideración de efectos e interacciones combinadas, esto es, el RIA destaca la importancia de considerar los efectos y las interacciones resultantes de la aplicación conjunta de los requisitos establecidos en el sistema de gestión. El objetivo es minimizar los riesgos de manera más efectiva al mismo tiempo que se logra un equilibrio adecuado en la aplicación de las medidas para cumplir con dichos requisitos. En otras palabras, se busca encontrar un equilibrio entre la eficacia en la gestión de riesgos y la aplicación equitativa de las medidas requeridas. Lo encontramos en el apartado 4 del artículo 9, cuando se refiere a los efectos y a la posible interacción derivada de la aplicación combinada de requisitos. En el apartado 5, en cambio, pasamos a la evaluación y consideración de los riesgos residuales, en el que se establece que las medidas de gestión de riesgos deben tener en cuenta los riesgos residuales pertinentes asociados a cada peligro, así como el riesgo residual general de los sistemas de IA de alto riesgo. Esto implica que, incluso después de aplicar medidas de mitigación, pueden persistir ciertos riesgos, y es importante evaluar y aceptar estos riesgos residuales de manera adecuada. Se exige que se establezcan allí donde sea técnicamente viable mecanismos de detección y evaluación en el diseño y desarrollo de soluciones de IA, que se implementen medidas de reducción y control y que se realice formación a los responsables de los controles y a los responsables del despliegue. Los sistemas de IA de alto riesgo serán sometidos a pruebas destinadas a determinar cuáles son las medidas de gestión de riesgos más adecuadas y específicas. Dichas pruebas comprobarán que los sistemas de IA de alto riesgo funcionan de manera coherente con su finalidad prevista y cumplen los requisitos mínimos obligatorios.

En definitiva, estos artículos del RIA enfatizan la importancia de una gestión integral de riesgos para los sistemas de IA de alto riesgo, que incluye la consideración de los efectos combinados de las medidas y la evaluación de los riesgos residuales asociados a estos sistemas. También subrayan la necesidad de diseñar y desarrollar

los sistemas de IA de manera que se minimicen los riesgos en la medida de lo posible, proporcionando información y formación adecuadas a los responsables del despliegue. Sistema de IA de alto riesgo afecte negativamente a las personas menores de dieciocho años y, en su caso, a otros grupos de personas vulnerables. Y abre la puerta a los entornos de prueba, indicando que las pruebas de los sistemas de IA de alto riesgo se realizarán, según proceda, en cualquier momento del proceso de desarrollo y, en todo caso, antes de su introducción en el mercado o puesta en servicio. Las pruebas se realizarán utilizando parámetros y umbrales de probabilidades previamente definidos que sean adecuados para la finalidad prevista del sistema de IA de alto riesgo.

2. SUJETOS OBLIGADOS. ¿QUIÉN ESTÁ OBLIGADO A CONTAR CON UN SISTEMA DE GESTIÓN DE RIESGOS? CUADRO RESUMEN DE LAS OBLIGACIONES VINCULADAS A LOS SISTEMAS DE RIESGO ALTO

Para comprender mejor la obligación de contar con un sistema de gestión de riesgos hay que entender, a su vez, el microcosmos de agentes o actores en el ámbito del RIA y las distintas obligaciones y requisitos que se proyectan sobre unos y otros; lo que para unos es una obligación o requisito mínimo para otros puede ser una exigencia de comprobar que un tercero ha obtenido una certificación, ha implementado un sistema de gestión o ha cumplido en tiempo y forma las derivadas del RIA. A continuación, se ofrece un cuadro resumen de las obligaciones para los sistemas de riesgo alto y se señalan en negrita aquellas derivadas o vinculadas a la existencia de un sistema de gestión, aunque pueda resultar o estar interrelacionada con otras variables.

<p>Sistemas de IA de alto riesgo Requisitos mínimos que cumplir por los sistemas</p>	<ul style="list-style-type: none"> • Los proveedores de sistemas de alto riesgo deberán: • Establecer, implementar, documentar i mantener un sistema de Gestión de Riesgos asociado al sistema de IA, con el objetivo de reducir al mínimo los riesgos para los usuarios y personas afectadas y demostrando que se cumplen los requisitos de la legislación vigente, incluso después de que los productos se hayan comercializado. Pondrá especial atención a los riesgos sobre la salud, seguridad y los derechos fundamentales. • Establecer un sistema de Gobernanza y Gestión de los Datos de entrenamiento y prueba, asegurando buenas prácticas en su diseño, recolección y preparación. Además, tendrán que asegurar su relevancia, corrección y sus apropiadas propiedades estadísticas, evitando sesgos que afecten negativamente a las personas. • Los sistemas de IA de Alto Riesgo deberán ir acompañados de Documentación Técnica Actualizada que demuestre que se cumplen los requisitos exigidos antes de su puesta en el mercado, y durante todo el tiempo que se encuentre en el mercado. • Tomarán Registros de Actividad del Sistema («logs») de forma automática durante toda la vida del sistema. • Los sistemas de IA de Alto Riesgo tendrán que ser diseñados y desarrollados de tal manera que garantice que su operación es suficientemente transparente (se tendrá en especial consideración en el diseño y desarrollo del sistema de IA en el marco de la evaluación y tratamiento de riesgos, en especial cuando haya potenciales colectivos vulnerables o menores de edad que puedan ser usuarios finales o destinatarios de esas herramientas) para habilitar a los usuarios a interpretar la salida del sistema y utilizar dicha información de forma apropiada. Se aportará información tal como las capacidades del sistema, sus requisitos de equipamiento, su ámbito de aplicación, su nivel de precisión, los sistemas de supervisión humana, etc.
---	---

- Deberán permitir que los sistemas de IA de Alto Riesgo puedan ser **supervisados por personas** durante su uso para minimizar los riesgos a la salud, seguridad y DDFF, con especial atención a aquellos riesgos residuales tras la aplicación de medidas de mitigación. Los usuarios podrán monitorizar los sistemas e interpretar su información de salida. Para la identificación biométrica remota en tiempo real, la salida requerirá de la verificación y confirmación separada por al menos dos personas físicas (con algunas excepciones contenidas en la ley). La supervisión humana es un elemento intrínseco a los sistemas de gestión, con definición de usuarios, asignación de roles con poderes distintos en función de su papel en la gestión y tratamiento de riesgos (en especial, responsables de los controles) y garantía mediante la comunicación y consulta del sistema.
- **Asegurar un nivel adecuado de precisión, robustez y ciberseguridad**, que se declarará en la documentación técnica que los acompaña. Al respecto nos remitimos al capítulo que en esta misma obra colectiva ha escrito la catedrática Francisca Ramón Fernández. El sistema de gestión implica supervisar que el diseño de la herramienta de IA se realice con la máxima resistencia posible frente a errores, sesgos, fallos o incoherencias que puedan producirse, especialmente en la interacción con otras personas o sistemas. Incorporarán en todo caso, medidas de ciberseguridad apropiadas y proporcionales a sus circunstancias, con especial atención a la protección contra la manipulación de los datos de entrenamiento.

<p>Sistemas de IA de alto riesgo</p> <p>Obligaciones de los proveedores, usuarios y terceras partes</p>	<p>Como consecuencia de la diversidad de sujetos que toman parte en la puesta en marcha, comercialización y funcionamiento de los sistemas de IA, y en especial, los de alto riesgo, en el RIA se establecen obligaciones diferenciadas para cada una de ellas.</p> <ul style="list-style-type: none"> • Proveedores: Se entiende por proveedores toda persona física o jurídica, autoridad pública, agencia u organismo de otra índole que desarrolle un sistema de IA o para el que se haya desarrollado un sistema de IA con vistas a introducirlo en el mercado o ponerlo en servicio con su propio nombre o marca comercial, ya sea de manera remunerada o gratuita. Tendrán las siguientes obligaciones: <ul style="list-style-type: none"> • Asegurarán que sus sistemas de IA cumplen con los requisitos del apartado anterior (requisitos mínimos que cumplir por los sistemas), informando además del nombre o marca comercial y la dirección donde pueda ser contactado. • Contarán con un sistema de gestión de calidad documentado y actualizado, manteniendo la documentación completa del sistema (de nuevo, nos remitimos a la contribución en esta obra colectiva de la profesora Francisca Ramón Fernández). • Custodiarán los registros («logs») del sistema que estén bajo su control. • Garantizarán que el sistema de IA se someta al correspondiente procedimiento de evaluación de conformidad antes de su comercialización o puesta en servicio. • Colaborarán con las autoridades registrando el sistema, demostrando el cumplimiento de todos los requisitos exigibles por el Reglamento cuando se les sea requerido y notificando los incumplimientos y riesgos que detecten, así como las acciones correctivas tomadas en consecuencia. • En el caso de tratarse de un Proveedor establecido fuera de la UE, antes de comercializar sus sistemas en el mercado de la UE, deberán designar mediante mandato escrito un representante autorizado que se encuentre en la UE. • Importadores: Se entiende por importador toda persona física o jurídica establecida en la Unión que introduzca en el mercado o ponga en servicio un sistema de IA que lleve el nombre o la marca comercial de una persona física o jurídica establecida fuera de la Unión. Antes de introducir el sistema en el mercado tendrán que asegurar que es conforme a la regulación verificando que:
--	---

	<ul style="list-style-type: none"> • El proveedor del sistema haya llevado a cabo el correspondiente procedimiento de evaluación de conformidad. • El proveedor haya elaborado la documentación técnica necesaria. • El sistema cuenta con el marcado CE de conformidad con lo exigido y vaya acompañado de la declaración de conformidad de la UE y sus instrucciones de uso. • El proveedor haya designado un representante autorizado en la UE. • En caso de no cumplir alguno de estos requisitos, o tener razones suficientes para pensar que dicha documentación es falsificada o acompañada de documentos falsos, el importador deberá abstenerse de introducir al mercado dicho sistema. • Los importadores deberán colaborar con las autoridades competentes e informar de su nombre o marca comercial en el producto, junto con la dirección donde pueda ser contactado. • Distribuidores: Se entiende por distribuidor toda persona física o jurídica que forme parte de la cadena de suministro, distinta del proveedor o el importador, que comercializa un sistema de IA en el mercado de la Unión sin influir sobre sus propiedades. Antes de hacer que un sistema de IA esté disponible en el mercado deberá asegurarse que: <ul style="list-style-type: none"> • El sistema cuenta con el marcado CE de conformidad con lo exigido y vaya acompañado de una copia de la Declaración de conformidad de la UE y sus instrucciones de uso. • Que el proveedor y el importador hayan cumplido con la obligación de indicar el nombre o marca comercial, así como dirección de contacto y que el proveedor cuente con un sistema de gestión de calidad. • En caso de no cumplir alguno de estos requisitos, o tener razones suficientes para pensar que dicha documentación es falsificada o acompañada de documentos falsos, el importador deberá abstenerse de disponer al mercado dicho sistema. • Los distribuidores deberán colaborar con las autoridades competentes e informar de su nombre o marca comercial en el producto, junto con la dirección donde pueda ser contactado. • Usuarios o responsables del despliegue: Se entiende por usuario toda persona física o jurídica, autoridad pública,
--	---

	<p>agencia u organismo de otra índole que utilice un sistema de IA bajo su propia autoridad, salvo cuando su uso se enmarque en una actividad personal de carácter no profesional. Tendrán las siguientes obligaciones:</p> <ul style="list-style-type: none">• Tomar medidas técnicas y organizativas apropiadas para asegurar que el uso de dicho sistema es conforme a las instrucciones de uso que lo acompañan.• Ejercer una supervisión humana del sistema, asegurando la persona encargada de realizarlo tenga la competencia, entrenamiento, autoridad y apoyo necesarios.• Monitorizar el funcionamiento del sistema.• Custodiar los registros («logs») del sistema que estén bajo su control.• Cooperar con las autoridades competentes.• Los usuarios o responsables del despliegue que sean o pertenezcan al sector público, así como los operadores privados que provean servicios públicos o aquellas empresas que evalúan solvencia crediticia y patrimonial y las que evalúan riesgos para fijar precios en seguros de salud y de vida, tendrán que realizar una evaluación adicional sobre el impacto algorítmico en los derechos fundamentales que puede provocar la utilización de dicho sistema (nos remitimos a la diferenciación hecha en el apartado I.2.1 del presente trabajo y al capítulo que rubrica Eduard Chaveli dentro de esta misma obra colectiva).
--	--

<p>Responsabilidades en la cadena de valor</p>	<ul style="list-style-type: none"> • Cualquier distribuidor, importador, usuario (o responsable del despliegue) o tercera parte podrá ser considerado proveedor de un sistema de IA de alto riesgo, y por lo tanto estar sujeto a las obligaciones que exige a los proveedores el Reglamento en los siguientes casos: <ul style="list-style-type: none"> • Si ponen su nombre o marca comercial en un sistema de IA de Alto Riesgo ya comercializado o puesto en servicio, sin perjuicio de los acuerdos contractuales que estipulen que las obligaciones se asiguen de otro modo. • Si introduce una modificación sustancial en un sistema de IA de Alto Riesgo que ya ha sido comercializado o puesto en servicio de forma que siga siendo un sistema de Alto Riesgo. También aplicará en aquellos casos en que se modifique la finalidad prevista por un sistema de IA, incluidos los sistemas de IA de uso general, que no hayan sido clasificados como de alto riesgo y que ya haya sido comercializado o puesto en servicio, de tal manera que el sistema de IA se convierta en uno de Alto Riesgo. • En el caso de sistemas de IA de Alto Riesgo que sean componentes de seguridad de productos, fabricante será considerado proveedor del sistema de IA en los siguientes supuestos: <ul style="list-style-type: none"> • El sistema de IA se comercializa junto con el producto bajo el nombre o marca comercial del fabricante del producto. • El sistema de IA se pone en servicio con el nombre o marca comercial del fabricante del producto. • El sistema de IA se pone en servicio bajo el nombre o la marca del fabricante del producto después de que el producto haya sido comercializado.
--	---

Figura 10. Cuadro resumen de obligaciones de los sistemas de riesgo alto e identificación de sujetos obligados para cada una de estas (obligaciones de los proveedores, usuarios o responsables del despliegue, importadores, distribuidores y terceras partes). Fuente: Font Advocats

III. RECAPITULACIÓN Y CONCLUSIONES

En el presente trabajo hemos analizado los sistemas de gestión riesgos como requisito mínimo obligatorio de los sistemas de IA clasificados como de riesgo alto y su relación con otras obligaciones específicas de los proveedores, resaltando en cualquier caso las notables diferencias entre esta obligación y otras, algunas únicamente aplicables a los proveedores, tales como la exigencia de someter al sistema de IA al correspondiente procedimiento de evaluación de conformidad o las relativas a mantener un sistema de gestión de calidad y conservación de la documentación técnica (arts. 11, 17 y 18 del RIA), y otras que despliegan efectos, en lugar de a los proveedores, únicamente a los «antiguos» usuarios, que, en la última versión del texto convertido en Ley europea, se refiere a los «responsables del

despliegue», entre las que destaca la significativa y preceptiva obligación de realizar evaluaciones de impacto algorítmico en los derechos fundamentales.

Con la pretensión de sintetizar al máximo posible, podemos señalar que las conclusiones que hemos alcanzado son las siguientes:

Primera. Sobre el contenido de un sistema de gestión de riesgos. El sistema de gestión de riesgos es un proceso iterativo, de mejora continua, planificado y ejecutado durante todo el ciclo de vida de un sistema de IA de alto riesgo, que requerirá revisiones y actualizaciones sistemáticas periódicas. Integrado como mínimo por tres fases (identificación, evaluación y tratamiento) exige incorporar responsables y usuarios con roles y funciones dentro del sistema, que permitan en especial monitorizar los niveles de riesgo residual y la eficacia y vulnerabilidad de los controles. Como es una obligación que se proyecta sobre los proveedores o fabricantes, contar con un sistema de gestión es una exigencia que nace desde el mismo diseño de la solución tecnológica y no finaliza con la primera comercialización, sino que también requiere de vigilancia poscomercialización (en relación con el artículo 72 del RIA). Como mínimo, el sistema de gestión debe identificar y los riesgos conocidos y previsibles que el sistema de IA de alto riesgo pueda conllevar para la salud, la seguridad o los derechos fundamentales, así como estimar y evaluar los niveles de riesgo en relación con los posibles usos y finalidades previstas. Ese análisis y evaluación permitirá detectar las necesidades y vulnerabilidades, que se cubrirán mediante la adopción de controles y medidas adecuadas y específicas de gestión de riesgos diseñadas para hacer frente a los riesgos detectados previamente.

Segunda. De la necesidad de diferenciar esta obligación de la llamada evaluación de impacto algorítmico en derechos fundamentales (artículo 27 RIA). Existe una diferencia sustantiva importante. El sistema de gestión se compone de tres fases y la evaluación de impacto algorítmico en derechos fundamentales debería enmarcarse práctica y exclusivamente en la fase de evaluación, puesto que si el resultado es que el despliegue de la IA supone un riesgo no aceptable desde la óptica de los derechos fundamentales entonces no hay ningún control que aplicar ni fase de tratamiento posible. Sólo una evaluación de impacto cuyo resultado sea favorable permitirá que se despliegue por parte del usuario o responsable del despliegue esa IA, clasificada como de alto riesgo, en el ámbito empresarial o público concreto (contexto, alcance y naturaleza) y para la finalidad concreta del supuesto. No es lo mismo disponer de un sistema de gestión desde la incepción o diseño de una tecnología, incluyendo todo el ciclo de vida del sistema de IA, que realizar una evaluación de impacto algorítmico en derechos fundamentales para un uso concreto en un contexto empresarial determinado. Las obligaciones, cuyo contenido es distinto, también tienen destinatarios o sujetos obligados distintos. La primera afecta a los proveedores (y a los distribuidores e importadores en la medida que deben comprobar y verificar que el proveedor cuenta con un sistema de gestión de riesgos operativo), mientras que la evaluación de impacto algorítmico en derechos fundamentales afecta exclusivamente a los «usuarios» o «responsables del despliegue». Y, además, se trata de una obligación que sólo nace o aplica para determinados responsables del despliegue, más concretamente, únicamente aplica a aquellos que son organismos de Derecho público, o a las entidades privadas que prestan servicios públicos, o a las empresas que evalúan solvencia crediticia y patrimonial y las que evalúan riesgos para fijar precios en seguros de salud y de vida. En lo sustantivo, la evaluación

de riesgos del sistema de gestión tiene un enfoque global *versus* el enfoque más selectivo sólo en derechos fundamentales de la evaluación de impacto. El enfoque más global del sistema de gestión está más concentrado en el diseño, comercialización y poscomercialización de la herramienta basada en IA, mientras que evaluación de impacto está centrado en el despliegue o uso concreto de una IA en el seno de una empresa u organismo público.

Tercera. De la inexistencia de un modelo ideal de sistema de gestión de riesgos en el ámbito de las IA clasificadas como de alto riesgo. Modelos disponibles aceptados por el mercado y en perspectiva comparada. A lo largo de este trabajo hemos analizado que el RIA concede un amplio margen de discreción al proveedor del sistema de inteligencia artificial para determinar cómo satisfacer y fijar los requisitos mínimos y componentes del sistema de gestión de riesgos de IA, tomando en consideración el estado actual de la tecnología y los avances científicos y tecnológicos. El objetivo, contar con un sistema de gestión, se pueden alcanzar de distintos modos: se pueden seguir los componentes, los requisitos y la metodología de ciertos estándares internacionales tales como la ISO/IEC 42001:2023, la ISO/IEC TR 24030:2021 y la ISO/IEC TR 5469:2024; o los modelos de la norma técnica NIST 800 218 PW.II y NIST 800 218RV.1.I.; o el modelo de gestión de riesgos de IA de la OCDE denominado *High-level AI risk-management interoperability framework*, entre muchos otros. Incluso en el caso de aceptar los modelos y metodologías indicadas anteriormente están habrán de ser adaptadas al contexto de cada organización, al alcance y naturaleza del ciclo de vida de la IA en cuestión y a los tratamientos u operaciones de procesamiento de datos (base de datos que nutre el algoritmo, extracción, datos de entrenamiento, etc.). No existe, pues, un modelo de gestión de riesgos único o «ideal». Además, los sistemas de gestión de riesgos de IA deben coordinarse con los otros modelos de gestión existentes en la organización (calidad, riesgos medioambientales, seguridad de la información, esquemas de seguridad, etc.), hasta el punto que en algunos casos puede producirse una «integración» del sistema de gestión de riesgos de IA dentro de otros modelos de gestión específicos (en especial, en aquellos de seguridad de la información o en los de cumplimiento normativo).

Datos y gobernanza de datos y conexiones con principios protección de datos en el artículo 10 del Reglamento

MARÍA LOZA CORERA

Doctora en Derecho. Lead Advisor en Govertis parte de Telefónica Tech
Profesora de la Universidad Internacional de La Rioja

I. INTRODUCCIÓN

En el mundo actual, ya nada puede entenderse sin datos, ni siquiera el pasado. Los datos son un activo esencial. En el contexto de la llamada economía digital, los datos juegan un papel de capital importancia, hasta el punto de hablar de la economía de los datos o economía basada en los datos¹. En este contexto, incluso se ha llegado a mencionar a la Inteligencia Artificial como de los activos intangibles más valiosos de cualquier empresa por ser un impulsor del valor organizacional². No obstante, la tecnología no es neutra³, ni la aproximación para la regulación del riesgo que se utilice⁴, por lo que el diseño y los datos, son absolutamente relevantes y buena prueba de ello es el RIA (en adelante, RIA). Las consecuencias de no contar con el tipo de datos adecuado, ni con la calidad requerida, podrían ser nefastas, al condicionar desde el diseño los resultados de la concreta solución de IA adoptada, siendo, por tanto, no válidos, y lo que es más importante, pudiendo afectar a la seguridad y/o los derechos fundamentales de las personas. La relación entre los datos y el sistema de IA es, por tanto, directamente proporcional a la calidad de los resultados obtenidos. No obstante lo anterior, no sólo será necesario disponer de conjuntos adecuados de datos y de calidad, sino que además, resulta imprescindible relacionar esos datos con la tecnología adecuada, los concretos procedimientos

1. LOZA CORERA, M., *De los microdatos a los datos masivos. Cuestiones legales*, Universidad de Valencia, 2017, p. 259.
2. WITZEL M. y BHARGAVA N., «AI-Related Risk The Merits of an ESG-Based Approach to Oversight», CIGI Papers N.º 279, agosto 2023. <https://www.cigionline.org/static/documents/no.279.pdf>
3. FLORIDI, L., «On Good and Evil, the Mistaken Idea That Technology is Ever Neutral, and the Importance of the Double-charge Thesis». *Philosophy & Technology*, septiembre 2023, disponible SSRN: <https://ssrn.com/abstract=4551487>
4. KAMINSKI M., «Regulating de risk of AI», 2022, Boston University Law Review, Vol. 103:1347, 2023, U of Colorado Law Legal Studies Research Paper No. 22-21, disponible SSRN: <https://ssrn.com/abstract=4195066> p. 1351.

internos y para unas determinadas finalidades determinadas por la organización, sin olvidar el cumplimiento de los diferentes marcos normativos aplicables, en otras palabras, establecer un sistema de gobernanza. En un momento en que ya se habla de la transición hacia la economía cuántica⁵, es *conditio sine qua non* establecer un adecuado sistema de gobernanza que permita dicha transición.

El término Gobernanza bien podría ser uno de esos «*suitcase words*» que Marvin Minsky definió como aquellas palabras que encierran múltiples significados. Por ello, resulta necesario clarificar los diferentes significados que dicho término encierra, aunque, como veremos, se encuentran plenamente relacionados, especialmente en el ámbito de la Inteligencia Artificial. En primer lugar, se abordará el concepto de gobernanza de datos, teniendo en cuenta la vital importancia de los datos para un sistema de inteligencia artificial. Posteriormente, se analizará la importancia de la gobernanza de datos en el contexto normativo europeo actual y el significado que alcanza en dicho contexto. Y finalmente, se analizará el concepto de gobernanza de datos en el RIA, más cercano al concepto de equidad de datos (*data equity*).

El RIA dedica todo un artículo (Artículo 10) dentro del capítulo III dedicado a los Sistemas de IA alto riesgo, a los *Datos y gobernanza de datos*, consciente de la vital trascendencia que los datos y la gobernanza de los mismos tienen, dentro de un sistema de IA. Podemos afirmar sin ningún género de dudas que se trata de uno de los artículos nucleares del Reglamento, ya que no disponer de conjuntos adecuados de datos, impedirá de inicio la puesta en marcha de un sistema de IA, no sólo por los posibles sesgos inherentes a los datos subyacentes, sino porque la IA también *aprende* de los datos. Se estudiará en detalle la evolución, tramitación y contenido final de dicho artículo 10, incluyendo todos los cambios y modificaciones ocurridos desde la Propuesta de Reglamento de la Comisión Europea de abril de 2021, pasando por el texto propuesto por el Consejo y las enmiendas aprobadas por el Parlamento Europeo en junio de 2023, hasta su versión definitiva. Indicar que no se tratarán en el presente capítulo las cuestiones relativas a precisión, solidez y sesgos, ya que se abordan de manera específica en el capítulo que rubrica Ana Aba Catoira. El estudio detallado anterior sobre las obligaciones en materia de gobernanza de datos establecidas por el Reglamento nos permitirá realizar una aproximación crítica a la versión final contenido.

Finalmente, realizaremos un breve análisis sobre la relación de la gobernanza de datos con los principios de protección de datos, sin perjuicio del análisis general más extenso en materia de protección de datos que se realiza en el capítulo que rubrica Jesús Jiménez López.

II. GOBERNANZA DE DATOS

1. CONCEPTO DE GOBERNANZA DE DATOS

El concepto de gobernanza no es exclusivo de la gestión de los datos, más bien tiene su origen en otros ámbitos, como la Gobernanza de las Tecnologías de la Información (Gobernanza TI). No obstante, dado el creciente protagonismo que los

5. World Economic Forum, «Quantum Economy Blueprint», enero 2024, disponible en https://www3.weforum.org/docs/WEF_Quantum_Economy_Blueprint_2024.pdf

datos han adquirido en las organizaciones, tanto públicas como privadas, el concepto de gobernanza de datos, de manera proporcional a dicho protagonismo, ha adquirido sustantividad propia configurándose como una verdadera necesidad.

Los datos son un elemento nuclear en la economía digital, hasta el punto de hablar de «economía de los datos», por lo que gestionar adecuadamente este activo empresarial es un presupuesto necesario para poder ser una compañía impulsada o basada en los datos (*data-driven*). Extraer valor de los datos para poder tomar decisiones más conscientes y eficaces es una posibilidad que no puede desconocerse en el contexto económico y tecnológico actual. Es en este punto donde el concepto de gobernanza de datos adquiere toda su importancia.

No existe una definición unívoca o normativa para el concepto de «Gobernanza de datos». En un primer momento, la gobernanza de datos se entendió referida al contexto interno de una organización, únicamente en lo relativo al control y gestión de sus datos y posteriormente ha evolucionado hacia un concepto más amplio y elaborado. Así, el Instituto de Gobernanza de Datos (*Data Governance Institute*) lo define⁶ como «el ejercicio de la toma de decisiones y la autoridad en asuntos relacionados con los datos», y de forma más amplia, como «un sistema de derechos de decisión y responsabilidades para los procesos relacionados con la información, ejecutados según modelos acordados que describen quién puede tomar qué acciones con qué información, y cuándo, bajo qué circunstancias, utilizando qué métodos».

Por su parte, la *Data Management Association* (DAMA) ha creado un marco de referencia para la gestión de datos, *Data Management Body of Knowledge* (DMBoK⁷), donde la Gobernanza de datos ocupa un lugar esencial dentro de la gestión de datos, evidenciando que no se trata de conceptos coincidentes. Es por ello que, la Gobernanza de datos o gobierno del dato se concibe como el «ejercicio de autoridad y control (planificación, monitorización y aplicación) sobre la gestión de los activos de datos».

La Agencia Española de Protección de Datos define⁸ la gobernanza de datos como «la estrategia para la correcta administración y gestión de la política de datos en la organización». La AEPD destaca que las políticas de protección de datos que debe adoptar el responsable del tratamiento en cumplimiento de lo establecido en el Considerando 78 y el artículo 24 del Reglamento General de Protección de Datos⁹ (RGPD), son una parte importante de la política de datos de la organización.

6. <https://datagovernance.com/the-data-governance-basics/definitions-of-data-governance/>

7. El DMBoK se centra en once temas principales: el Gobierno del dato; Arquitectura de datos; Modelado y diseño de datos; Almacenamiento y operaciones de datos; Seguridad de datos; Integración e interoperabilidad de datos; documentos y contenidos; datos maestros y de referencia; Almacenamiento de datos e inteligencia empresarial; Metadatos y Calidad de datos.

8. AEPD, «Gobernanza y política de protección de datos», 2020 <https://www.aepd.es/prensa-y-comunicacion/blog/gobernanza-y-politica-de-proteccion-de-datos>

9. REGLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos).

También indica que, cuando se traten datos personales, deberán añadirse a los objetivos de la Gobernanza de datos:

- Cumplir con los principios de protección de datos.
- Garantizar que los interesados puedan ejercer sus derechos.
- Garantizar la protección de la protección de los datos personales desde el diseño y por defecto, mediante una gestión del riesgo para los derechos y libertades.
- Cumplir con las restantes obligaciones legales derivadas de la normativa de protección de datos.

Salvador Serna¹⁰ pone de relieve que, a pesar de las múltiples aproximaciones al concepto de gobernanza de datos, «existe cierto consenso en asociar la gobernanza de datos a las ideas de: (1) poner en valor los datos como un activo de la organización que debe gestionarse (2) establecer responsabilidades en la toma de decisiones (derechos) y las tareas asociadas (deberes) y (3) establecer pautas y normas para velar por la calidad de los datos y su uso adecuado». A dichas características, añadimos una cuarta, la necesidad de un liderazgo estratégico de la dirección para el establecimiento de un sistema de gobernanza de datos, no dependiendo de un exclusivo departamento o área de la compañía, de forma que, como sistema transversal, sea coherente con los objetivos y cultura de la organización y, por supuesto, con la normativa vigente.

Resulta por tanto esencial contar con un sistema de gobernanza de datos, ya que a través del mismo se logra la gestión integral del dato durante todo su ciclo de vida, tanto a nivel de calidad, protección, seguridad y mantenimiento, como de cumplimiento normativo. Además de obtener el máximo valor de los datos que ayude en la toma de decisiones más eficientes, con una adecuada gobernanza de datos se consigue minimizar riesgos, ahorro de costes al centralizar la gestión de la información, eliminación de silos, mayor calidad del dato y mejora de procesos gracias al sistema de monitorización y mejora continua y, algo muy importante, se establecen las condiciones para permitir la escalabilidad de diferentes soluciones de IA que se puedan adoptar. Por ello, pasamos del concepto de gobernanza de datos a gobernanza de IA, pero siendo el primero un presupuesto necesario para poder hablar del segundo. La industria es bien consciente de la necesidad imperiosa de contar con un sistema de gobernanza en materia de IA, no solo para cumplir con la normativa, sino para impulsar el valor empresarial¹¹.

2. CONTEXTO EUROPEO

El concepto de Gobernanza de datos anterior, al cual podemos denominar «micro», debe necesariamente ponerse en relación con el contexto actual político y normativo de la UE, en concreto, con los mecanismos de gobernanza de datos o requerimientos normativos a nivel «macro» necesarios para posibilitar el mercado único de datos.

10. SALVADOR SERNA, M., (2021), Inteligencia artificial y gobernanza de datos en la Administración Pública: sentando las bases para su integración a nivel corporativo, en *Repensando la administración pública: administración digital e innovación pública*, (pp. 126-148), INAP, 2021.

11. IBM, The urgency of AI governance, 2023. <https://www.ibm.com/downloads/cas/MV9EXNV8>

En 2018 la Comisión Europea lanzó su *Estrategia de Inteligencia Artificial*¹² donde estableció las bases para asegurar que el potencial de la IA sirva para el progreso humano, potenciando la capacidad tecnológica e industrial de la Unión, preparándose para las transformaciones socioeconómicas que originará la IA y mediante el establecimiento de un marco ético y jurídico apropiado, basado en los valores de la Unión y en consonancia con la Carta de los Derechos Fundamentales de la UE. En este camino a seguir, un objetivo claro e imprescindible es incrementar el volumen de datos disponible y facilitar el acceso a los mismos. Así, la Comisión Europea, consciente del valor de los datos tanto para la economía como para la sociedad y, sin renunciar a la protección de los datos personales, ha impulsado la *Estrategia de Datos de la UE*¹³, en el marco de las prioridades políticas fijadas para el período 2019-2024 (*Una Europa adaptada a la era digital*)¹⁴ y de la *Brújula Digital 2030: el enfoque de Europa para el Decenio Digital*¹⁵.

En la *Estrategia Europea de Datos* la Comisión afirma que «El objetivo es crear un espacio único europeo de datos, un verdadero mercado único de datos, abierto a datos procedentes de todo el mundo, en el que los datos personales y no personales, incluidos los datos sensibles de empresas, estén seguros y las empresas también tengan fácil acceso a una cantidad casi infinita de datos industriales de alta calidad, de manera que se impulse el crecimiento y se cree valor, minimizando al mismo tiempo la huella humana medioambiental y de carbono».

Para conseguir ese espacio único europeo de datos que garantice la competitividad mundial¹⁶ y la soberanía¹⁷ de los datos de Europa, tal y como se afirma en la *Estrategia Europea de Datos*, la legislación de la UE debe aplicarse con eficacia de forma que todos los productos y servicios basados en datos cumplan las normas del mercado único de datos. Junto con la legislación adecuada, deberán adoptarse mecanismos «claros y fiables» de gobernanza que permitan el acceso y utilización de los datos, para garantizar el cumplimiento de los objetivos del espacio europeo de datos.

El marco normativo europeo proyectado para posibilitar la materialización de la *Estrategia Europea de Datos* se conforma, entre otros, por el Reglamento 2022/868 de Gobernanza de Datos¹⁸ y el Reglamento 2023/2854 sobre normas armonizadas

12. COM(2018) 237 final, *Inteligencia artificial para Europa*.

13. COM(2020) 66 final, *Una Estrategia Europea de Datos*, Comisión Europea <https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:52020DC0066>

14. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age_es

15. COM(2021) 118 final <https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:52021DC0118>

16. COM(2020) 66 final «Sin embargo, las fuentes de competitividad para las próximas décadas en la economía de los datos se determinan ahora. Esta es la razón por la que la UE debe actuar ya.»

17. El funcionamiento del espacio europeo de datos dependerá de la capacidad de la UE para invertir en la próxima generación de tecnologías e infraestructuras, así como en las competencias digitales, como la alfabetización en materia de datos. Esto, a su vez, incrementará la soberanía tecnológica de Europa en cuanto a tecnologías facilitadoras esenciales y las infraestructuras correspondientes para la economía de los datos.

18. Reglamento (UE) 2022/868 del Parlamento Europeo y del Consejo de 30 de mayo de 2022 relativo a la gobernanza europea de datos y por el que se modifica el Reglamento (UE) 2018/1724 (Reglamento de Gobernanza de Datos).

para un acceso justo a los datos y su utilización (Reglamento de Datos)¹⁹, sin olvidar el Reglamento relativo a un marco para la libre circulación de datos no personales en la Unión Europea²⁰, en coherencia con el significado que al concepto de «datos» se le otorga por los Reglamentos anteriormente mencionados, cuyo significado es mucho más amplio que el concepto de «dato personal».

Debe ponerse de relieve que el mercado único de datos europeo no desconoce que los flujos internacionales de datos son indispensables en los mercados y entornos competitivos actuales, por lo que tiene un enfoque abierto, pero sin renunciar a la protección y los valores europeos.

Vemos, por tanto, cómo se ha ido evolucionando progresivamente, de una regulación centrada en la protección de los datos personales y derechos y libertades de las personas, a una estrategia centrada en el dato (no necesariamente personal) como activo empresarial, centro de la economía de los datos, que necesita de normativa que garantice su disponibilidad, compartición y reutilización segura, pero siempre preservando los valores europeos. Es por ello que, para garantizar el mercado único de datos (gobernanza a nivel «macro») es imprescindible que las organizaciones cuenten con una sólida gobernanza de datos a nivel interno (nivel «micro»), la cual además permitirá avanzar hacia la gobernanza de la inteligencia artificial.

3. CONCEPTO DE GOBERNANZA EN EL ÁMBITO DE LA INTELIGENCIA ARTIFICIAL

El RIA no ofrece una definición de gobernanza aplicada al ámbito de IA. Tampoco encontramos una definición en la ISO *Information technology-Artificial intelligence — Artificial intelligence concepts and terminology ISO/IEC 22989*. La IAPP²¹ define «IA Governance» como «*A system of policies, practices and processes organizations implement to manage and oversee their use of AI technology and associated risks to ensure the AI aligns with an organization's objectives, is developed and used responsibly and ethically, and complies with applicable legal requirements*». De forma muy similar, la industria la define como «*AI governance is a system of rules, practices, processes and tools that help an organization use AI in alignment with its values and strategies, address compliance requirements and drive trustworthy performance*»²². Se afirma que la gobernanza de la IA es probable que sea tan importante como la gobernanza específica de los componentes del propio algoritmo²³.

19. Reglamento (UE) 2023/2854 del Parlamento Europeo y del Consejo de 13 de diciembre de 2023 sobre normas armonizadas para un acceso justo a los datos y su utilización, y por el que se modifican el Reglamento (UE) 2017/2394 y la Directiva (UE) 2020/1828 (Reglamento de Datos).

20. <https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:32018R1807&qid=1696786250350>

21. IAPP, Key Terms for AI Governance, junio 2023. <https://iapp.org/resources/article/key-terms-for-ai-governance/>

22. Op. Cit. IBM, The urgency of AI governance, 2023.

23. De hecho, en 2022, la gobernanza de la IA era la novena prioridad estratégica más importante para las funciones de privacidad. En 2023 es la segunda configurándose como una prioridad estratégica, IAPP-EY Professionalizing Organizational AI Governance Report, p. 9, 2023.

No obstante, con independencia de la existencia de una definición normativa o no, es incuestionable que el concepto de gobernanza adquiere toda su importancia en el ámbito de la IA hasta el punto de trascender al concepto de gobernanza de datos y hablar de gobernanza de la IA. Toda organización debe establecer los procedimientos necesarios para garantizar el cumplimiento de la normativa aplicable, de las medidas de seguridad necesarias y el respeto a los derechos y libertades fundamentales, así como para garantizar la responsabilidad proactiva de la organización y sus órganos de gobierno en la utilización de las diferentes soluciones de IA que decida implementar. De hecho, si tuviéramos que resumir en una palabra el RIA, ésta sería «Gobernanza».

No debe caerse en el error de que dichas obligaciones solo recaen en las entidades que desarrollan los sistemas de IA, sino que aquellas que los diseñan o los despliegan (implementadores o responsables del despliegue) también tienen responsabilidades, por lo que, aunque a diferentes niveles, es necesario que todas las organizaciones establezcan mecanismos de gobernanza de la IA.

Existen diferentes sistemas o marcos de Gobernanza de la IA. En el ámbito del *soft law* destacan en EE.UU., el *Artificial Intelligence Risk Management Framework*²⁴ (AIRMF) del *National Institute of Standards and Technology* (NIST) y en Japón las *Governance Guidelines for the Implementation of AI Principles*²⁵, sin que hasta la fecha exista un marco normativo vinculante en la materia, aunque ya se vislumbra su próxima aprobación en ambos países. Por el contrario, en Europa no existía un marco específico para la gobernanza de la IA hasta la aprobación del Reglamento europeo.

Con independencia de las diferentes aproximaciones en materia de Gobernanza de la IA, el propio concepto de Gobernanza es plenamente coincidente con la afirmación de Novelli, Taddeo y Floridi²⁶ de que la responsabilidad proactiva es una piedra angular de la gobernanza de la inteligencia artificial.

El RIA se refiere específicamente a la gobernanza en relación a los datos. Así, tanto su Considerando 67, como el artículo 10, se refieren a las «prácticas adecuadas de gobernanza y gestión de datos» que analizaremos a continuación. Por tanto, dejando a un lado el Capítulo VII dedicado a la Gobernanza institucional tanto a nivel europeo (Comité Europeo de Inteligencia Artificial), como nacional (autoridades nacionales competentes), el Reglamento se refiere al concepto de gobernanza en relación a los datos, por tanto, a nivel «micro». Ello no quiere decir en absoluto que la gobernanza de IA que establece el Reglamento europeo se agote en dicho artículo más bien dedicado a la gobernanza de datos, sino que debe ponerse en relación con el resto de obligaciones establecidas para los sistemas de IA de alto riesgo, que exigen la implementación de otros procedimientos, tales como procedimientos de evaluación de la conformidad, declaración de conformidad y marcado CE, sistemas de gestión de la calidad que incluyan obligatoriamente, procedimientos de gestión de las modificaciones, técnicas, procedimientos y acciones sistemáticas que se utilizarán

24. <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>

25. AI Governance in Japan, REPORT FROM THE EXPERT GROUP ON HOW AI PRINCIPLES SHOULD BE IMPLEMENTED, 2021, disponible en https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20210709_8.pdf

26. NOVELLI C., TADDEO M., FLORIDI L., Accountability in artificial intelligence: what it is and how it Works, *AI & Soc* (2023) <https://doi.org/10.1007/s00146-023-01635-y>

para el diseño, el control del diseño y la verificación del diseño y control de calidad, sistemas y procedimientos de gestión de datos, sistema de gestión de riesgos, seguimiento poscomercialización, procedimiento de notificación de incidentes graves, procedimientos de registro de toda la documentación y el establecimiento de un marco de rendición de cuentas. Todas estas obligaciones conforman lo que entendemos por gobernanza de la IA en el marco del Reglamento europeo.

Finalmente, cabe citar una perspectiva más amplia de la gobernanza en materia de IA, dirigida a los reguladores, en el sentido de que hay autores que consideran que la normativa que se está proponiendo tanto en Europa, como en Canadá u otros lugares, no es suficiente para prevenir otros riesgos que pueden ocurrir más a largo plazo. Así, KOLT²⁷ afirma que las propuestas normativas para regular la IA «se centran principalmente en los riesgos inmediatos de la IA, en lugar de en riesgos más amplios y a más largo plazo» por lo que «ofrece una hoja de ruta para la “preparación algorítmica”: un conjunto de cinco principios con visión de futuro para guiar el desarrollo de normativas que afronten la perspectiva de cisnes negros algorítmicos y mitiguen los daños que plantean a la sociedad».

III. EVOLUCIÓN, TRAMITACIÓN Y CONTENIDO FINAL DEL ARTÍCULO 10

El RIA dedica el artículo 10, dentro de la Sección 2 del capítulo III dedicado a los Sistemas de IA alto riesgo, a los *Datos y gobernanza de datos*, consciente de la vital trascendencia que, los datos y la gobernanza de los mismos, tienen dentro de un sistema de IA. Para abordar el análisis de su contenido, analizaremos primero los roles involucrados, para adentrarnos después en las obligaciones asociadas a cada uno de ellos.

1. ROLES INVOLUCRADOS

Debe ponerse de relieve que las obligaciones establecidas en el artículo 10 relativas a los datos y su gobernanza se enuncian sin mencionar al concreto sujeto obligado a su cumplimiento, ya que se configuran como requisitos del propio sistema de alto riesgo.

Por ello, para poder establecer los sujetos obligados a implementar prácticas adecuadas de gobernanza y gestión de datos, debemos atender primero a los conjuntos de datos sobre los que recaen dichas obligaciones, para ver a qué integrante de la cadena de valor corresponde. Así, el artículo 10 distingue entre los conjuntos de datos utilizados para el entrenamiento, validación y prueba de sistemas de IA de alto riesgo, distintos de aquellos otros que no utilizan técnicas que implican el

27. KOLT, N., *Algorithmic Black Swans* (octubre, 2023). Washington University Law Review, Vol. 101, Forthcoming, disponible SSRN: <https://ssrn.com/abstract=4370566> p.42. Estos principios son: Principio 1: La gobernanza de la IA debe tratar de anticipar y mitigar los daños a gran escala de los sistemas de IA; Principio 2: La gobernanza de la IA debe adoptar un enfoque de cartera compuesto por estrategias reguladoras diversas y no correlacionadas; Principio 3: La gobernanza de la IA debe ser altamente escalable; Principio 4: La gobernanza de la IA debe explorar y evaluar continuamente nuevas estrategias reguladoras; Principio 5: El análisis coste-beneficio de las intervenciones de la gobernanza de la IA debe tener más en cuenta los peores resultados.

entrenamiento de modelos²⁸. Si atendemos a las diversas figuras que conforman la cadena de valor, ya analizadas a lo largo de la presente obra, dejando a un lado aquellos roles que no influyen directamente en el desarrollo del sistema de IA, tales como el distribuidor²⁹ o el importador³⁰, destaca el proveedor³¹. El proveedor es aquella entidad que *desarrolla o para la que se desarrolla* un sistema de IA o un modelo de IA de uso general y lo introduce en el mercado o pone en servicio con su propio nombre o marca comercial, por lo que deberá disponer de conjuntos de datos utilizados para el entrenamiento, validación y prueba del sistema. No obstante, el proveedor, por definición, puede desarrollar directamente el sistema de IA o contratar a terceros para realizar dicho desarrollo, en cuyo caso, deberán regularse contractualmente las correspondientes obligaciones. En este sentido, el Parlamento Europeo recordaba en sus Considerandos³² que los *desarrolladores* de algoritmos tienen especial relevancia ya que han podido utilizar datos subyacentes (históricos) los cuales pueden no reunir los requisitos de calidad deseables por incluir sesgos, o bien, haber generado estos datos en entornos reales y, por tanto, estar sesgados por defecto. Finalmente, se ha omitido la referencia expresa a los desarrolladores en dicho Considerando, pero manteniendo la importancia de la calidad de los datos subyacentes.

En relación a la figura del *proveedor*, debe tenerse presente que, en materia de responsabilidades a lo largo de la cadena de valor de la IA, el Reglamento en determinados casos considera³³ «proveedor» de un sistema de IA de alto riesgo a cualquier distribuidor, importador, implementador u otro tercero y, por tanto, sujeto a las obligaciones establecidas en el artículo 16 para proveedores e implementadores de sistemas de IA de alto riesgo y de otras partes. En este sentido, destacar que el Parlamento propuso³⁴ modificar el título del artículo de 16 de forma que incluyera no sólo a los proveedores, sino también a los implementadores y otras partes, pero dicha enmienda no fue finalmente aceptada. No obstante, a pesar de la evidente coherencia de la enmienda propuesta por el Parlamento, ello no afecta al fondo, ya que el artículo 25.1 establece expresamente la responsabilidad de estas figuras. Por tanto, cualquier distribuidor, importador, implementador u otro tercero que (i) pongan su nombre o marca en un sistema de IA de alto riesgo ya comercializado o puesto en servicio (ii) realice una modificación sustancial en el mismo o (iii) realice una modificación de tal manera que el sistema de IA se convierta en un sistema de IA de alto riesgo, quedará sujeto a las obligaciones establecidas en el artículo 16 y, por tanto, al cumplimiento de todos los requisitos exigidos para los sistemas de alto riesgo, entre los que se encuentra los relativos a la gobernanza de datos. La versión final³⁵ ha incluido la definición de «proveedor intermedio», definido como aquel proveedor de un sistema de IA, incluido un sistema de IA de propósito general, que

28. Artículo 10.6.

29. Artículo 3. 7).

30. Artículo 3. 6).

31. Artículo 3. 3).

32. Enmienda 78 sobre el Considerando 44 (actual Considerando 67).

33. Artículo 25.1.

34. Enmienda 331, título del artículo 16: «Obligaciones de los proveedores e implementadores de sistemas de IA de alto riesgo y de otras partes».

35. Artículo 3, 68).

integra un modelo de IA, independientemente de si el modelo es proporcionado por ellos mismos e integrado verticalmente o proporcionado por otra entidad basada en relaciones contractuales.

Por su parte, el implementador o responsable del despliegue (*deployer*)³⁶ es aquella entidad que utiliza un sistema de IA bajo su propia autoridad, siempre que no aplique la «excepción doméstica», es decir, que su uso se enmarque en una actividad personal de carácter no profesional. Aunque no se mencione expresamente, entendemos que el implementador será sujeto obligado siempre que reentrene el sistema facilitado por el proveedor. Ahondaremos en la cuestión del reentrenamiento más adelante.

El artículo 10 únicamente menciona expresamente al proveedor en relación a la posibilidad de tratar excepcionalmente categorías especiales de datos en la medida en que sea estrictamente necesario para garantizar la detección y la corrección de los sesgos negativos. El Parlamento añadió³⁷ una segunda referencia expresa al proveedor, al configurar su posible exención de responsabilidad por infracción de cualquiera de las obligaciones establecidas en el artículo 10, trasladando dicha responsabilidad al implementador, en el caso de que el proveedor no tuviera acceso a los datos, por estar en poder exclusivamente del implementador y así se hubiera establecido en un contrato. Este apartado no ha sido acogido en la versión final, pero cabe preguntarse qué sentido tendría que un implementador tuviera acceso exclusivo a los datos de un sistema introducido en el mercado por un proveedor, pero sin que el implementador lo utilice bajo su propia autoridad, pues en ese caso ya tendría responsabilidad sobre el mismo.

En todo caso, resulta de especial interés la mención³⁸ realizada por el Parlamento expresamente en relación a la posibilidad de externalizar los requisitos relacionados con la gobernanza de datos «recurriendo a terceros que ofrezcan servicios certificados de cumplimiento, incluida la verificación de la gobernanza de los datos, la integridad de los conjuntos de datos y las prácticas de entrenamiento, validación y prueba de datos», la cual ha sido acogida en el texto definitivo. Por ello, pensamos que irrumpirá en la cadena de valor una nueva figura («verificadores de datos» o «proveedores de servicios certificados de datos») que se encargue precisamente de proporcionar a proveedores o implementadores, conjuntos de datos para el desarrollo de los sistemas de IA, que cumplan con los requisitos establecidos por el RIA.

2. OBLIGACIONES

El artículo 10 relativo a los datos y su gobernanza es de una importancia capital³⁹ ya que, del cumplimiento de las obligaciones establecidas en el mismo, deriva

36. Artículo 3. 4). Destacar el relevante cambio introducido por el Parlamento Europeo (a través de la Enmienda 172 que modifica la definición de usuario del artículo 3, 4) en coherencia con el Considerando 59) que prescinde del término «usuario» para denominarlo «implementador», lo cual entendemos resulta más clarificador ya que descarta confusiones con el usuario final del sistema, persona física.

37. Enmienda 291 que introducía un nuevo apartado 6 bis.

38. Enmienda 78 que modifica Considerando 44 *in fine* (actual Considerando 67).

39. El Considerando 67 afirma (traducción no oficial) «Los datos de alta calidad y el acceso a datos de alta calidad desempeñan un papel esencial a la hora de proporcionar una estructura y garantizar el funcionamiento de muchos sistemas de IA, en especial

disponer de datos de alta calidad y, por tanto, el correcto funcionamiento de los sistemas de IA, especialmente, los de alto riesgo.

Así, los sistemas de IA de alto riesgo que hagan uso de técnicas que impliquen el entrenamiento de modelos con datos, deberán desarrollarse a partir de conjuntos de datos que cumplan con los *criterios de calidad* que se especifican en los párrafos 2 a 5. Frente a la exigencia de utilizar conjuntos de datos de entrenamiento, validación y prueba que cumplan los criterios de calidad indicados, debe resaltarse la modulación que propuso el Parlamento Europeo⁴⁰ de que dichos criterios de calidad se cumplieran «*en la medida en que esto sea técnicamente posible de conformidad con el segmento de mercado o ámbito de aplicación de que se trate*». Puntualizaba también el Parlamento, que estos criterios debían cumplirse en el caso de las técnicas que no requieran datos de entrada etiquetados, como el aprendizaje no supervisado y el aprendizaje de refuerzo. Ninguna de estas dos propuestas del Parlamento fue finalmente acogida, por lo que la versión final ha eliminado cualquier suerte de modulación de responsabilidad. Analizaremos a continuación los criterios de calidad establecidos en cada uno de los párrafos.

En el párrafo segundo se concretan las *prácticas adecuadas* de gobernanza y gestión de datos a las que deberán someterse los conjuntos de datos de entrenamiento, validación y prueba que se utilicen para el entrenamiento de modelos de sistemas de IA de alto riesgo. Podemos clasificar dichas prácticas en torno a diferentes acciones:

- la elección de un diseño adecuado: a) decisiones pertinentes relativas al diseño;
- su recopilación: b) procesos de recogida de datos y el origen de los mismos y, en el caso de los datos personales, la finalidad original de la recogida de datos;
- la preparación de los datos: c) operaciones de tratamiento oportunas para la preparación de los datos, como la anotación, el etiquetado, la depuración, la actualización, el enriquecimiento y la agregación;
- d) la formulación de hipótesis, en particular en lo que respecta a la información que los datos miden y representan;
- al estudio previo de los conjuntos de datos: e) evaluación de la disponibilidad, la cantidad e idoneidad de los conjuntos de datos necesarios);
- la calidad de los datos:
- f) examen atendiendo a posibles sesgos que puedan afectar a la salud y la seguridad de las personas, afectar negativamente a los derechos fundamentales o dar lugar a algún tipo de discriminación prohibida por el Derecho de la Unión, especialmente cuando las salidas de datos influyan en las informaciones de entrada de futuras operaciones;
- g) medidas adecuadas para detectar, prevenir y mitigar los posibles sesgos detectados con arreglo a la letra f); y

cuando se emplean técnicas que implican el entrenamiento de modelos, con vistas a garantizar que el sistema de IA de alto riesgo funcione del modo previsto y en condiciones de seguridad y no se convierta en una fuente de algún tipo de discriminación prohibida por el Derecho de la Unión (...).

40. Enmienda 278, que modifica el artículo 10.1.

— h) la identificación de las lagunas o deficiencias de datos pertinentes que impidan el cumplimiento del presente Reglamento, y la forma de subsanar dichas lagunas y deficiencias.

En primer lugar, destacar que, a diferencia de las propuestas de la Comisión y del Consejo que hablaban de «*prácticas* adecuadas de gobernanza y gestión de datos», el Parlamento⁴¹ propuso sustituir dicha expresión por «una *gobernanza adecuada* al contexto del uso, así como a la finalidad prevista del sistema de IA» que implicaba la adopción de una serie de *medidas*. La versión final no acoge la propuesta del Parlamento y vuelve a hablar de «*prácticas* adecuadas de gobernanza y gestión de datos apropiadas para la finalidad prevista del sistema de IA».

En relación a las concretas prácticas, la mayoría de las enmiendas introducidas por el Parlamento han sido finalmente acogidas. Así, el Parlamento incluyó⁴² una nueva práctica relativa a la transparencia en relación a la finalidad original de la recopilación de datos, que el texto final⁴³ concreta aún más al distinguir entre procesos de recogida de datos no personales, en cuyo caso deberá indicarse su origen y, de datos personales, sobre los que debe indicarse la finalidad original de su recogida.

En la práctica relativa a las operaciones de preparación⁴⁴ de los datos, el Parlamento añadió la actualización⁴⁵ de los mismos.

Respecto al estudio previo de los conjuntos de datos⁴⁶, el Parlamento⁴⁷ eliminó el requerimiento de que la evaluación de la disponibilidad, la cantidad e idoneidad de los conjuntos de datos necesarios fuese previa, lo cual, en nuestra opinión, no tiene demasiado sentido, pues, aunque dicha evaluación es obvio que deberá ser previa, reforzaba su carácter *ad hoc*.

Pero es en las medidas relativas a la calidad de los datos donde el Parlamento introdujo mayores modificaciones. Así, en cuanto al examen de posibles sesgos⁴⁸, el Consejo añadió la precisión de que pudieran «afectar a la salud y la seguridad de las personas físicas o dar lugar a algún tipo de discriminación prohibida por el Derecho de la Unión». Por su parte, el Parlamento añadió⁴⁹ que pudieran «afectar negativamente a los derechos fundamentales». En relación a que puedan dar lugar a una discriminación prohibida por el Derecho de la Unión, el Parlamento añadió «especialmente cuando los datos de salida influyan en los datos de entrada en futuras operaciones (“bucle de retroalimentación”), puntualización que finalmente no ha sido incluida en el articulado», pero sí en el Considerando 67. Asimismo, el Parlamento introdujo una nueva práctica⁵⁰, consistente en llevar a cabo las «medidas adecuadas para detectar, prevenir y mitigar posibles sesgos», lo que implica una labor más allá de realizar un examen *ex ante* sobre conjuntos de datos determinados, obligando

41. Enmienda 279, que modifica el artículo 10.2.

42. Enmienda 280 que incluye un nuevo apartado a bis).

43. Artículo 10.2, apartado b).

44. Artículo 10.2 c).

45. Enmienda 282.

46. Artículo 10.2 e).

47. Enmienda 284.

48. Artículo 10.2 f).

49. Enmienda 285.

50. Enmienda 286 que introdujo un nuevo apartado f bis), actual apartado g).

a establecer medidas para poder detectar, prevenir y mitigar posibles sesgos que pudieran detectarse o manifestarse posteriormente.

En relación a las prácticas relativas a la detección de posibles lagunas o deficiencias en los datos y la forma de subsanarlas⁵¹, el Parlamento introdujo⁵², y así lo recoge la versión final, el matiz de que dichas lagunas o deficiencias serán las «pertinentes que impidan el cumplimiento del presente Reglamento», pareciendo reducir, por tanto, el ámbito objetivo de las mencionadas lagunas o deficiencias subsanables.

En el párrafo tercero se concretan una serie de *obligaciones que* comenzaron siendo de resultado, pero que finalmente han acabado siendo moduladas. Así, se establece que los conjuntos de datos que se utilicen para el entrenamiento, validación y prueba «serán pertinentes, suficientemente representativos y, en la medida de lo posible, estarán exentos de errores y serán completos con vistas a la finalidad prevista».

En relación a esta primera obligación, el Consejo introdujo una primera modulación al incluir «en la mayor medida posible» de forma previa al imperativo («carecerán de errores y estarán completos»). Posteriormente, el Parlamento⁵³ modifica significativamente la redacción y a la obligación de que los datos sean representativos, añade el adverbio «suficientemente» representativos.

En segundo lugar, la obligación de resultado consistente en carecer de errores y estar completos, se transforma por el Parlamento en que serán «debidamente evaluados por lo que respecta a los errores y ser tan completos como sea posible en vista de la finalidad prevista»⁵⁴. El texto final de forma similar indica «en la medida de lo posible, estar exentos de errores y ser completos con vistas a la finalidad prevista». Dicha evolución también se aprecia de manera correlativa en el Considerando 67.

Finalmente, se añade la obligación de que los conjuntos de datos tendrán las propiedades estadísticas adecuadas, en relación a las personas o los grupos de personas sobre los que se pretenda utilizar el sistema de IA de alto riesgo. Los conjuntos de datos *podrán reunir* estas características individualmente para cada dato o para una combinación de estos. El Parlamento corrige que los conjuntos de datos *reunirán* estas características, no individualmente para cada dato, sino para cada conjunto de datos o para una combinación de estos, tal y como reza el texto final.

En el párrafo cuarto se establece que los datos «tendrán en cuenta, en la medida en que lo exija la finalidad prevista, las características o los elementos propios del entorno geográfico, contextual, conductual o funcional específico en el que esté previsto utilizar el sistema de IA de alto riesgo». El Parlamento propuso añadir⁵⁵ que también deberán tenerse en cuenta los usos indebidos razonablemente previsibles del sistema de IA, que no será recogida por la versión final. Por otro lado, esta obligación debe conectarse con la presunción establecida en el artículo 42 a través de la cual se presumirá que se cumplen los requisitos del párrafo cuarto siempre que los sistemas

51. Artículo 10.2 h).

52. Enmienda 287.

53. Enmienda 288.

54. En coherencia con el Considerando 44 que establece que «(...) deben ser lo suficientemente pertinentes y representativos, adecuadamente examinados en busca de errores y tan completos como sea posible en vista de la finalidad prevista del sistema (...)».

55. Enmienda 289.

de IA de alto riesgo hayan sido entrenados y probados con datos que reflejen el entorno geográfico, conductual, contextual y funcional específico en el que esté previsto su uso.

En el párrafo quinto se establece la posibilidad para los proveedores de tratar categorías especiales de datos «en la medida en que sea estrictamente necesario para garantizar la detección y corrección de los sesgos». Indicar que el Parlamento denominaba a dichos sesgos *negativos* y los definía⁵⁶ como aquellos que «crea(n) un efecto discriminatorio directo o indirecto contra una persona física», pero finalmente, el concepto de «sesgo negativo» no se recogió en la versión final. Se establece la posibilidad de establecer una base de legitimación⁵⁷ adecuada para poder tratar categorías especiales de datos que, en aplicación de la normativa de protección de datos, no eximirá de la obligación de adoptar las salvaguardias adecuadas para los derechos y libertades de las personas físicas. El Considerando 70 habla expresamente de «cuestión de interés público esencial» y rescata la referencia expresa al artículo 9, apartado 2, letra g) del Reglamento (UE) 2016/679 y del artículo 10, apartado 2, letra g) del Reglamento (UE) 2018/1725, que el Consejo había introducido en un primer momento. En este punto, para que el tratamiento de datos pueda ampararse en el artículo 9.2 g) del RGPD (el tratamiento es necesario por razones de un interés público esencial), debe recordarse que es necesario que así se prevea en una norma de Derecho nacional o europeo, que además especifique el interés público esencial que justifica el tratamiento de dichos datos, en qué circunstancias puede limitarse el derecho de protección de datos, unas reglas precisas y las garantías adecuadas tanto a nivel técnico como organizativo para proteger los intereses y derechos fundamentales del interesado. En este punto, el Parlamento, en lugar de enumerar a modo de ejemplo una serie de medidas, introdujo⁵⁸ un catálogo de condiciones necesarias que deberán aplicarse para que se pueda realizar el tratamiento, entre las que se incluye, que el tratamiento de datos sintéticos o anonimizados no permita alcanzar eficazmente la detección y corrección de sesgos; que los datos que se utilicen sean seudonimizados o se sujeten a limitaciones técnicas en cuanto a la reutilización de los datos personales y a las medidas más avanzadas de seguridad y de preservación de la intimidad; o que se eliminen una vez se ha corregido el sesgo o cuando los datos personales lleguen al final de su período de conservación, las cuales han sido recogidas en el texto final.

El Parlamento europeo remarca la excepcionalidad de que los proveedores de dichos sistemas puedan tratar categorías especiales de datos al introducir el adverbio «excepcionalmente». En este sentido, el Parlamento introdujo la exigencia de que los proveedores que recurrieran a esta disposición debían elaborar documentación que explicase por qué el tratamiento de categorías especiales de datos personales era necesario para detectar y corregir sesgos. En la versión final esta obligación no menciona expresamente a los proveedores y se limita a indicar que los registros de las actividades de tratamiento de conformidad con el Reglamento (UE) 2016/679,

56. Enmienda 78 al Considerando 44 *in fine*.

57. Enmienda 160 por la que se introduce un nuevo artículo 2.5 bis: «El presente Reglamento no afectará a los Reglamentos (UE) 2016/679 (...), sin perjuicio de los mecanismos previstos en el artículo 10, apartado 5 (...)» que finalmente recoge el texto final.

58. Enmienda 290.

la Directiva (UE) 2016/680 y el Reglamento (UE) 2018/1725 deberán incluir dicha justificación.

Una vez vistos los criterios de calidad establecidos para los conjuntos de datos de entrenamiento, validación y prueba que sirvan para el desarrollo de sistemas de IA de alto riesgo que utilicen técnicas que implican el entrenamiento de modelos con datos, el párrafo sexto establece que dichos criterios de calidad, con respecto al desarrollo de sistemas de IA de alto riesgo que no empleen técnicas que impliquen el entrenamiento de modelos, solo se aplicarán a los conjuntos de datos de prueba. Resulta interesante destacar que, mientras la Comisión establecía para estos sistemas que debían garantizar el cumplimiento de las prácticas adecuadas de gobernanza y gestión de datos establecidas en el párrafo segundo, el Consejo, el Parlamento y la versión final amplían dicha obligación a todos los criterios de calidad (párrafos 2 a 5) pero limitando dicho cumplimiento solo a los conjuntos de datos de prueba.

IV. APROXIMACIÓN CRÍTICA

Vista la evolución de la propuesta normativa del artículo 10, en el presente apartado realizaremos una evaluación crítica del contenido final de dicho artículo.

En primer lugar, en relación a los roles, una crítica general es la ausencia de definición de usuario final o «afectado» por el sistema de IA, sobre todo si tenemos en cuenta que el Parlamento propuso introducir una definición⁵⁹ de «persona afectada» y que éstas entran dentro del ámbito de aplicación del RIA⁶⁰. En relación a la cadena de valor de la IA y los roles que intervienen en la misma, tras la importante referencia⁶¹ introducida por el Parlamento en relación a la posibilidad de externalizar los requisitos relacionados con la gobernanza de datos, tal y como comentábamos en el apartado dedicado a los roles, entendemos que se dan todas las circunstancias para la irrupción en la cadena de valor de nuevas figuras que exclusivamente provean de datos «verificados» y certifiquen que los datos cumplen con los requisitos de gobernanza establecidos, así como su integridad y entrenamiento («verificadores de datos» o «proveedores de servicios certificados de datos»). Resultará, por tanto, clave la regulación y posible transmisión de responsabilidad. No obstante, cabe cuestionarse si este modelo de provisión de datos «verificados» puede ofrecer el cumplimiento individualizado de dichos requisitos de gobernanza al que aspira el Reglamento ya que, en primer lugar, la gobernanza debe adecuarse al contexto del uso, así como a la finalidad prevista del sistema de IA⁶² y, en segundo lugar, los

59. Enmienda 174. «Persona afectada: toda persona física o grupo de personas que se vean expuestos a un sistema de IA o resulten afectados de algún otro modo por un sistema de IA». No se entienden las razones que hayan podido conducir a la no aceptación de dicha enmienda, sobre todo cuando sí se introduce la referencia a los mecanismos de garantía o tutela en caso de infracción del Reglamento, los cuales, por otro lado, no se reservan al afectado por el sistema de IA, sino a cualquier persona que considere que ha habido una infracción del Reglamento; y, finalmente, porque sí aporta otras definiciones que no parecen tan relevantes, como la del «sujeto» que participa en pruebas en condiciones reales o el «consentimiento informado» de dicha persona.

60. Artículo 2.1 g).

61. Enmienda 78 que modifica Considerando 44 *in fine*.

62. Artículo 10, párrafo segundo.

conjuntos de datos deberán tener en cuenta, en la medida en que lo exija la finalidad prevista, las características o elementos propios del entorno geográfico, contextual conductual o funcional en el que está previsto utilizar el sistema de IA de alto riesgo⁶³. De este modo, tal y como advierte PEGUERA POCH⁶⁴, la cadena de valor podría adquirir «configuraciones diversas a las consideradas por el legislador en función de la evolución de los modelos de negocio que se acaben consolidando».

Por otro lado, tal y como el Supervisor Europeo de Protección de Datos (SEPD) recomendaba⁶⁵, debería especificarse que los operadores de IA que reentrenen sistemas de IA pre-entrenados queden incluidos dentro del concepto de proveedores, ya que los sistemas de IA pueden entrenarse más de una vez a lo largo de su ciclo de vida o bien pueden aplicar técnicas de aprendizaje continuo. El reentrenamiento puede deberse, indica el SEPD, bien por la falta de grandes conjuntos de datos para entrenamiento, o bien porque se vuelven a entrenar con el fin de utilizarlos para una tarea similar en un ámbito diferente (aprendizaje por transferencia). El RIA tampoco aclara si las actividades de reentrenamiento o de aprendizaje continuo se consideran parte del «desarrollo» del sistema de IA, pues en tal caso claramente se considerarían proveedores. Indica el SEPD que este punto es especialmente relevante en relación a los modelos fundacionales y la posibilidad generalizada de reentrenamiento de los mismos. El Reglamento no incluye una definición de las operaciones que se incluyen en «desarrollo» de un sistema de IA, y en la definición de proveedor, si bien incluye la referencia al desarrollo o la comercialización bajo su nombre o marca de un modelo de IA de propósito general, no se menciona el reentrenamiento. No obstante, en un único Considerando⁶⁶ se menciona expresamente al reentrenamiento como un proceso que puede ser incorporado por el proveedor al sistema de IA. Por ello, una interpretación sistemática y teleológica nos llevaría a considerar proveedor a aquel que introduzca en el mercado un sistema reentrenado, aunque la precisión realizada por el SEPD hubiera sido lo conveniente.

En segundo lugar, el artículo 10 se refiere a los requisitos que deben cumplir los conjuntos de datos de entrenamiento, validación y prueba que se utilizarán para el desarrollo de sistemas de IA de alto riesgo que utilizan técnicas que implican el entrenamiento de modelos con datos. Se ha puesto de relieve por determinados autores⁶⁷ que se ignoran otras etapas del *machine learning* que también deberían estar

63. Artículo 10, párrafo cuarto.

64. PEGUERA POCH, M. «La propuesta de reglamento de IA: una intervención legislativa insoslayable en un contexto de incertidumbre», en PEGUERA POCH (coords.) *Perspectivas regulatorias de la Inteligencia Artificial en la Unión Europea*, Madrid: Reus, 2023.

65. EDPS, *Opinion 44/2023 on the Proposal for Artificial Intelligence Act in the light of legislative developments*, p. 8. https://edps.europa.eu/system/files/2023-10/2023-0137_d3269_opinion_en.pdf

66. Considerando 88: «Dentro de la cadena de valor de la IA, múltiples partes suministran a menudo sistemas, herramientas y servicios de IA, pero también componentes o procesos que son incorporados por el proveedor al sistema de IA con diversos objetivos, incluido el entrenamiento del modelo, el reentrenamiento del modelo, la prueba y evaluación del modelo, la integración en programas informáticos u otros aspectos del desarrollo del modelo (...)».

67. EBERS, M., HOCH, V. R. S., ROSENKRANZ, F., RUSCHEMEIER, H., & STEINRÖTTER, B. «The european Commission's proposal for an artificial intelligence Act-A

sujetas a criterios de calidad de datos y a prácticas de gobernanza de datos y también con respecto a las licencias de datos que permiten el acceso a los mismos.

El párrafo primero establece una suerte de obligación de resultado, al establecer que los sistemas de IA de alto riesgo que hagan uso de técnicas que impliquen el entrenamiento de modelos con datos «se desarrollarán sobre la base de conjuntos de datos de entrenamiento, validación y prueba que cumplan los criterios de calidad establecidos en los párrafos 2 a 5 (...)». El Parlamento propuso introducir una modulación de responsabilidad, o más bien, la eliminación de dicha obligación de resultado con respecto a todas las obligaciones en materia de gobernanza, consistente en condicionar la obligatoriedad de cumplimiento de dichos requisitos a que fuera «técnicamente posible de conformidad con el segmento de mercado o ámbito de aplicación de que se trate». Dicha modulación era una modificación significativa, pero que, en la práctica, podría no ser tal, pues al basarse en criterios exclusivamente técnicos, la justificación de la imposibilidad de cumplir con algunos de los criterios de calidad requeridos debería evidenciar, precisamente, la imposibilidad técnica que aconteciera en cada caso concreto. Lo relevante es que la versión final ha eliminado cualquier suerte de modulación de responsabilidad, con independencia del segmento o ámbito de aplicación concreto o de la imposibilidad técnica, lo cual refuerza la importancia de cumplir con los criterios de calidad en todo caso.

El párrafo segundo introduce las prácticas de gobernanza y gestión que deberán cumplir los conjuntos de datos para el entrenamiento, validación y prueba de los sistemas de alto riesgo, que suponen todo un sistema de gestión de datos. Dichas prácticas de gobernanza de datos deben conectarse necesariamente con el sistema de gestión de la calidad y, en especial, con el sistema de gestión de riesgos, aunque no se indique expresamente, lo cual hubiera sido deseable, ya que remarcaría la importancia del cumplimiento del artículo 10, que, como decimos, es esencial. El sistema de gestión de la calidad sí que menciona⁶⁸ «los sistemas y procedimientos de gestión de datos lo que incluye su adquisición, recopilación, análisis, etiquetado, almacenamiento, filtrado, prospección, agregación, conservación y cualquier otra operación relacionada con los datos que se lleve a cabo antes de la introducción en el mercado o puesta en servicio de sistemas de IA de alto riesgo», pero entendemos que hubiera sido deseable hacer referencia expresa al sistema de gobernanza de datos completo establecido en el artículo 10, de la misma forma que se incluye la referencia expresa al sistema de gestión de riesgos. En relación al sistema de gestión de riesgos establecido en el artículo 9, se establece que «Los riesgos a que se refiere el presente artículo son únicamente aquellos que pueden reducirse o eliminarse razonablemente mediante el desarrollo o el diseño del sistema de IA de alto riesgo o el suministro de información técnica adecuada», lo cual parece excluir los riesgos derivados del incumplimiento de los criterios de calidad de los datos. No obstante, el mismo artículo precisa que se deberán determinar y analizar los riesgos conocidos y previsibles que el sistema de IA de alto riesgo pueda conllevar para la salud, la seguridad o los derechos fundamentales, lo cual implica que no podrán desconocerse los riesgos derivados del incumplimiento de los criterios de calidad de los datos y

critical assessment by members of the robotics and AI law society (RIALS)», 2021, J, 4(4), p. 595. doi: <https://doi.org/10.3390/j4040043>
68. Artículo 17.1 f).

las prácticas de gobernanza. En cualquier caso, el sistema de gobernanza de datos configurado por el RIA tiene entidad propia suficiente que trasciende al sistema de gestión de riesgos, pero ello no implica que sea desconocido por este último.

La importancia del sistema de gobernanza de datos, queda evidenciada por el hecho de formar parte de la documentación técnica (Anexo IV) que deberá conservar el proveedor durante diez años, si bien es cierto que no menciona expresamente el artículo 10, sino que se refiere, en relación a los datos, a una descripción general de los conjuntos de datos de entrenamiento utilizados e información acerca de su procedencia, su alcance y sus características principales; la manera en que se obtuvieron y seleccionaron los datos; los procedimientos de etiquetado (p. ej., para el aprendizaje supervisado) y las metodologías de depuración de datos (p. ej., la detección de anomalías); y los procedimientos de validación y prueba utilizados, incluida la información acerca de los datos de validación y prueba empleados y sus características principales. Hubiera sido deseable incluir una referencia expresa a los procedimientos de gobernanza de datos del artículo 10 para poder disponer de la trazabilidad suficiente a efectos de posibles exigencias de responsabilidades.

En la propuesta de la Comisión, el párrafo tercero establecía una obligación de resultado consistente en que los conjuntos de datos que se utilizasen para el entrenamiento, validación y prueba «serán pertinentes y representativos, carecerán de errores y estarán completos». La industria o incluso algunos gobiernos como el Noruego⁶⁹ y algunos autores, se mostraron reacios a su redacción como un «requerimiento absoluto», ya que es una tarea imposible que los datos puedan estar siempre libres de errores siendo tal nivel de perfección «técnicamente inviable» y podrían obstaculizar la innovación⁷⁰. Otros autores⁷¹ han puesto de relieve la existencia de condicionamientos frente al cumplimiento de estas obligaciones aparentemente estrictas, que rebajan de hecho el grado de exigencia. Así, las sucesivas versiones han introducido fórmulas que han rebajado el grado de exigencia en la obtención de dichos resultados, de forma que la versión final establece que los datos deberán ser pertinentes, suficientemente representativos *y en la medida de lo posible*, estarán exentos de errores y serán completos, teniendo en cuenta la finalidad prevista. Hubiera sido aconsejable introducir también, junto con la finalidad, la referencia a los usos indebidos razonablemente previsibles⁷², por una cuestión de coherencia ya que éstos sí son tenidos en cuenta en la evaluación de riesgos del artículo 9⁷³.

Esta aparente modulación de responsabilidad, consideramos que debe conectarse con el concepto de responsabilidad proactiva, por lo que se deberá poder acreditar la pertinencia, representatividad suficiente, análisis de posibles errores y la completitud

69. <https://www.regjeringen.no/contentassets/939c260c81234eae96b6a1a0fd32b6de/norwegian-position-paper-on-the-ecs-proposal-for-a-regulation-of-ai.pdf>

70. Cit. Ebers, M. *et alia*.

71. VEALE M. y BORGESIU F., «Demystifying the Draft EU Artificial Intelligence Act», *Computer Law Review International*, 2021, 22(4), pp. 97-112, párrafo 41. DOI <https://doi.org/10.48550/arXiv.2107.03721>

72. Artículo 3. 13).

73. Artículo 9.2b).

de los datos, si bien es cierto que por la propia naturaleza de la IA pueda resultar problemático evaluar la responsabilidad de los resultados obtenidos⁷⁴.

Por otro lado, entendemos que resulta un tanto paradójico hablar de «criterios de calidad» cuando no se concretan criterios para medir la calidad de los conjuntos de datos, refiriéndose únicamente al resultado deseable⁷⁵. En otras palabras, el RIA deja dicha concreción al ámbito de la normalización, lo cual resulta en cierto modo comprensible al tratarse de aspectos mayormente técnicos, pero a la vez deja a la norma un tanto vacía de contenido sustantivo⁷⁶. Así, se indica⁷⁷ que «la normalización debe desempeñar un papel fundamental para proporcionar soluciones técnicas a los proveedores a fin de garantizar el cumplimiento del presente Reglamento (...)». Debe destacarse en este punto las modificaciones⁷⁸ realizadas por el Parlamento, que suponen un rol activo por parte de la Comisión y no una mera «externalización» de la cuestión a organismos de normalización. Así, la Comisión, teniendo en cuenta la importancia de las normas para garantizar la conformidad con los requisitos del Reglamento y la competitividad de las empresas, establece que en la elaboración de normas deberá haber una representación equilibrada de los intereses fomentando la participación de todas las partes interesadas pertinentes. Para facilitar el cumplimiento normativo, en el plazo máximo de dos meses desde la aprobación del RIA, la Comisión deberá emitir las primeras peticiones de normalización a las organizaciones europeas de normalización⁷⁹.

En este punto, debe ponerse de manifiesto que el recurso a organismos privados para la elaboración de normas es criticado por determinados autores⁸⁰, máxime cuando dichas normas aparentemente «técnicas» tienen repercusión en valores o derechos fundamentales. Lo anterior se evidencia cuando el RIA⁸¹ establece que la Comisión estará facultada para adoptar especificaciones comunes cuando las normas armonizadas pertinentes no aborden suficientemente los problemas relacionados con los derechos fundamentales. Recordemos que aquellos sistemas de IA de alto riesgo o los modelos de IA de propósito general que sean conformes con las normas armonizadas que se aprueben, se *presumirá*⁸² que cumplen con los

74. Op. Cit. NOVELLI C., TADDEO M., FLORIDI L., Accountability in artificial intelligence.

75. *Cit.* Ebers, M. *et alia* mencionan como posibles criterios la precisión predictiva, robustez y la imparcialidad de los modelos de aprendizaje automático entrenados.

76. En la Propuesta de cláusulas contractuales tipo para la contratación de inteligencia artificial por parte de organismos públicos, versión de septiembre de 2023, el artículo 3 (características de los conjuntos de datos) es exactamente igual tanto para los sistemas de IA de alto riesgo como para el resto de sistemas. Disponible en <https://public-buyers-community.ec.europa.eu/communities/procurement-ai/resources/eu-model-contractual-ai-clauses-pilot-procurements-ai>

77. Considerando 121.

78. Enmiendas 103 a 107, relativas al Considerando 61 (actual Considerando 121).

79. CEN (European Committee for Standardisation), CENELEC (European Committee for Electrotechnical Standardisation) <https://www.cencenelec.eu/>

80. VEALE M., y BORGESIU F., «Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach», *Computer Law Review International*, vol. 22, núm 4, pág 105.

81. Artículo 41.1 a (iii).

82. Artículo 40.1.

requisitos⁸³ establecidos para los sistemas de IA de alto riesgo, por lo que para obtener la evaluación de la conformidad bastará seguir un procedimiento basado en el control interno (Anexo VI), que no prevé la participación de un organismo notificado. Por tanto, solo cuando no existan normas armonizadas o especificaciones comunes, o no se hayan aplicado, se seguirá un procedimiento de evaluación de la conformidad que implique la intervención de un organismo notificado (Anexo VII). Serán los proveedores de estos sistemas, antes de su introducción en el mercado o puesta en servicio, quienes se asegurarán de que han sido sometidos al procedimiento de evaluación de la conformidad que corresponda⁸⁴ y, de ser positiva, elaborarán la declaración UE de conformidad⁸⁵ y colocarán el marcado CE⁸⁶. Huelga decir que la «autoevaluación» de la conformidad supone a la postre menos garantías precisamente en lo que respecta a la verificación del cumplimiento de los requisitos, recordemos, para sistemas de IA de alto riesgo, por lo que sería deseable que el previo procedimiento de evaluación de la conformidad para sistemas de IA de alto riesgo se realizase siempre por una tercera parte diferente al proveedor. Este punto también ha sido reclamado tanto por el Comité Europeo de Protección de Datos (CEPD) como por el Supervisor Europeo de Protección de Datos (SEPD), los cuales afirman⁸⁷ que, aunque en el RGPD no se establece la obligación de realizar una evaluación de la conformidad por terceros para tratamientos de datos de alto riesgo, aún no se conocen plenamente los riesgos en el ámbito de la IA. Es por ello que abogan por introducir de forma general, y no sólo para determinados sistemas de alto riesgo, la evaluación de la conformidad *ex ante* por terceros ya que ello «reforzaría aún más la seguridad jurídica y la confianza en todos los sistemas de IA de alto riesgo». El SEPD se reafirma posteriormente⁸⁸ en dicha postura y añade que, teniendo en cuenta la legislación sectorial aplicable a la actividad en cuyo contexto se utilizará el sistema de IA, la evaluación por terceros del sistema de IA de alto riesgo, para garantizar la fiabilidad de la IA, requerirá la participación de la autoridad de supervisión que tenga conocimientos específicos en la materia.

Por tanto, consideramos que no puede dejarse a criterio del proveedor el someterse o no a la verificación por parte de un tercero tal y como han mantenido tanto el Parlamento⁸⁹, como el texto final. Por otro lado, tampoco se entiende por qué la referencia a la existencia o no de normas armonizadas o especificaciones comunes, solo se tiene en cuenta respecto a un tipo de sistemas de IA de alto riesgo (concretamente los relativos a identificación biométrica y categorización de personas físicas) y no se tiene presente de forma generalizada para todos ellos.

En relación al requisito de que los datos sean completos y, en la medida de lo posible, carezcan de errores, se ha puesto de manifiesto que el uso de técnicas como

83. Capítulo IV.

84. Artículo 43.

85. Artículo 47.

86. Artículo 48.

87. EDPB-EDPS *Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*, 18 junio de 2021, párrafo 37. Disponible en https://edps.europa.eu/system/files/2021-06/2021-06-18-edpb-edps_joint_opinion_ai_regulation_en.pdf

88. *Cit. EDPS, Opinion 44/2023*, párrafo 28.

89. Artículo 43.2 y Enmienda 453 en lo relativo al Artículo 43.1 d).

la privacidad diferencial, implica la introducción de ruido para evitar la revelación involuntaria de datos sensibles. Por este motivo algunos autores⁹⁰ abogan por que el artículo 10 permita la utilización de estas técnicas de mejora de la privacidad (*PET*) en las prácticas de gobernanza de datos de los sistemas de alto riesgo. A este respecto, destacar que el Consejo introdujo en el Considerando 44 (actual Considerando 67) la puntualización de que el requisito de que los conjuntos de datos sean completos y carezcan de errores, no debía afectar al uso de técnicas de protección de la privacidad en el contexto del desarrollo y la prueba de sistemas de IA. En la versión posterior del Parlamento desapareció dicha puntualización, pero finalmente ha sido recuperada por el texto definitivo, lo cual entendemos es positivo.

Como se ha comentado en el epígrafe anterior, el Consejo introdujo una primera modulación de dicha obligación de resultado, no con respecto a la pertinencia y representatividad de los datos, sino con respecto a la exigencia de carencia de errores y la completitud de los datos, al establecer que «en la mayor medida posible, carecerán de errores y estarán completos». El Parlamento por su parte, valida que los conjuntos de datos sean «suficientemente representativos», «debidamente evaluados en relación a los errores» y «tan completos como sea posible en vista de la finalidad prevista», por lo que es clara la dilución de la exigencia en cuanto a la calidad de los datos introducida por el Parlamento, que se recoge de forma similar en el texto final. En relación a la calidad de los datos, ya que ésta puede depender del contexto, es bienvenida la introducción por parte del Parlamento de dicha referencia a la finalidad del tratamiento prevista. Así lo propuso el gobierno noruego, al recomendar incluir en el párrafo tercero una referencia a la finalidad del tratamiento en el sentido de relacionar la pertinencia, necesidad y exactitud con dicha finalidad, tal y como realiza el RGPD al definir los Principios de Minimización y exactitud de datos en el artículo 5, apartado 1, letras c) y d), respectivamente.

Respecto a la *presunción*⁹¹ de que aquellos sistemas de IA de alto riesgo «que hayan sido entrenados y probados con datos relativos al entorno geográfico, conductual, contextual y funcional específico en el que esté previsto su uso» cumplen con los requisitos establecidos en el artículo 10, apartado 4, en nuestra opinión, resulta cuestionable. Según dicha presunción, bastará «entrenar y probar» un sistema con dichos datos para considerar que los conjuntos de datos utilizados tienen en cuenta «las características o los elementos propios del entorno geográfico, contextual, conductual o funcional específico en el que esté previsto utilizar el sistema de IA de alto riesgo» en la medida en que lo exija la finalidad prevista, lo cual parecen cosas bien distintas, pues deberá tenerse en cuenta, en todo caso, la finalidad prevista del sistema, por lo que dichas características o elementos variarán en cada caso. En todo caso, resulta una presunción un tanto difusa y genérica como para inferir el cumplimiento de un requisito tan importante como el establecido en el 10.4.

Además del contenido sustantivo, no podemos desconocer que, para comprobar su cumplimiento, en este caso, de los requisitos de gobernanza de datos, será necesario que el organismo competente disponga de las competencias necesarias para realizar inspecciones *in situ* y a distancia sin previo aviso, así como para acceder a los datos de entrenamiento, validación y prueba y código fuente de los sistemas de IA de alto

90. Cit. Ebers, M. *et alia*.

91. Artículo 42.1.

riesgo. Así lo había reclamado el SEPD⁹² y el Parlamento⁹³ lo propuso. Para ello, será requisito necesario que el proveedor, o sujeto obligado, esté en condiciones de ofrecer dichas muestras que la autoridad nacional de supervisión está habilitada para solicitar. El Parlamento propuso que el sujeto obligado deberá conservar evidencias y muestras suficientes para que la autoridad pueda «someter a ingeniería inversa a los sistemas de IA y adquirir pruebas para detectar los incumplimientos». No obstante, la versión final⁹⁴ no ha acogido dicha literalidad, pero establece⁹⁵ que el proveedor concederá a las autoridades de vigilancia del mercado pleno acceso a la documentación, así como a los conjuntos de datos de entrenamiento, validación y prueba utilizados e incluso, cuando proceda y con sujeción a las salvaguardias de seguridad, a través de interfaces de programación de aplicaciones («API») u otros medios técnicos y herramientas pertinentes que permitan el acceso a distancia. En determinados casos, se concederá acceso al código fuente⁹⁶. Por tanto, es claro que el proveedor deberá conservar los conjuntos de datos utilizados para el desarrollo del sistema⁹⁷, razón por la cual entendemos que hubiera sido deseable establecer claramente dicha obligación en el artículo 10, máxime si tenemos en cuenta la presunción comentada anteriormente, aunque el RIA no la configura expresamente como *iusuris tantum*.

En lo que respecta al *régimen sancionador* en esta materia, el Parlamento introdujo importantes cambios. Las versiones de la Comisión y del Consejo, establecían las mayores sanciones, por un lado, para las infracciones relativas a las prácticas de inteligencia artificial prohibidas (artículo 5) y las relativas al incumplimiento de los requisitos establecidos para los datos y gobernanza de datos (Artículo 10), con multas de hasta 30 000 000 EUR o, si el infractor es una empresa, de hasta el 6 % del volumen de negocio total anual mundial del ejercicio financiero anterior, si esta cuantía fuese superior, y por otro, el incumplimiento del resto de los requisitos u obligaciones establecidos en el Reglamento, con multas administrativas de hasta 20 000 000 EUR o, si el infractor es una empresa, de hasta el 4 % del volumen de negocio total anual mundial. El Parlamento propuso aumentar las sanciones relativas a las prácticas de IA prohibidas hasta 40.000.000 euros pero, curiosamente, eliminando de dicho rango las infracciones relativas al artículo 10, y creando un nuevo rango sancionador por el incumplimiento de los requisitos establecidos a los datos y gobernanza de datos y las obligaciones de transparencia⁹⁸ con sanciones de 20 000 000 EUR o, si el infractor

92. Cit. EDPS, *Opinion 44/2023*, párrafo 45.

93. Enmienda 587, que introduce un nuevo apartado 3 bis) en el artículo 63.

94. Artículo 74.5.

95. Artículo 74.12.

96. Artículo 74.13.

97. Ex artículo 18, deberá conservarse durante diez años la documentación técnica (Anexo XI) que incluye (Sección 1, punto 2c): «información sobre los datos utilizados para el entrenamiento, las pruebas y la validación, cuando proceda, incluidos el tipo y la procedencia de los datos y los métodos de gestión (por ejemplo, limpieza, filtrado, etc.), el número de puntos de datos, su alcance y sus principales características; cómo se obtuvieron y seleccionaron los datos, así como cualquier otra medida que permita detectar que las fuentes de datos no son idóneas y los métodos para detectar sesgos identificables, cuando proceda». Obsérvese que no se habla de la totalidad de los conjuntos de datos.

98. Enmienda 650, artículo 71, nuevo apartado 3 bis.

es una empresa, de hasta el 4 % del volumen de negocio total anual mundial del ejercicio financiero anterior. Para el resto de infracciones de determinados artículos, propuso reducir las sanciones a la mitad. También propuso reducir a la mitad⁹⁹ las infracciones por presentar información inexacta, incompleta o engañosa a organismos notificados y a las autoridades nacionales competentes, lo cual, en un sistema basado en la «autoevaluación» del cumplimiento de los requisitos, tiene especial relevancia. Finalmente, las sanciones¹⁰⁰ más graves son únicamente por infracción del artículo 5 (prácticas prohibidas) y supondrán multas de hasta 35.000.000 EUR o de hasta el 7 % de su volumen de negocios total anual a escala mundial correspondiente al ejercicio anterior, si esta cifra es superior. Se incluye un catálogo de determinadas disposiciones, entre las que no se encuentra el artículo 10, cuya infracción se sanciona con multas de hasta 15 000 000 EUR o, si el infractor es una empresa, de hasta el 3% de su volumen de negocios total anual. Por tanto, la infracción del artículo 10 ha pasado de ser una de las infracciones más graves, a no aparecer en el régimen sancionador, quizá por error al haber eliminado la versión final el nuevo rango sancionador que propuso el Parlamento para las infracciones del artículo 10 y 13 del RIA.

De otro lado, la sanción por suministro de información incorrecta, incompleta o engañosa a los organismos notificados y a las autoridades nacionales competentes se eleva a multas administrativas de hasta 7.500.000 EUR o, si el infractor es una empresa, de hasta el 1 % de su volumen de negocios total anual, por lo que tampoco se acogió la propuesta del Parlamento en este punto.

Asimismo, llama la atención que, a pesar del mandato¹⁰¹ general de que las sanciones tendrán particularmente en cuenta los intereses de las pymes y las empresas emergentes, así como su viabilidad económica, el Parlamento propusiera eliminar la modulación de responsabilidad introducida por el Consejo en relación a las Pymes y empresas emergentes, estableciendo un porcentaje inferior en cuanto a su volumen de negocios anual mundial en todas las sanciones. La versión final recupera la mención¹⁰² a las PYMES y *start ups* y recoge una modulación de la responsabilidad consistente en aplicar el porcentaje o el importe de la sanción, según cuál de ellos sea menor, al contrario de lo establecido en el régimen sancionador general, en el que se deberá optar por la mayor cuantía. Consideramos que, a pesar de que la introducción de dicha modulación es positiva, pero solo beneficiará a aquellas PYMES y *start ups* cuyo volumen de negocios total anual sea muy elevado.

Entendemos, al igual que otros autores¹⁰³, que se ha construido un sistema de cumplimiento basado en la «autoevaluación» sin la intervención obligatoria de organismos externos, lo cual, unido a la disminución de las sanciones, incluso por facilitar información inexacta, incompleta o engañosa a las autoridades u organismos notificados, reduce notablemente el grado de seguridad jurídica esperado a alcanzar con el Reglamento. Incluso si pensamos en los sistemas de IA de alto riesgo, que han

99. Enmienda 652.

100. Artículo 99.2.

101. Artículo 99.1.

102. Artículo 99.6.

103. Cit. Ebers, M. *et alia*, p. 601; PEGUERA POCH, M., *La propuesta de reglamento de IA: una intervención legislativa insoslayable en un contexto de incertidumbre*, Capítulo cerrado a 20 de mayo de 2023, p. 24. Publicado en: Peguera Poch, Miquel (coord.) «Perspectivas regulatorias de la Inteligencia Artificial en la Unión Europea», Madrid: Reus, 2023.

pasado de ser un listado de *numerus clausus* a, con las modificaciones introducidas por el Consejo, tener que cumplir el criterio cumulativo de suponer «un riesgo importante para la salud, la seguridad o los derechos fundamentales», en último término, también está en manos de los proveedores determinar si estamos ante un sistema de alto riesgo o no. EBERS *et alia*¹⁰⁴ lo resumen muy bien al indicar que «en contraste con la inminente sobrerregulación atribuible a la amplia definición de IA, el enfoque de autocumplimiento plantea problemas de infrarregulación» (la traducción es nuestra).

V. CONFLUENCIA DE LA NORMATIVA DE PROTECCIÓN DE DATO

En el presente apartado abordaremos las conexiones de los requisitos en materia de gobernanza de datos con los principios de protección de datos, ya que la interacción del RIA con la normativa de protección de datos ha sido tratada a nivel general en otro capítulo de la presente obra rubricado por Jiménez López.

Si tenemos en cuenta que una de las bases jurídicas del RIA es el artículo 16 del Tratado de Funcionamiento de la Unión Europea (TFUE), la importancia de la normativa de protección de datos en el RIA queda fuera de toda duda. Debe tenerse presente que muchos sistemas de IA serán entrenados o tratarán datos personales, o bien ayudarán a las personas a la toma de decisiones o directamente podrán tomar y ejecutar la decisión, por lo que el RGPD será plenamente de aplicación. No obstante, el RIA no incluye dentro de su articulado la obligación general de cumplir con la normativa de protección de datos, sin perjuicio de menciones a obligaciones concretas. Lo más parecido es la exigencia¹⁰⁵, introducida por el Parlamento y recogida en el texto definitivo, de que la declaración de conformidad incluya una declaración de que el sistema de IA cumple el RGPD.

Y no es una cuestión baladí. No en vano, en un principio la infracción de los requisitos de gobernanza de datos se configuró al mismo nivel sancionador que las prácticas prohibidas. A estas alturas no debemos partir de la premisa de que la tecnología es neutra, sino más bien de todo lo contrario, tal como afirma Floridi¹⁰⁶. Ni siquiera lo es la aproximación para la regulación del riesgo que se utilice¹⁰⁷, por lo que tanto el diseño como los datos que se utilicen, son absolutamente relevantes como podemos ver en el RIA. Las consecuencias de no contar con el tipo de datos adecuado, ni con la calidad requerida, podrían ser nefastas, al condicionar desde el diseño los resultados, siendo, por tanto, no válidos, y lo que es más importante, pudiendo afectar a los derechos fundamentales de las personas. La relación entre los datos y el sistema de IA es, por tanto, directamente proporcional a la calidad de los resultados obtenidos. Es por ello que el artículo 10 recoge, entre las prácticas adecuadas de gobernanza y gestión de datos, cuestiones relativas al diseño del sistema y a la transparencia y calidad de los datos.

104. Cit. Ebers, M. *et alia*, p. 601.

105. Anexo V, punto 5.

106. Op. Cit. FLORIDI, L., «On Good and Evil...».

107. Op. Cit. KAMINSKI M., «Regulating de risk of AI»..., p. 1351.

El CEPD y el SEPD afirmaban¹⁰⁸ que «la propuesta (de reglamento) carece de una relación clara con la legislación en materia de protección de datos». Otros autores¹⁰⁹ indicaban que el RIA «debería procurar una mejor armonización y coordinación con la normativa de protección de datos». Esta problemática se ha visto en parte reducida gracias a las enmiendas introducidas en esta materia por el Parlamento Europeo, al positivizar en el RIA la importancia del cumplimiento de la normativa de protección de datos que, no por no mencionarse dejaba de ser de obligado cumplimiento, pero pone de relieve la importancia del mismo en el ámbito de la IA.

A ello debemos añadir que el RIA carece de principios informadores que guíen a los diferentes sujetos obligados en la aplicación del mismo y que presidan cualquier interpretación que los operadores jurídicos realicen. En este sentido, el Parlamento propuso introducir¹¹⁰ una serie de principios generales aplicables a todos los sistemas de IA que, al informar la aplicación del RIA, podrían ser exigibles a todos los operadores dentro de su ámbito de aplicación, tal y como ocurre con el RGPD. No obstante, por alguna razón que desconocemos, esta propuesta no fue recogida en la versión final. Entre los principios que proponía el Parlamento figuraba el de «*Privacidad y gobernanza de datos*: los sistemas de IA se desarrollarán y utilizarán de conformidad con las normas vigentes en materia de privacidad y protección de datos, y tratarán datos que cumplan normas estrictas en términos de calidad e integridad». Dicho principio evidencia el mutuo condicionamiento existente entre el derecho de protección de datos y las obligaciones en materia de gobernanza de datos.

Desde el punto de vista de las obligaciones relativas a la gobernanza de datos, se ha puesto en evidencia posibles carencias de la regulación actual. No obstante, las obligaciones establecidas en el artículo 10 aplican con independencia de si se trata de datos personales o no, y sin perjuicio de las obligaciones en su caso derivadas de la aplicación del RGPD. De la misma forma, prácticas admitidas por el RIA podrían no resultar viables de no cumplir con los requisitos establecidos por la normativa de protección de datos¹¹¹. Por tanto, es claro que los principios de protección de datos aplicarán en todo caso. No obstante, en este punto el CEPD y el SEPD¹¹², en relación al sistema de certificación, proponían que se incluyeran los principios de minimización y protección de datos desde el diseño como uno de los requisitos a tener en cuenta para obtener el marcado CE, debido al «posible alto nivel de interferencia de los sistemas de IA de alto riesgo con los derechos fundamentales a la privacidad y a la protección de los datos personales, y la necesidad de garantizar un alto nivel de confianza en el sistema de IA». Opinión que posteriormente reitera el SEPD¹¹³.

108. CEPD-SEPD, Dictamen conjunto 5/2021 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial), 18 de junio de 2021, párrafo 76.

109. COTINO L., CASTILLO J.A., SALAZAR I., BENJAMINS R., CUMBRERAS M., ESTEBAN A., «Un análisis crítico constructivo de la Propuesta de Reglamento de la Unión Europea por el que se establecen normas armonizadas sobre la Inteligencia Artificial (*Artificial Intelligence Act*)», en Diario La Ley, Wolters Kluwer, 2 de julio de 2021.

110. Enmienda 213. Artículo 4 bis:

111. Vid. Considerando 63.

112. Op. Cit. CEPD-SEPD, Dictamen conjunto 5/2021.. párrafo 76.

113. SEPD, Opinion 44/2023 on the Proposal for Artificial Intelligence Act in the light of legislative developments, 23 octubre 2023, párrafo 27.

A pesar de que el artículo 10, ni el RIA en general, no recojan expresamente el cumplimiento de ningún principio de protección de datos, los Considerandos sí lo hacen. Así, el Considerando 67 indica que, para facilitar el cumplimiento de la normativa de protección de datos, las prácticas de gobernanza y gestión de datos deben incluir, en el caso de los datos personales, la transparencia sobre la finalidad original de la recopilación de datos. Por tanto, el principio de transparencia en protección de datos se convierte en una condición para cumplir con este requisito en materia de gobernanza de datos, y viceversa, pues como afirma la AEPD¹¹⁴ «la información disponible en el marco de la Transparencia-RIA debería ser lo suficientemente completa como para permitir a los responsables y encargados del tratamiento cumplir sus diferentes obligaciones con arreglo al RGPD». Por su parte, el Considerando 69 establece que «el derecho a la intimidad y a la protección de los datos personales debe garantizarse a lo largo de todo el ciclo de vida del sistema de IA. A este respecto, los principios de minimización de datos y de protección de datos desde el diseño y por defecto, establecidos en el Derecho de la Unión en materia de protección de datos, son aplicables cuando se tratan datos personales». El Considerando 67 también aclara que el requisito de que los conjuntos de datos sean, en la medida de lo posible, completos y estén libres de errores «no debe afectar al uso de técnicas de preservación de la intimidad en el contexto del desarrollo y las pruebas de los sistemas de IA» y en el mismo sentido el Considerando 69, cuando indica que los proveedores para garantizar el cumplimiento de dichos principios podrán utilizar «tecnología que permita llevar algoritmos a los datos y permita el entrenamiento de los sistemas de IA sin la transmisión entre las partes o la copia de los propios datos en bruto o estructurados, sin perjuicio de los requisitos sobre gobernanza de datos previstos en el presente Reglamento».

Cuando el artículo 10 exige que los datos estén exentos de errores y sean completos con vistas a la finalidad prevista, entendemos se está haciendo directa referencia al principio de exactitud. Tal y como afirma la AEPD¹¹⁵ «el comportamiento de un algoritmo, entre otros los algoritmos de inteligencia artificial (IA), podría verse comprometido por la inexactitud de los datos de entrada utilizados en la ejecución del mismo, no solo por los datos utilizados en su desarrollo», razón por la cual «es necesario evaluar la exactitud de los datos de entrada, ya que podría introducir sesgos y comprometer el rendimiento no solo del algoritmo, sino de todo el tratamiento».

Por tanto, deben establecerse controles que eviten la introducción de datos de entrada inexactos y también controles para establecer las adecuadas medidas de protección en el caso de introducirse datos inexactos. A esto responde la obligación establecida en el artículo 10 de establecer las medidas adecuadas para detectar, prevenir y mitigar los posibles sesgos que se detecten.

Precisamente para garantizar la detección y corrección de sesgos en relación con los sistemas de IA de alto riesgo, se habilita excepcionalmente a los proveedores de dichos sistemas a tratar categorías especiales de datos siempre que se cumplan una

114. AEPD, «Inteligencia artificial: Transparencia», 20 de septiembre de 2023. <https://www.aepd.es/prensa-y-comunicacion/blog/inteligencia-artificial-transparencia>

115. AEPD, «Inteligencia Artificial: principio de exactitud en los tratamientos», 31 de mayo de 2023 <https://www.aepd.es/prensa-y-comunicacion/blog/inteligencia-artificial-principio-de-exactitud-en-los-tratamientos>

serie de condiciones (i) que no pueda realizarse mediante datos sintéticos, anónimos u otros datos; (ii) que las categorías especiales de datos personales tratados se sujeten a limitaciones técnicas en cuanto a la reutilización y a las medidas más avanzadas de seguridad (iii) se sujeten a medidas que garanticen la seguridad y protección de los datos personales tratados (iv) no se transmitirán, transferirán ni serán accesibles de otro modo a terceros; (v) se suprimirán una vez que se haya corregido el sesgo o los datos personales hayan llegado al final de su período de conservación. Este artículo hace referencia al principio de licitud para el tratamiento de dichas categorías especiales de datos. Hay autores¹¹⁶ que sostienen que supone una excepción al RGPD por constituir en sí mismo una base de legitimación. Por el contrario, entendemos que será precisa una base de legitimación adecuada, en primer lugar, por la propia aplicación del RGPD y, en segundo lugar, porque el propio párrafo 5, cuando enumera las condiciones necesarias para que se pueda producir el tratamiento expresamente indica que deberán tenerse en cuenta las disposiciones establecidas en el Reglamento (UE) 2016/679 (..).

Por tanto, queda patente la importancia de los principios de protección de datos en relación a la gobernanza de datos, lo cual pone de manifiesto la importantísima interrelación e interdependencia entre ambos marcos normativos. Tan es así, que si atendemos al ámbito de aplicación tanto subjetivo como material de la Evaluación de Impacto en Derechos Fundamentales¹¹⁷ que ha quedado totalmente desdibujado hasta el punto, en nuestra opinión, de no poder cumplir la finalidad para la que fue concebida, la normativa de protección de datos y, especialmente, sus principios y la Evaluación de Impacto en Protección de Datos, se erige como guardiana en último término de los citados derechos fundamentales, sin perjuicio de que el análisis de riesgos los incluya dentro de su ámbito objetivo.

VI. CONCLUSIONES

Primera. A nivel de gobernanza de la IA, consideramos que es necesario establecer un marco de gobernanza internacional. Las iniciativas a nivel de regulación en materia de IA en diferentes continentes evidencian la necesidad de una regulación internacional y, por tanto, del establecimiento de los necesarios mecanismos de coordinación¹¹⁸. No obstante lo anterior, Europa, consciente de sus carencias en materia de soberanía tecnológica y buscando salvaguardar la salud, seguridad y derechos fundamentales, ha establecido su propio marco de gobernanza de IA mediante el que aspira a repetir el «efecto Bruselas» que obtuvo con el Reglamento General de Protección de Datos. Para garantizar el mercado único de datos (gobernanza a nivel «macro») es imprescindible que las organizaciones cuenten con una sólida gobernanza de datos a nivel interno (nivel «micro»), la cual además permitirá avanzar hacia la gobernanza de la inteligencia artificial. No debe caerse en el error de que dichas obligaciones solo recaen en las entidades que desarrollan los sistemas de IA, sino que aquellas que los diseñan o los despliegan (implementadores) también

116. Op. Cit. Ebers, M. *et alia*, p. 600.

117. Artículo 27.

118. ROBERTS, H., HINE, E., TADDEO, M. y FLORIDI, L., «Global AI governance: barriers and pathways forward», 29 septiembre de 2023. <http://dx.doi.org/10.2139/ssrn.4588040>

tienen responsabilidades, por lo que, aunque a diferentes niveles, es necesario que todas las organizaciones establezcan mecanismos de gobernanza de la IA. Nunca ha sido tan importante la gobernanza, no sólo a nivel de implementación y gestión, sino que ha de comenzar por los órganos de dirección que son los responsables de marcar y liderar la estrategia en materia de IA, así como supervisar su aplicación. Si tuviéramos que resumir en una palabra el RIA, ésta sería «Gobernanza». En relación a la gobernanza de datos consideramos que debe utilizarse un concepto amplio, que no se refiera únicamente al establecido en el artículo 10, sino que incluya también el seguimiento posterior¹¹⁹ a la comercialización y, además, un seguimiento a largo plazo para detectar riesgos sistémicos en relación a la erosión gradual de las instituciones y los valores sociales y políticos¹²⁰.

Segunda. El artículo 10 relativo a los datos y su gobernanza es de una importancia capital ya que, del cumplimiento de las obligaciones establecidas en el mismo, deriva disponer de datos de alta calidad y, por tanto, el correcto funcionamiento de los sistemas de IA, especialmente, los de alto riesgo. Establece los requisitos («criterios de calidad») que deben cumplir los conjuntos de datos que se utilicen para el entrenamiento, validación y prueba de los sistemas de alto riesgo. Resulta de una importancia capital contar con datos de calidad, tanto para el entrenamiento, como para el desarrollo del sistema, ya que, de lo contrario, tanto el propio sistema, como sus resultados podrán verse afectados, cuestión de vital importancia cuando estamos hablando de seguridad y derechos fundamentales. Por esta razón, contar con un robusto sistema de gobernanza de datos resulta imperativo y trascendental, tanto para garantizar el correcto funcionamiento del sistema, como para acreditar la necesaria responsabilidad proactiva. Las obligaciones en materia de gobernanza de datos deben conectarse necesariamente con el sistema de gestión de la calidad y, en especial, con el sistema de gestión de riesgos, aunque no se indique expresamente. Ciertamente es que el sistema de gobernanza de datos configurado por el RIA tiene entidad propia de forma que trasciende al sistema de gestión de riesgos, pero ello no implica que sea desconocido por este último. La inclusión de la referencia expresa al artículo 10 hubiera sido deseable, tanto en el sistema de gestión de la calidad como en el análisis de riesgos, no sólo por el hecho de remarcar la importancia del cumplimiento del artículo 10, sino por razones de coherencia sistemática.

Tercera. Entendemos que se dan todas las circunstancias para la irrupción en la cadena de valor de nuevas figuras que exclusivamente provean de datos «verificados» y certifiquen que los datos cumplen con los requisitos de gobernanza establecidos, así como su integridad y entrenamiento («verificadores de datos» o «proveedores de servicios certificados de datos»). Resultará, por tanto, clave la regulación y posible transmisión de responsabilidad. Hemos cuestionado si este modelo de provisión de datos «verificados» puede ofrecer el cumplimiento individualizado de dichos requisitos de gobernanza al que aspira el Reglamento ya que, en primer lugar, la gobernanza debe adecuarse al contexto del uso, así como a la finalidad prevista del sistema de IA¹²¹ y, en segundo lugar, los conjuntos de datos deberán tener en cuenta, en la medida en que lo exija la finalidad prevista, las características o elementos

119. Anexo IV, 2. d) y g).

120. Op. Cit. KOLT, N., *Algorithmic Black Swans*, p.37.

121. Artículo 10, párrafo segundo.

propios del entorno geográfico, contextual conductual o funcional en el que está previsto utilizar el sistema de IA de alto riesgo. Por tanto, será responsabilidad última del implementador valorar la adecuación de dichos conjuntos de datos al caso de uso para el que utilizará el sistema de IA. En otras palabras, el hecho de que puedan irrumpir en la cadena de valor nuevas figuras consecuencia de la externalización de los requisitos relacionados con la gobernanza de datos, no exime al implementador (y, en su caso, responsable del tratamiento) del cumplimiento del resto de obligaciones, pues «la responsabilidad proactiva es una piedra angular de la gobernanza de la inteligencia artificial»¹²² tanto proactiva (*ex ante*) como reactiva (*ex post*). Sin perjuicio del cuestionamiento de dicho modelo, resultará clave la regulación y posible transmisión de responsabilidad a nivel contractual.

Cuarta. Para comprobar el cumplimiento de los requisitos en materia de gobernanza de datos por parte de las autoridades competentes, el proveedor o el sujeto obligado, deberá conservar los conjuntos de datos utilizados para el desarrollo y entrenamiento del sistema. Así se desprende de las medidas de vigilancia post mercado¹²³ al indicar que «los proveedores concederán a las autoridades de vigilancia del mercado pleno acceso a la documentación, así como a los conjuntos de datos de entrenamiento, validación y prueba utilizados para el desarrollo de los sistemas de IA de alto riesgo». Por esta razón entendemos que hubiera sido deseable establecer claramente dicha obligación en el propio artículo 10. Además, deberá evidenciarse que el sistema de IA cumple con el RGPD, con todo lo que ello implica, tal y como se afirma en la declaración de conformidad.

Quinta. A pesar de que no se incluya expresamente, la aplicación de la normativa de protección de datos va a jugar un papel absolutamente necesario en lo que respecta a los criterios de calidad que deben cumplir los conjuntos de datos, pues deben observarse los principios en materia de protección de datos. Además, la normativa de protección de datos, mediante su Evaluación de Impacto en Protección de Datos, va a jugar un papel fundamental para salvaguardar los derechos y libertades de los interesados, cubriendo en parte el espacio que debería ocupar la Evaluación de Impacto en Derechos Fundamentales.

Sexta. El RIA supone un reto de cumplimiento, ya que aún una compleja parte de requisitos y normas técnicas y/o armonizadas, junto con la aplicación de la normativa de protección de datos y derechos fundamentales, a lo que debemos sumar la interacción de diferentes roles cuyo cumplimiento debe ser supervisado en última instancia por el implementador. Si bien es cierto que la naturaleza de la IA hace que sea problemático evaluar la responsabilidad de los resultados obtenidos¹²⁴, ello traslada el peso de la responsabilidad proactiva a poder acreditar la pertinencia, representatividad suficiente, análisis de posibles errores y la completitud de los datos, lo cual se logrará con unos robustos sistemas de gobernanza de datos.

122. Op. Cit. NOVELLI C., TADDEO M., FLORIDI L., Accountability in artificial intelligence.

123. Artículo 74.12).

124. Op. Cit. NOVELLI C., TADDEO M., FLORIDI L., Accountability in artificial intelligence.

Sistemas de gestión de calidad, documentación técnica y conservación en el Reglamento

FRANCISCA RAMÓN FERNÁNDEZ

Catedrática de Derecho Civil de la Universitat Politècnica de València¹

I. INTRODUCCIÓN

En el presente estudio nos vamos a ocupar de la regulación de los sistemas de gestión de calidad, documentación técnica y conservación de los sistemas de inteligencia artificial de alto impacto en las diferentes propuestas, así como en el texto definitivo de la Ley de Inteligencia Artificial en relación con los sistemas de Inteligencia Artificial de alto riesgo que se contemplan en la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión COM(2021) 206 final 2021/0106 (COD), de 21 de abril de 2021², junto con los Anexos³, y en el texto definitivo P9_TA(2024)0138 referente a la Resolución legislativa del Parlamento Europeo, de 13 de marzo de 2024, sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión (COM(2021)0206 — C9-0146/2021 —2021/0106(COD)), RIA⁴.

1. Trabajo realizado en el marco del Grupo de Investigación de Excelencia Generalitat Valenciana «Algorithmical Law» (Proyecto Prometeu 2021/009, 2021-2024), y Proyecto «Derechos y garantías públicas frente a las decisiones automatizadas y el sesgo y discriminación algorítmicas» 2023-2025 (PID2022-136439OB-I00) financiado por MCIN/AEI/10.13039/501100011033/ FEDER, UE.
2. Texto de la Propuesta disponible en: https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0008.02/DOC_1&format=PDF (Consultado el 25 de julio de 2023).
3. Disponible en: https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0008.02/DOC_2&format=PDF (Consultado el 25 de julio de 2023).
4. Texto del Reglamento de Inteligencia Artificial disponible en: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_ES.pdf (Consultado el 15 de marzo de 2024). Véase también: Mühlhoff, R. y Ruschemeier, H., «Regulating AI with purpose limitation for models», *Journal of AI Law and Regulation*, n.º 1 (2024), pp. 24-39. Disponible en: https://aire.lexxion.eu/data/article/19395/pdf/aire_2024_01-006.pdf (Consultado el 18 de marzo de 2024).

La normativa objeto de análisis se recoge en el capítulo (antes en las versiones anteriores se refería a título) III dedicado a regular los sistemas de inteligencia artificial de alto riesgo, en la sección (antes en las versiones anteriores se refería a capítulo) 2 dedicado a los requisitos para los sistemas de inteligencia artificial de alto riesgo, en el artículo 11 que hace referencia a la documentación técnica. Estos requisitos se derivan de las directrices éticas para una inteligencia artificial fiable que fueron elaboradas por el grupo independiente de expertos de alto nivel sobre inteligencia artificial que se creó por la Comisión Europea, en junio de 2018⁵. Se contempla la flexibilidad respecto a las soluciones técnicas que se necesitan para lograr el cumplimiento de los requisitos indicados que podrán proceder de normas o especificaciones técnicas o ser objeto de desarrollo de acuerdo con los conocimientos científicos o de ingeniería, de forma discrecional por parte del proveedor del sistema de inteligencia artificial. Así, se permite a los proveedores de los sistemas decidir de qué forma quieren cumplir los requisitos teniendo en cuenta el estado de la técnica y los avances tecnológicos y científicos.

En la sección (antes en las versiones anteriores se refería a capítulo) 3 que se ocupa de las obligaciones de los proveedores y responsables del despliegue (antes usuarios) de sistemas de inteligencia artificial de alto riesgo y de otras partes, en concreto el artículo 17 sobre el sistema de gestión de calidad, y el artículo 18 inicialmente destinado a contemplar la obligación de elaborar documentación técnica, y que después en la posterior versión de la Propuesta de Reglamento de 2022 y la Resolución legislativa del Parlamento Europeo de 2024, pasa a denominarse conservación de la documentación, y recoger el contenido del artículo 50 que se contiene en la sección (antes en las versiones anteriores se refería a capítulo) 5 sobre las normas, evaluación de la conformidad, certificados, registro y que se ocupaba de la conservación de los documentos, quedando posteriormente suprimido este artículo 50.

Se trata de un conjunto de obligaciones horizontales que se imponen a los proveedores de sistemas de inteligencia artificial de alto riesgo⁶, y también se establecen obligaciones para los usuarios y otros participantes de la cadena de valor de la inteligencia artificial, como pueden ser los importadores, distribuidores y representantes autorizados⁷.

Como indica el considerando 9 «el presente Reglamento tiene por objeto reforzar la eficacia de tales derechos y vías de recurso vigentes mediante el establecimiento

5. Unión Europea, *Directrices Éticas para una IA fiable. Grupo de expertos de alto nivel sobre inteligencia artificial*, Comisión Europea, Bruselas 2019.
6. Véase: Cotino Hueso, L., «Los usos de la inteligencia artificial en el sector público, su variable impacto y categorización jurídica», *Revista Canaria de Administración Pública*, n.º 1 (2023), pp. 211-242. Disponible en: <https://revistacanarias.tirant.com/index.php/revista-canaria/article/view/7/7> (Consultado el 24 de julio de 2023).
7. Cfr. Ramón Fernández, F., «Inteligencia artificial y transparencia en relación con la regulación de los servicios y mercados digitales», *Equidad y transparencia en la prestación de servicios*, María Elena Cobas Cobiella y Raquel Guillén Catalán, directoras, Dykinson, Madrid (2023), pp. 147-169. También puede resultar de interés: Argelich Comelles, C., «Gobernanza de las plataformas en línea ante la DSA y las Propuestas de Reglamento de Mercados Digitales e Inteligencia Artificial (DMA y RIA). (Gobernanza de plataformas en línea frente a DSA, DMA y RIA de la UE)», *Anuario de Derecho civil*, tomo LXXV, fasc. II, pp. 501-530. Disponible en: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4434522 (Consultado el 11 de noviembre de 2023).

de requisitos y obligaciones específicos, en particular en lo que respecta a la transparencia, la documentación técnica y la conservación de registros de los sistemas de IA». De igual modo, el considerando 66 nos precisa que «Deben aplicarse a los sistemas de IA de alto riesgo requisitos referentes a la gestión de riesgos, la calidad y la pertinencia de los conjuntos de datos utilizados, la documentación técnica y la conservación de registros, la transparencia y la comunicación de información a los responsables del despliegue, la supervisión humana, la solidez, la precisión y la ciberseguridad. Dichos requisitos son necesarios para reducir de forma efectiva los riesgos para la salud, la seguridad y los derechos fundamentales, y no se dispone razonablemente de otras medidas menos restrictivas del comercio, con lo que se evitan restricciones injustificadas de este».

En el antes capítulo 5 donde se ubicaba el artículo 50, se explicaba de forma detallada los procedimientos de evaluación de la conformidad que debía seguirse para cada tipo de sistema de IA de alto riesgo. Esto tenía como finalidad la reducción de la carga que soportarían los operadores económicos y los organismos notificados. Los sistemas de IA que se destinen a ser utilizados como componentes de seguridad de productos regulados por la legislación del nuevo marco normativo, por ejemplo, máquinas, productos sanitarios o juguetes se sujetarán a los mismos mecanismos de cumplimiento antes y posteriormente que los productos en los que se integran⁸.

8. Como indicaba la Propuesta de Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de IA (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión, de 2021, «se establecerá un nuevo sistema de cumplimiento y aplicación para los sistemas de IA de alto riesgo independientes que se mencionan en el anexo III. Este seguirá el modelo de la legislación del nuevo marco legislativo que los proveedores aplicarán mediante controles internos, con la salvedad de los sistemas de identificación biométrica remota, que se someterán a evaluaciones de la conformidad efectuadas por terceros. Una solución efectiva y razonable para dichos sistemas podría consistir en realizar una evaluación integral de la conformidad ex ante mediante controles internos, combinada con una supervisión ex post estricta, dado que la intervención reguladora se encuentra en una fase temprana, que el sector de la IA es muy innovador y que apenas están empezando a acumularse los conocimientos necesarios para llevar a cabo auditorías. Para evaluar los sistemas de IA de alto riesgo “independientes” mediante controles internos, sería necesario cumplir ex ante de manera plena, efectiva y debidamente documentada todos los requisitos del Reglamento, así como los sólidos sistemas de gestión de la calidad y los riesgos y el seguimiento posterior a la comercialización. Una vez que el proveedor haya llevado a cabo la evaluación de la conformidad oportuna, deberá registrar dichos sistemas de IA de alto riesgo independientes en una base de datos de la UE que la Comisión gestionará con el propósito de redoblar la transparencia y la vigilancia públicas y de fortalecer la supervisión ex post por parte de las autoridades competentes. En cambio, por motivos de coherencia con la legislación vigente relativa a la seguridad de los productos, la evaluación de la conformidad de los sistemas de IA que son componentes de seguridad de productos seguirá un sistema en el que terceros llevarán a cabo procedimientos de evaluación de la conformidad ya definidos en la legislación sectorial sobre seguridad de los productos pertinente. Si se realizan modificaciones sustanciales en los sistemas de IA (fundamentalmente cambios que trasciendan los aspectos predeterminados por el proveedor en su documentación».

Hay que tener en cuenta lo indicado en el Reglamento (UE) 2023/1230 del Parlamento Europeo y del Consejo, de 14 de junio de 2023, relativo a las máquinas⁹. Resaltar la importancia de la regulación objeto de estudio. La inteligencia artificial además es uno de los cinco objetivos específicos interrelacionados del Programa Europea Digital que establece el Reglamento (UE) 2021/694 del Parlamento Europeo y del Consejo de 29 de abril de 2021 y por el que se deroga la Decisión (UE) 2015/2240¹⁰.

En el título III se contienen las normas específicas para los sistemas de inteligencia artificial que conllevan un alto riesgo para la salud y la seguridad o los derechos fundamentales de las personas. Hay que ponderar los riesgos que pueden producir con la implementación de estos sistemas en el mercado europeo siempre que cumplan unos requisitos de obligado cumplimiento y que sean evaluados anteriormente a su introducción en el mercado de la Unión. La función, la finalidad y las modalidades de uso del sistema serán los factores para la calificación de un sistema de inteligencia artificial de alto riesgo.

El Informe de la Comisión al Parlamento Europeo, al Consejo y al Comité Económico y Social Europeo sobre las repercusiones en materia de seguridad y responsabilidad civil de la inteligencia artificial, el internet de las cosas y la robótica que se anexa al Libro Blanco sobre la inteligencia artificial-un enfoque europeo orientado a la excelencia y la confianza, de 19 de febrero de 2020 [COM (2020) 65 final]¹¹ indica que «el comportamiento autónomo de algunos sistemas de IA a lo largo de su ciclo de vida puede conllevar importantes cambios en los productos y tener repercusiones en la seguridad, lo que puede requerir una nueva evaluación de riesgos. Además, es probable que se requiera la supervisión humana como garantía desde la fase de diseño y a lo largo de todo el ciclo de vida de los productos y sistemas de IA».¹²

Respecto a ese alto riesgo y la necesidad de control y valoración previa cabe mencionar lo indicado en el Libro Blanco sobre la inteligencia artificial-un enfoque europeo orientado a la excelencia y la confianza COM(2020) 65 final, de 19 de febrero de 2020.¹³ Sobre un control objetivo previo de la conformidad para verificar y garantizar el cumplimiento de algunos de los requisitos obligatorios previamente mencionadas por parte de las aplicaciones de elevado riesgo, y ese control puede incluir procedimientos de ensayo, inspección o certificación, así como también disponer de controles de los algoritmos y de los conjuntos de datos utilizados en la fase de desarrollo.

9. Disponible en: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0132_ES.pdf (Consultado el 15 de marzo de 2024).

10. DOUE L 166, de 11 de mayo de 2021. Disponible en: <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32021R0694> (Consultado el 24 de julio de 2023).

11. Disponible en: <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52020DC0064> (Consultado el 25 de julio de 2023).

12. Ramón Fernández, F., «El robot como producto defectuoso y responsabilidad civil», *Derecho Digital e Innovación*, n.º 14 (2022), pp. 1-28.

13. Disponible en: <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52020DC0065> (Consultado el 25 de julio de 2023).

Nos remitimos a lo indicado en la Decisión núm. 768/2008/CE del Parlamento Europeo y del Consejo, de 9 de julio de 2008 («Reglamento sobre la Ciberseguridad»). También interesa destacar la Decisión de la Comisión, de 24 de enero de 2024, por la que se crea la Oficina Europea de Inteligencia Artificial.

También debemos mencionar, en el ámbito español, el Real Decreto 817/2023, de 8 de noviembre, que establece un entorno controlado de pruebas para el ensayo del cumplimiento de la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial en el que se indica que se pone en «marcha el primer entorno controlado de pruebas para comprobar la forma de implementar los requisitos aplicables a los sistemas de inteligencia artificial de alto riesgo de la propuesta de reglamento europeo de inteligencia artificial con el ánimo de obtener, como resultado de esta experiencia, unas guías basadas en la evidencia y la experimentación que faciliten a las entidades, especialmente las pequeñas y medianas empresas, y a la sociedad en general, el alineamiento con la propuesta del Reglamento Europeo de Inteligencia Artificial. Durante el desarrollo de este entorno controlado de pruebas, se utilizará como referencia la posición del Consejo de la Unión Europea del 25 de noviembre de 2022».¹⁴

Las principales cuestiones que vamos a abordar van a ser las siguientes:

a) Se van a analizar los distintos cambios y modificaciones operados en las sucesivas versiones propuestas después del texto inicial del año 2021, en relación con el texto definitivo aprobado en el año 2024, con la finalidad de observar cuál ha sido su propósito y finalidad por lo que respecta a los artículos 11, 17, 18 y 50 (actualmente suprimido el contenido inicial y que se ocupa actualmente de regular Obligaciones de transparencia de los proveedores y usuarios de determinados sistemas de IA).

b) Determinar cuáles son los principales motivos de los cambios realizados, así como los aspectos más relevantes de su aplicación.

c) Establecer los contextos donde los sistemas de inteligencia artificial de alto riesgo pueden operar y cómo establecer el sistema de gestión de la calidad, problemas referentes a la documentación técnica y aspectos relativos a la conservación documental.

La metodología que vamos a utilizar es realizar un análisis comparativo de la distinta normativa aplicable en materia de IA según las diferentes versiones de propuestas, así como la doctrina que se ha pronunciado sobre la materia con el

14. Señala también el Real Decreto 817/2023, «La inteligencia artificial es una tecnología disruptiva con una alta capacidad de impacto en la economía y la sociedad. En el plano económico, y junto a otras tecnologías digitales, presenta un alto potencial para el aumento de la productividad, la apertura de nuevas líneas de negocio, el desarrollo de nuevos productos o servicios —basados, por ejemplo, en la personalización, la optimización de los procesos industriales o las cadenas de valor—, la mejora en la facilidad de realización de tareas cotidianas, la automatización de ciertas tareas rutinarias y el desarrollo de la innovación. Este potencial incide positivamente en el crecimiento económico, la creación de empleo y el progreso social.

No obstante, los sistemas de inteligencia artificial también pueden suponer riesgos sobre el respeto de los derechos fundamentales de la ciudadanía, como por ejemplo los relativos a la discriminación y a la protección de datos personales, o incluso causar problemas graves sobre la salud o la seguridad de la ciudadanía».

objeto de obtener unas conclusiones válidas aplicables a la comunidad científica internacional.

II. EL ARTÍCULO 17 DEL REGLAMENTO SOBRE SISTEMA DE GESTIÓN DE LA CALIDAD

En el primer texto preparado por la Comisión Europea en el año 2021 se establecía que los proveedores de sistemas de IA de alto riesgo establecerán un sistema de gestión de calidad, y que se documentará el mismo de forma sistemática y ordenada mediante políticas, procedimientos e instrucciones por escrita que incluirán unos aspectos que se precisan en el propio precepto (técnicas, procedimientos, examen, prueba y validación, especificaciones técnicas, gestión de datos, gestión de riesgos, notificaciones, registro, rendición de cuentas, entre otros), y que será proporcional al tamaño de la organización del proveedor.

Respecto a la gestión de los datos hay que tener en cuenta lo indicado en la Propuesta de Reglamento del Parlamento Europeo y del Consejo sobre normas armonizadas para un acceso justo a los datos y su utilización (Ley de Datos), de 23 de febrero de 2022, [COM(2022) 68 final 2022/0047 (COD)]¹⁵, y la Resolución legislativa del Parlamento Europeo, de 9 de noviembre de 2023, sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo sobre normas armonizadas sobre el acceso equitativo a los datos y su uso (Ley de Datos [COM(2022)0068 — C9-0051/2022 — 2022/0047(COD)]).¹⁶

En los casos en que se tratara de proveedores que fueran entidades de créditos que se regularan por la Directiva 2013/36/UE del Parlamento Europeo y del Consejo de 26 de junio de 2013 relativa al acceso a la actividad de las entidades de crédito y a la supervisión prudencial de las entidades de crédito y las empresas de inversión, por la que se modifica la Directiva 2002/87/CE y se derogan las Directivas 2006/48/CE y 2006/49/CE, se considerarán que cumplen la obligación de establecer un sistema de gestión de calidad cuando cumplan las normas referentes a los sistemas, procedimientos y mecanismos de gobernanza a los que hace referencia el artículo 74 de la mencionada norma. En este contexto, se tendrán en cuenta todas las normas armonizadas que se indican en el artículo 40 del RIA, en el que se menciona el Reglamento (UE) núm. 1025/2012 del Parlamento Europeo y del Consejo de 25 de octubre de 2012 sobre la normalización europea, por el que se modifican las Directivas 89/686/CEE y 93/15/CEE del Consejo y las Directivas 94/9/CE, 94/25/CE, 95/16/CE, 97/23/CE, 98/34/CE, 2004/22/CE, 2007/23/CE, 2009/23/CE y 2009/105/CE del Parlamento Europeo y del Consejo y por el que se deroga la Decisión 87/95/CEE del Consejo y la Decisión n o 1673/2006/CE del Parlamento Europeo y del Consejo.

La Propuesta de RIA introduce un nuevo apartado 2 bis con la finalidad de garantizar una mayor armonización con la legislación sectorial respecto a las obligaciones relacionadas con los sistemas de gestión de la calidad.

15. Disponible en: <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52022PC0068> (Consultado el 13 de noviembre de 2023).

16. Disponible en: https://www.europarl.europa.eu/doceo/document/TA-9-2023-0385_EN.pdf (Consultado el 13 de noviembre de 2023).

Este nuevo apartado indica que para los proveedores de sistemas de inteligencia artificial de alto riesgo que estén sujetos a obligaciones relativas a los sistemas de gestión de la calidad en virtud de la legislación sectorial pertinente de la Unión, los aspectos descritos en el apartado 1 pueden formar parte de los sistemas de gestión de calidad en virtud de dicha ley.

En el apartado 3 del artículo 17 del RIA en la Propuesta de 2022, el Texto de compromiso de la Cuarta Presidencia, se realizan diversos ajustes mencionando solamente las entidades financieras sin especificar las de crédito, así como la supresión de la referencia de la Directiva 2013/36/UE, y la remisión de la excepción del apartado 1, letras g), h) e i) y la mención a la legislación de servicios financieros de la Unión pertinente.

La versión que incorpora las enmiendas aprobadas por el Parlamento Europeo es el texto denominado Ley de Inteligencia Artificial Enmiendas aprobadas por el Parlamento Europeo el 14 de junio de 2023 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión (COM(2021)0206 — C9-0146/2021 — 2021/0106(COD)).¹⁷

En la enmienda 346 sobre la propuesta de Reglamento, respecto al artículo 17, apartado 1, parte introductoria del texto de la Comisión, se establece una diferencia respecto a la redacción por cuanto los proveedores de sistemas de IA de alto riesgo ya no indica que establecerán, que implicaba una obligación, sino que se sustituye por «dispondrán» entendiéndose como una facultad de disposición del sistema de gestión de calidad. También se introduce una alternativa que no se encontraba en la redacción inicial, ya que ahora se refiere a procedimientos o instrucciones escritas, y se añade la posibilidad de introducir en un sistema de gestión de la calidad ya existente con arreglo a los actos legislativos sectoriales de la Unión. En la anterior versión del texto aprobado se indicaba que se implantarán los sistemas de gestión, con lo que se vuelve a la obligatoriedad del sistema de gestión de calidad, y ya no menciona como alternativa procedimientos o instrucciones escritas, sino tanto procedimientos como instrucciones escritas al modificarse la redacción e incluirlas a ambas. No obstante, el texto definitivo recupera la redacción inicial y se expresa de la siguiente forma: «Los proveedores de sistemas de IA de alto riesgo establecerán un sistema de gestión de la calidad que garantice el cumplimiento del presente Reglamento. Dicho sistema deberá consignarse de manera sistemática y ordenada en documentación en la que se recojan las políticas, los procedimientos y las instrucciones».

En la enmienda 347, relativa al artículo 17, apartado 1, letra a, se suprimía la indicación de incluir «a) una estrategia para el cumplimiento reglamentario, incluido

17. Disponible en: https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_ES.pdf (Consultado el 24 de julio de 2023). Se puede consultar también: BARRIO ANDRÉS, M., «Novedades en la tramitación del próximo Reglamento europeo de inteligencia artificial», *Real Instituto Elcano*, (2023), pp. 1-10. Disponible en: <https://www.realinstitutoelcano.org/analisis/novedades-en-la-tramitacion-del-proximo-reglamento-europeo-de-inteligencia-artificial/> (Consultado el 24 de julio de 2023).

el cumplimiento de los procedimientos de evaluación de la conformidad y de los procedimientos de gestión de las modificaciones de los sistemas de IA de alto riesgo».

En el texto aprobado se vuelve a incluir el apartado 1, letra a, anteriormente suprimido, y se indica que se incluirá «(a) una estrategia de cumplimiento de la normativa, incluido el cumplimiento de los procedimientos de evaluación de la conformidad y los procedimientos de gestión de las modificaciones del sistema de IA de alto riesgo». Claramente se amplía la referencia a cumplimiento de normativa y no meramente de la reglamentaria.

En la enmienda 348, respecto al artículo 17, apartado 1, letra e) que trata sobre las especificaciones técnicas, incluidas las normas, que se aplicarán y, cuando las normas armonizadas pertinentes no se apliquen en su totalidad, se añade que «no cubran todos los requisitos pertinentes», los medios que se utilizarán para velar por que el sistema de IA de alto riesgo cumpla los requisitos establecidos en el capítulo 2 del presente título.

El texto aprobado incluye lo señalado en la enmienda 348, de tal forma que se refiere a las especificaciones técnicas, incluidas las normas, que deban aplicarse y, cuando las normas armonizadas pertinentes no se apliquen en su totalidad o no cubran todos los requisitos pertinentes establecidos en el capítulo II, los medios que deban utilizarse para garantizar que el sistema de IA de alto riesgo cumple los requisitos establecidos, remitiéndose a lo mencionado en el citado capítulo. La reformulación de la redacción inicial dota al precepto de una mayor agilidad técnica y mejor comprensión, con la finalidad de evitar reiteraciones legislativas.

El artículo 17, apartado 1, letra f, se enmienda incluyendo respecto de los sistemas y procedimientos de gestión de datos, además de los mencionados como la recopilación, análisis, etiquetado, almacenamiento, filtrado, prospección, agregación, conservación y cualquier otra operación relacionada con los datos que se lleve a cabo antes de la introducción en el mercado o puesta en servicio de sistemas de IA de alto riesgo y con ese fin, la adquisición. El texto aprobado mantiene la referencia a la adquisición de datos y a la recopilación de datos que ya mencionaba la enmienda.

En la enmienda 350, referente al artículo 17, apartado 1, letra j, se elimina la referencia a autoridades nacionales competentes, indicando solamente que la gestión de la comunicación se realizará con las autoridades nacionales pertinentes, incluidas las sectoriales. Se considera más que un aspecto competencial, un aspecto de adecuación o pertinencia, y se elimina la indicación que se contenía inicialmente de que permitían acceder a datos o facilitar el acceso a ellos; y también la referencia a los organismos notificados; otros operadores; los clientes, u otras partes interesadas. En el texto aprobado se clarifica indicando autoridades nacionales, eliminando la indicación de competentes, otras autoridades competentes pertinentes, incluidas las sectoriales.

En la enmienda 351 que se presenta al artículo 17, apartado 2, respecto a la inclusión de los aspectos mencionados en el apartado 1 que será proporcional al tamaño de la organización del proveedor, se añade un párrafo indicando que «los proveedores respetarán en todo caso el grado de rigor y el nivel de protección requerido para garantizar la conformidad de sus sistemas de AI con el presente Reglamento».

Se mantiene en el texto aprobado dicha enmienda.

Respecto a la aplicación de la gestión de la calidad, la doctrina ha indicado un caso¹⁸. Sería el supuesto de la gestión de la calidad aplicada a la producción de billetes. Así, se indica que a través de la aplicación del concepto de calidad 4.0 en el que se engloban distintos aspectos en los que las tecnologías facilitadoras de la Industria 4.0 pueden mejorar los sistemas de gestión de la calidad de los productos. Las herramientas a implementar sería mejorar la conectividad, el análisis de datos, la inteligencia artificial y la automatización.

Este autor destaca varios puntos a tener en cuenta¹⁹: a) El Edge Computing y las redes IoT (Internet of Things — Internet de las cosas) proporcionan un mayor volumen de datos fiables para su análisis; b) El análisis de los datos permite tomar decisiones de acuerdo con ellos con mayor precisión y elaborar modelos que favorezcan la predicción de eventos y la previsión en la producción.; c) El 5G y otras mejoras de la conectividad favorecen la velocidad en el intercambio de información dentro y fuera del entorno; d) La colaboración y la conformidad entre los diferentes agentes ligados a la producción se consiguen mediante la intranet y la gestión documental, y se combinan con redes *blockchain* que consolidan la seguridad del intercambio de información, y e) La promoción de la cultura de la calidad a todos los estamentos de la empresa o entidad es clave para asegurar un cumplimiento por parte de todos los eslabones que intervienen en la cadena.

El artículo 17, apartado 2 bis, se introduce por Mandato del Consejo y se mantiene su misma redacción en el texto del Proyecto señalando que: «2a. Para los proveedores de sistemas de IA de alto riesgo que estén sujetos a obligaciones relativas a los sistemas de gestión de la calidad o a su función equivalente en virtud de la legislación sectorial pertinente de la Unión, los aspectos descritos en el apartado 1 podrán formar parte de los sistemas de gestión de la calidad con arreglo a dicha legislación».

Este apartado se convierte en el apartado 3 del artículo 17 con algunas modificaciones de interés en su definitiva redacción referentes a la normativa aplicable no limitándose a la legislación sino al Derecho: «Los proveedores de sistemas de IA de alto riesgo que estén sujetos a obligaciones relativas a los sistemas de gestión de la calidad o una función equivalente con arreglo al Derecho sectorial pertinente de la Unión podrán incluir los aspectos enumerados en el apartado 1 como parte de los sistemas de gestión de la calidad con arreglo a dicho Derecho».

El artículo 17, apartado 4 (antes apartado 3), mantiene la redacción el texto del Proyecto del Mandato del Consejo respecto a los proveedores que sean entidades financieras de crédito a las que se les aplica la Directiva 2013/36/UE, están obligadas a establecer una gestión de la calidad con la excepción de lo indicado en el apartado 1, g), h) e i) se considerarán cumplidas si se respetan las normas sobre disposiciones, mecanismos o procesos de gobernanza interna de conformidad con la legislación pertinente de la Unión en materia de servicios financieros. Teniendo en cuenta las normas armonizadas a las que se refiere el artículo 40.

18. Sigo la exposición de López González, A., «Inteligencia artificial aplicada al control de calidad en la producción de billetes», *Papel ocasional del Banco de España*, n.º 2303 (2023), pp. 1 y sigs. Disponible en: <https://ssrn.com/abstract=4451046> (Consultado el 11 de noviembre de 2023).

19. *Ibidem*, pp. 12 y 13.

El artículo 16 referente a las obligaciones de los proveedores de sistemas de IA de alto riesgo se remite al artículo 17, ya que estos proveedores deberán disponer de un sistema de gestión de la calidad ajustado a lo indicado en el precepto citado.

El artículo 63 sobre las excepciones para operadores específicos y respecto de las microempresas definidas como indica la Recomendación 2003/361/CE de la Comisión siempre que no tengan empresas asociadas o vinculadas podrán cumplir determinados elementos del sistema de gestión de la calidad que exige el artículo 17 del RIA de una forma simplificada. Para ello la Comisión elaborará las directrices sobre los elementos del sistema que se puedan cumplir de dicha forma sin que ello afecte al nivel de protección y a la necesidad de cumplir los requisitos de los sistemas de inteligencia artificial de alto riesgo.

El anexo VI se remite al artículo 17 respecto al procedimiento de evaluación de la conformidad basado en el control interno, ya que el proveedor verificará que el sistema de gestión de la calidad establecido cumple los requisitos del citado precepto. También el Anexo VII sobre conformidad basada en la evaluación del sistema de gestión de calidad y la evaluación de la documentación técnica en el que se refiere el artículo 17. La solicitud del proveedor deberá incluir la documentación referente al sistema de gestión de calidad que abarcará todos los aspectos indicados en el artículo 17. Este sistema de gestión de la calidad se evaluará por el organismo notificado que determinará si cumple con lo indicado en el artículo 17.

III. EL ARTÍCULO 11 DEL REGLAMENTO SOBRE DOCUMENTACIÓN TÉCNICA CON ANEXOS

La Propuesta de Reglamento del Parlamento Europeo y del Consejo de 2022 incorporaba algunas diferencias con el texto inicial de 2021 ya que se establecía respecto a la documentación técnica en los sistemas de inteligencia artificial de alto riesgo que se facilite toda la información a las autoridades nacionales competentes y a los organismos notificados de forma clara y completa, que no se contenía en la redacción de 2021. Se incluye a las pyme y se hace mención a empresas de nueva creación, que no se especificaba en la redacción de 2021, y en este caso se contendría, como mínimo, cualquier documentación equivalente a los elementos que establece el anexo IV que cumpla los mismos objetivos, a menos que se considere inadecuada, y se elimina la indicación de que sea previa aprobación de la autoridad competente.

La versión de junio de 2023 de Ley de Inteligencia artificial incorpora la enmienda 292 respecto al artículo 11, apartado 1, párrafo 1, que en su versión inicial disponía: «La documentación técnica se redactará de modo que demuestre que el sistema de IA de alto riesgo cumple los requisitos establecidos en el presente capítulo y proporcionará a las autoridades nacionales competentes y los organismos notificados *toda* la información que necesiten para evaluar si el sistema de IA de que se trate cumple dichos requisitos. Contendrá, como mínimo, los elementos contemplados en el anexo IV» y que se sustituye por «la documentación técnica se redactará de modo que demuestre que el sistema de IA de alto riesgo cumple los requisitos establecidos en el presente capítulo y proporcionará a las autoridades nacionales de supervisión y los organismos notificados la información que necesiten para evaluar si el sistema de IA de que se trate cumple dichos requisitos. Contendrá, como mínimo, los elementos contemplados en el anexo IV o, en el caso de las pymes y las empresas emergentes,

cualquier documentación equivalente que cumpla los mismos objetivos, previa aprobación de la autoridad nacional competente».

Destaca en esta enmienda la referencia a las pequeñas y medianas empresas y a las empresas emergentes por lo que habrá que atender a dicha consideración como tales a lo indicado en el Reglamento (UE) núm. 651/2014 de la Comisión, de 17 de junio de 2014, por el que se declaran determinadas categorías de ayudas compatibles con el mercado interior en aplicación de los artículos 107 y 108 del Tratado que considera a la empresa según el artículo 1 del anexo I, a «toda entidad», independientemente de su forma jurídica, que ejerza una actividad económica. En particular, se considerarán empresas las entidades que ejerzan una actividad artesanal u otras actividades a título individual o familiar, así como las sociedades de personas y las asociaciones que ejerzan una actividad económica de forma regular, respecto a qué se considera como pyme, el artículo 2 del anexo I incluye a las pequeñas empresas que ocupa a menos de 50 personas y cuyo volumen de negocios anual o cuyo balance anual no es superior a los 10 millones de euros.

Del mismo modo, la mención específica de las pymes en la enmienda puede tener su fundamento en el Reglamento (UE) 2021/694 que, en su artículo 5, centrado en el objetivo de la inteligencia artificial, se persigue como objetivo operativo hacer accesibles las capacidades de desarrollo y refuerzo y los conocimientos básicos de inteligencia artificial a las empresas, y en especial a las pymes y las empresas emergentes.

En el texto aprobado se mantiene la redacción del Mandato del Consejo, y se añade en el caso de las PYME, incluidas las de nueva creación utilizando el texto la denominación de «emergentes» que podrán facilitar los elementos de la documentación técnica que se especifican en el anexo IV de forma simplificada. Para ello, la Comisión establecerá un formulario simplificado de documentación técnica orientado a las necesidades de las pequeñas empresas y microempresas. En el caso de que una PYME ya creada o de nueva creación, en términos del texto «emergente», opte por facilitar la información que se requiere en el anexo IV de manera simplificada utilizará el formulario mencionado en el precepto. Los organismos notificados aceptarán el formulario a efectos de evaluación de la conformidad.

En el texto también observamos distintas referencias al artículo 11 como es el caso del artículo 22 que se refiere Representantes autorizados de los proveedores de sistemas de IA de alto riesgo en las que antes de que se comercialice un sistema de inteligencia artificial de alto riesgo, aquéllos se asegurarán que es conforme al Reglamento y verificarán verificar que se han elaborado la declaración UE de conformidad y la documentación técnica a que se refiere el artículo 11 y que el proveedor ha llevado a cabo un procedimiento de evaluación de la conformidad adecuado. Se elimina la referencia al anexo IV que se refiere a la documentación técnica que indica el artículo 11 del RIA, en su apartado 1, y contendrá unos mínimos relativos a la información sobre el sistema de inteligencia artificial pertinente tales como al descripción detallada de los elementos y del proceso para su desarrollo, así como información detallada sobre la supervisión, funcionamiento y control del sistema de inteligencia artificial. Dentro de cada uno de estos bloque se contiene una especificación detallada sobre la versión del sistema, software, interfaz, la lógica del

sistema y los algoritmos, los requisitos de datos en fichas técnicas, procedimientos de validación y ensayo, medidas de ciberseguridad, entre otros.

También se refiere al artículo 11 el artículo 97 sobre el ejercicio de la delegación respecto a la delegación de poderes dentro del Capítulo XI referente a la delegación de poderes y procedimiento de comité.

También hay que atender a lo indicado en nuestro país en la Ley 29/2022, de 21 de diciembre, de fomento del ecosistema de las empresas emergentes y a la Ley 18/2022, de 28 de septiembre, de creación y crecimiento de empresas, así como en el Decreto-ley 2/2023, de 8 de marzo, de medidas urgentes de impulso a la inteligencia artificial en Extremadura y el Real Decreto 729/2023, de 22 de agosto, por el que se aprueba el Estatuto de la Agencia Española de Supervisión de Inteligencia Artificial.

Todo ello de conformidad con el documento elaborado por el Gobierno de España *España Digital 2026*²⁰ cuyo propósito para dicha fecha es la aceleración de la digitalización de las empresas atendiendo principalmente a las pymes y start-ups y crear las condiciones favorables para el surgimiento y maduración de empresas emergentes de base tecnológica, siguiendo las indicaciones de la Estrategia Nacional de Inteligencia Artificial (ENIA)²¹ que es uno de los componentes del Plan de Recuperación, Transformación y Resiliencia.²² Se inserta en el eje 6 de la Estrategia que corresponde al Componente 16 en el Plan de Recuperación, y que tiene como uno de los objetivos el apoyo del despliegue y uso masivo de la inteligencia artificial por parte de las grandes empresas, las Administraciones Públicas, las pequeñas y medianas empresas y empresas emergentes y la sociedad civil.

Cabe mencionar el indicado Real Decreto 817/2023 que dicta en conformidad con lo indicado en la Ley 28/2022, de 21 de diciembre, de fomento del ecosistema de las empresas emergentes²³, en su artículo 16, se contemplan la creación de entornos

20. Disponible en: https://espanadigital.gob.es/sites/espanadigital/files/2022-07/EspanaDigital_2026.pdf (Consultado el 24 de julio de 2023). Anteriormente cabe señalar el documento también elaborado por el Gobierno de España denominado *España Digital 2025*. Disponible en: https://avancedigital.mineco.gob.es/programas-avance-digital/Documents/EspanaDigital_2025_TransicionDigital.pdf (Consultado el 6 de noviembre de 2023), que ya mencionaba la aceleración de la digitalización de las empresas, con especial atención a las microPYMEs, así como la Estrategia de Agenda Digital 2021-2027 (<https://www.europarl.europa.eu/factsheets/es/sheet/64/una-agenda-digital-para-europa>, consultado el 6 de noviembre de 2023), donde se aborda la conectividad, las infraestructuras de la tecnología, el talento digital y la economía digital. Véase también la *Carta de Derechos Digitales*, 2021. Disponible en: https://www.lamoncloa.gob.es/presidente/actividades/Documents/2021/140721-Carta_Derechos_Digitales_RedEs.pdf (Consultado el 6 de noviembre de 2023).
21. Disponible en: <https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/ENIA2B.pdf> (Consultado el 24 de julio de 2023).
22. Disponible en: https://portal.mineco.gob.es/es-es/ministerio/plan_recuperacion/Documents/Plan-de-Recuperacion-Transformacion-Resiliencia.pdf (Consultado el 24 de julio de 2023).
23. También hay que indicar la Orden PCM/825/2023, de 20 de julio, por la que se regulan los criterios y el procedimiento de certificación de empresas emergentes que dan acceso a los beneficios y especialidades reconocidas en la Ley 28/2022, de 21 de diciembre, de fomento del ecosistema de las empresas emergentes (BOE n.º 173, de 21 de julio de 2023).

controlados, por periodos limitados de tiempo con la finalidad de evaluar la utilidad, la viabilidad y el impacto de innovaciones tecnológicas aplicadas a actividades reguladas, a la oferta o provisión de nuevos bienes o servicios, a nuevas formas de provisión o prestación de los mismos o a fórmulas alternativas para su supervisión o control por parte de las autoridades competentes. Se indica la creación de los entornos controlados de pruebas para la evaluación de su impacto que se justifica por razones imperiosas de interés general.

Este Real Decreto 817/2023, como indica el artículo 1, «tiene por objeto establecer un entorno controlado de pruebas para ensayar el cumplimiento de ciertos requisitos por parte de algunos sistemas de inteligencia artificial que puedan suponer riesgos para la seguridad, la salud y los derechos fundamentales de las personas. Asimismo, se regula el procedimiento de selección de los sistemas y entidades que participarán en el entorno controlado de pruebas».

Menciona la gestión de la calidad en la referencia a la «autoevaluación de cumplimiento», en el artículo 3, considerado como el procedimiento de verificación del cumplimiento de los requisitos, del sistema de gestión de la calidad, de la documentación técnica y del plan de seguimiento posterior a la comercialización. En el artículo 13 se indica que tanto el proveedor IA participante como, en su caso, el usuario participante deberá realizar las siguientes acciones para completar la autoevaluación siendo una de estas acciones la comprobación de que el sistema de gestión de la calidad es acorde con las especificaciones que ofrezca el órgano competente. El órgano competente examinará los documentos asociados a la declaración de cumplimiento presentada por el proveedor IA, principalmente los que describen el sistema de gestión de la calidad, la documentación técnica o el plan de seguimiento posterior a la comercialización, como también dispone el artículo 13 del Real Decreto 817/2023.

Se introdujo en el artículo 11, un apartado 3 bis (nuevo) en la enmienda 294 respecto del texto de la Comisión. Este nuevo apartado dispone que «En el caso de los proveedores que sean entidades de crédito reguladas por la Directiva 2013/36/UE, la documentación técnica formará parte de la documentación relativa a los sistemas, procedimientos y mecanismos de gobierno interno que figuran en el artículo 74 de dicha Directiva».

Se recoge en este nuevo apartado lo que se indicaba en el artículo 18, apartado 2, que es suprimido por la enmienda 354.

Dicha incorporación ya no se encuentra en el texto definitivo del artículo 11. El Considerando 158 del texto definitivo especifica lo siguiente: «Con vistas a aumentar la coherencia entre el presente Reglamento y las normas aplicables a las entidades de crédito reguladas por la Directiva 2013/36/UE, conviene igualmente integrar algunas de las obligaciones procedimentales de los proveedores relativas a la gestión de riesgos, la vigilancia poscomercialización y la documentación en las obligaciones y los procedimientos vigentes con arreglo a la Directiva 2013/36/UE. Para evitar solapamientos, también se deben contemplar excepciones limitadas en relación con el sistema de gestión de la calidad de los proveedores y la obligación de vigilancia impuesta a los responsables del despliegue de sistemas de IA de alto riesgo, en la medida en que estos se apliquen a las entidades de crédito reguladas por la Directiva 2013/36/UE».

Se refiere a los sistemas, procedimientos y mecanismos de las entidades centrándose en el gobierno interno y planes de rescate y resolución haciendo mención de la autoridad bancaria europea (ABE).

En la enmienda 293 respecto al artículo 11, apartado 2, se cambia la redacción respecto a que cuando se introduzca en el mercado o se ponga en servicio un sistema de IA de alto riesgo asociado a un producto al que se apliquen los actos legislativos mencionados en el anexo II, sección A, se elaborará una única documentación técnica que contenga toda la información estipulada en el anexo 1, en vez del anexo IV a que hacía referencia la redacción inicial, así como la información que exijan dichos actos legislativos.

El Real Decreto 817/2023, en su anexo VI menciona la documentación técnica a presentar a la finalización de la implantación de los requisitos, al que cabe remitir. Toda la información proporcionada será tratada con la debida confidencialidad en concordancia con el artículo 18 del presente real decreto.

Como indica el artículo 11 del Real Decreto 817/2023, esta documentación técnica del sistema de inteligencia artificial que se indica en el anexo VI se elaborará conforme a las especificaciones que facilitará el órgano competente, y dicha documentación deberá actualizarse a lo largo de la duración del entorno controlado de pruebas.

Según el artículo 13 del Real Decreto 817/2023, tanto el proveedor IA participante como, en su caso, el usuario participante deberá realizar las siguientes acciones para completar la autoevaluación, y entre ellas se indica la verificación de que el diseño y desarrollo del proceso del sistema de inteligencia artificial y su seguimiento posterior a la comercialización son coherentes con lo establecido en la documentación técnica y con las especificaciones ofrecidas por el órgano competente, y verificará también que la documentación técnica de su sistema de inteligencia artificial incluye el contenido según especificaciones del anexo VI del mencionado Real Decreto, y además de la documentación de comprobación del cumplimiento de los puntos anteriores que se indican en el artículo 13 indicado.

Según lo indicado en el artículo 14 del Real Decreto 817/2023, respecto del seguimiento posterior a la comercialización, se basará en un plan de monitorización posterior a la comercialización que se incluirá en la documentación técnica a aportar que se contempla en el anexo VI del presente real decreto. Para su redacción se seguirán las especificaciones que el órgano competente proporcione a tal efecto.

El artículo 21 del Real Decreto 817/2023 respecto a la obtención de información sobre el desarrollo del entorno indica que Durante el transcurso del entorno controlado de pruebas, Subdirección General de Inteligencia Artificial y Tecnologías Habilitadoras Digitales recabará información tanto de los proveedores IA participantes como de los usuarios participantes sobre la manera en que se han implementado las actuaciones pertinentes en cada sistema de inteligencia artificial; la forma en que se ha efectuado la autoevaluación de cumplimiento; la documentación técnica asociada a cada sistema de inteligencia artificial; y sobre los sistemas de gestión de la calidad o del riesgo descritos en los anexos o guías.

IV. EL ARTÍCULO 18 DEL REGLAMENTO SOBRE CONSERVACIÓN DE LA DOCUMENTACIÓN

La Propuesta de Reglamento del Parlamento Europeo y del Consejo de 2022 deja sin contenido el artículo 18 en su redacción inicial sobre la obligación de elaborar la documentación técnica y pasa a recoger el contenido del artículo 50 sobre conservación de los documentos, que se denomina como «Conservación de la documentación». Y actualiza el apartado 2 del artículo 18 de acuerdo con los cambios introducidos en el tercer texto de compromiso en relación con las instituciones financieras y se reflejan dichos cambios en el apartado 2 del artículo 20.

El artículo 18 que en su redacción por la Propuesta de Reglamento, en su apartado 1, indicaba que los proveedores de sistemas de IA de alto riesgo elaborarán la documentación técnica mencionada en el artículo 11 con arreglo al anexo IV, es suprimido por la enmienda 353, y también se suprime el apartado 2 del artículo 18, por la enmienda 354, cuya redacción establecía que en el caso de los proveedores que sean entidades de crédito reguladas por la Directiva 2013/36/UE, la documentación técnica formará parte de la documentación relativa a los sistemas, procedimientos y mecanismos de gobernanza interna que figuran en el artículo 74 de dicha Directiva.

El texto aprobado mantiene la redacción del Mandato del Consejo.

El artículo 16 respecto de las obligaciones de los proveedores de sistemas de inteligencia artificial de alto riesgo menciona como una de ellas la conservación de la documentación a la que hace referencia el artículo 18.

V. EL INICIAL ARTÍCULO 50 DEL REGLAMENTO SOBRE CONSERVACIÓN DE LOS DOCUMENTOS

El artículo 50²⁴ referente a «Conservación de los documentos» en la Propuesta de Reglamento de 2022 es eliminado su contenido y pasa a reproducirse el mismo en el artículo 18, que queda sin el contenido inicial referente a la obligación de elaborar documentación técnica, y pasa a denominarse «Conservación de la documentación».²⁵

24. Este precepto en la redacción de la Propuesta de Reglamento disponía que:
- «Durante un período que finalizará diez años después de la introducción del sistema de IA en el mercado o su puesta en servicio, el proveedor mantendrá a disposición de las autoridades nacionales competentes:
 - a) la documentación técnica a que se refiere el artículo 11;
 - b) la documentación relativa al sistema de gestión de la calidad a que se refiere el artículo 17;
 - c) la documentación relativa a los cambios aprobados por los organismos notificados, si procede;
 - d) las decisiones y otros documentos expedidos por los organismos notificados, si procede;
 - e) la declaración UE de conformidad contemplada en el artículo 48».
25. Se puede consultar: Zapata Cárdenas, C. A. y Giménez Chornet, V., «Retos de los archivos ante los derechos digitales», *Los nuevos retos de los Derechos Digitales*, Ramón Fernández, F. (coord.), Tirant lo Blanch, Valencia, 2022, pp. 313 y sigs.; y Giménez Chornet, V., «La problemática de la inteligencia artificial en la gestión documental archivística», *Ciencia de Datos y Perspectivas de la Inteligencia Artificial*, Ramón Fernández, F. (Coord.), Tirant lo Blanch, Valencia, 2024, pp. 181 y sigs.

En la enmienda 477 sobre la propuesta de Reglamento se añadía al artículo 50, párrafo primero, la indicación de la autoridad nacional de supervisión, además de a las autoridades nacionales competentes durante un período que finalizará diez años después de introducir el sistema de inteligencia artificial en el mercado o su puesta en servicio.

En el texto aprobado se mantiene la eliminación del precepto, aunque se recupera la numeración del precepto dentro del Capítulo IV «Obligaciones de transparencia de los proveedores y responsables del despliegue de determinados sistemas de IA» bajo la denominación «Obligaciones de transparencia de los proveedores y usuarios de determinados sistemas de IA».

VI. CONCLUSIONES

En el presente trabajo hemos analizado los sistemas de gestión de calidad, documentación técnica y conservación, en particular los artículos 17, 11, 18 y 50 de la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión, y en comparación con el texto definitivo de la Resolución legislativa del Parlamento Europeo, de 13 de marzo de 2024, señalando los cambios más relevantes que se han operado en toda la trayectoria legislativa. Se aprecian cambios en la numeración del articulado y también en la terminología utilizada siendo de destacar la mención a «responsables del despliegue» cuando antes se refería a «usuarios», y la estructura en capítulos y secciones (antes a título y capítulo), entre otros.

A través del análisis del texto de la propuesta de la Comisión y de las enmiendas acordadas por el Parlamento, en las que se realizan algunos cambios en la redacción inicial, hemos observado algunos aspectos que pueden resultar de interés. Además, hay que tener en cuenta la reciente publicación del Real Decreto 817/2023, que nos aporta una valiosa perspectiva en relación al sistema de gestión de la calidad y de la documentación técnica. El entorno controlado de pruebas, tal y como indica el citado texto legal, posibilita la cooperación entre los usuarios y los proveedores de inteligencia artificial, validando desde ambos aspectos la implementación de los requisitos tanto de sistemas de inteligencia artificial de alto riesgo, como de los sistemas de propósito general y los modelos fundacionales en relación al cumplimiento de los requisitos comunitarios.

En el caso del artículo 11, relativo a los requisitos para los sistemas de AI de alto riesgo, enfocado a la documentación técnica, se pretende que la misma esté disponible antes de que el sistema de IA de alto riesgo se comercialice o se ponga en servicio. La idea es garantizar la máxima seguridad del sistema y la conformidad con el sistema con los requisitos que se exigen, y destaca la referencia a las PYMEs y a las empresas emergentes siendo de aplicación la normativa que hemos mencionado en el ámbito comunitario.

Una de las modificaciones de interés es la que se incorpora en relación con las PYMEs respecto a la documentación técnica pudiendo hacer uso del formulario simplificado que establecerá la Comisión para facilitar dicha labora a las pequeñas y microempresas.

El artículo 17 del texto definitivo se focaliza en el sistema de gestión de la calidad y la adopción del mismo por parte de los proveedores de sistemas de IA, y que según lo indicado en el Real Decreto 817/2023 es toda persona jurídica privada, entidad del sector público en España, u organismo de otra índole, que ha desarrollado o para quien se ha desarrollado un sistema de inteligencia artificial, y que lo introduce en el mercado o lo pone en servicio bajo su propio nombre o marca comercial, ya sea de forma onerosa o gratuita. El proveedor AI será designado de las formas indicadas a continuación según la fase del proceso en la que se encuentre.

Destaca la referencia a las entidades financieras y la indicación de que los proveedores respetarán en todo caso el grado de rigor y el nivel de protección requerido para garantizar la conformidad de sus sistemas de AI con el Reglamento.

A ello hay que tener en cuenta en relación al acceso y uso de los datos, la reciente Resolución legislativa del Parlamento Europeo, de 9 de noviembre de 2023, a la que hemos hecho mención en el presente estudio.

El artículo 18 del texto final se relaciona con lo indicado con los artículos 11 y 17 del mismo texto legal y se refiere a la conservación de la documentación durante el plazo fijado por la norma. Esta preservación de la documentación de cualquier producto supone un paso más en la trazabilidad y seguridad para el consumidor.

En las enmiendas al artículo 18 de la Propuesta de Reglamento se suprimen los apartados 1 y 2 del citado precepto.

El artículo 50 de la Propuesta de Reglamento que se refería a la conservación de documentos queda sin contenido, aunque el precepto existe con dicha numeración pero referido a «Obligaciones de transparencia de los proveedores y usuarios de determinados sistemas de IA», y pasa el contenido inicial al actual artículo 18 anteriormente mencionado.

La obligación de conservar registros de los sistemas de alto riesgo en el Reglamento de Inteligencia Artificial

WILMA ARELLANO TOLEDO¹

Doctora por la Universidad Complutense de Madrid. OdiseIA

ANTONIO MERCHÁN MURILLO

Doctor. Abogado. OdiseIA. Profesor ayudante doctor (acreditado a Prof. Titular) en la Universidad de Cádiz²

I. INTRODUCCIÓN

Como es bien sabido,³ el Reglamento Europeo de Inteligencia Artificial clasifica a los sistemas de IA en varias categorías (*vid. supra*) y una de éstas es la de los sistemas considerados como de alto riesgo, que son aquellos que entre otras cosas «debe[n] someterse a una evaluación de la conformidad realizada por un organismo independiente para su introducción en el mercado o puesta en servicio con arreglo a los actos legislativos de armonización de la Unión enumerados en el anexo I» (antes en el II) del propio Reglamento y los contemplados en el Anexo III.

1. Doctora por la Universidad Complutense de Madrid. OdiseIA.
2. Doctor Universidad Pablo de Olavide. Abogado. OdiseIA. Co-editor de *INTELETTICA*. Profesor ayudante doctor (acreditado a Prof. Titular) en la Universidad de Cádiz. Miembro del equipo de investigación del proyecto «Hacia una transición digital centrada en la persona en la Unión Europea».
3. Artículo realizado con la financiación del Ministerio de Universidades y los Fondos *NextGenerationEU* de la Unión Europea a través del programa María Zambrano. También realizado en el marco del Proyecto «Derechos y garantías públicas frente a las decisiones automatizadas y el sesgo y discriminación algorítmicas» 2023-2025 (PID2022-136439OB-I00) financiado por MCIN/AEI/10.13039/501100011033/FEDER, UE y del Proyecto de I+D+i «Derechos y garantías públicas frente a las decisiones automatizadas y el sesgo y discriminación algorítmicas» 2023-2025 (PID2022-136439OB-I00) Ministerio de Ciencia e Innovación (MICINN), Proyectos de Generación de Conocimiento 2022 (modalidad orientada) y del proyecto «Hacia una transición digital centrada en la persona en la Unión Europea», TED2021-129307A-I00, financiado por MCIN/AEI/10.13039/501100011033 y por la Unión Europea *NextGenerationEU/PRTR*.

Es importante destacar que debe comprobarse que todos los intervinientes en el ciclo de vida del sistema de IA de alto riesgo asumen la parte de obligaciones que les corresponden en materia de registros, ya que esta imposición puede ser aplicable a los desarrolladores, fabricantes, proveedores, importadores, responsables del despliegue (antes denominados usuarios), involucrados en el seguimiento postcomercialización, etcétera⁴ (*vid. supra e infra*). Esto se explica porque la fortaleza de un registro o *log*⁵ está, entre otras cosas, en la posibilidad de probar que no han sido alterados por entes no autorizados y así convertirse evidencia. El objetivo es proteger la integridad de la información, puesto que ello es sumamente relevante para que los sistemas de alto riesgo cumplan con sus obligaciones en materia de registros.

En adelante, abordaremos los Considerandos y el articulado referente a las obligaciones en materia de registro de los sistemas de IA de los que hablaremos aquí.

II. ALGUNAS NOCIONES PREVIAS RELACIONADAS CON LAS OBLIGACIONES EN MATERIA DE REGISTROS

El RIA que estamos comentando, contiene una numerosa relación de Considerandos y en muchos de ellos encontramos elementos que pueden estar vinculados con la obligación de los sistemas de alto riesgo. El primero de ellos es el Considerando 46 (antes en el 27) que explica que los sistemas de IA de los que estamos hablando sólo pueden comercializarse o ponerse en servicio si cumplen con determinados requisitos, entre los cuales está el de los registros. El objetivo es el de proteger de dos esferas: la pública y la privada. La primera está concebida para evitar «riesgos inaceptables para intereses públicos importantes de la UE, reconocidos y protegidos por el Derecho de la Unión».

En la última versión, de mayo de 2024, el Considerando 47 hacía referencia puntual a los «efectos adversos» que puedan tener para la seguridad y salud de las personas los componentes de seguridad de los sistemas de IA.

La dimensión privada se expone ahora en el considerando 66, cuando se dispone que las obligaciones de los sistemas de IA de alto riesgo (en especial la de conservación de registros) tienen que ser cumplidas porque dichos sistemas pueden tener un efecto lesivo en la salud, la seguridad y los derechos fundamentales. De igual manera, la protección de la esfera privada ante el uso de sistemas de Inteligencia Artificial de alto riesgo aparece en el Considerando 46 (antes 43) y se especifica que ésta y otras obligaciones también persiguen evitar los riesgos para los tres aspectos antedichos (salud, seguridad y garantías individuales).

De este modo, tanto en la protección de la esfera pública como en la protección de la dimensión privada, los registros tienen y tendrán un importante papel, para verificar e incluso servir como medios de prueba y para distintos casos de uso. Desde luego también para aquellas circunstancias en las que un sistema de alto riesgo pueda llegar a dañar a persona o patrimonio, por poner un ejemplo.

4. Se puede acudir al capítulo que aborda el artículo 3 del Reglamento, referente a las definiciones sobre quiénes son cada uno de estos intervinientes.

5. No deben confundirse con el mencionado en el Considerando 131 (antes en el 69) referente a que los proveedores de sistemas de alto riesgo deben formar parte de un registro (que líneas después llama base de datos) que gestionará la Comisión.

Por otro lado, tenemos el Considerando 71 (antes en el 46) que plantea que todas las obligaciones, incluida la de generar registros, tienen que darse en toda la vida útil (antes llamado ciclo de vida) del desarrollo de Inteligencia Artificial del que se trate. Y esto es de vital importancia, pues se establece que para cumplir con los objetivos de trazabilidad de los sistemas de IA se debe «disponer de información comprensible sobre el modo en que se han desarrollado y sobre su funcionamiento durante toda su vida útil» por lo que «es preciso llevar registros y disponer de documentación técnica que contenga la información necesaria para evaluar si el sistema de IA en cuestión cumple los requisitos pertinentes y facilitar la vigilancia poscomercialización». Es decir, el ciclo de vida del sistema entendido desde su concepción y hasta el momento de ser vigilado e inspeccionado en su etapa ulterior a la comercialización (en un capítulo de esta obra se aborda este tema muy puntualmente).

Por lo tanto, como puede observarse, se puede deducir una referencia indirecta a diferentes intervinientes en los procesos de todo el ciclo de vida de los sistemas de Inteligencia Artificial de alto riesgo, lo cual les obligaría a adoptar las medidas impuestas, entre las que se encuentra la de guardar los *logs* generados automáticamente por dichos sistemas. Al hacer referencia al ciclo de vida completo, los intervinientes pueden ser numerosos⁶ y el conjunto de *logs*, también.

En las versiones de 2021 y de 2023 del que fuera el Considerando 46 se exponía que debían conservarse los «registros generados automáticamente por el sistema de IA de alto riesgo, incluidos, por ejemplo, los datos de salida, la fecha y la hora de inicio, etcétera, en la medida en que dicho sistema y los correspondientes registros estén bajo su control, durante un período adecuado que les permita cumplir sus obligaciones».

En las últimas versiones las de 2024, el Considerando modifica, ampliándola, la relación de informaciones que deben constar en la documentación técnica (lo

6. Y no solamente pueden ser diversos los intervinientes, sino que también pueden adoptar distintos papeles según de qué situaciones se trate, puesto que como reza el Considerando 84 (en las versiones anteriores en el 57): «Para garantizar la seguridad jurídica, es necesario aclarar que, en determinadas condiciones específicas, debe considerarse proveedor de un sistema de IA de alto riesgo a cualquier distribuidor, importador, responsable del despliegue u otro tercero que, por tanto, debe asumir todas las obligaciones pertinentes. Este sería el caso si, por ejemplo, esa persona pone su nombre o marca en un sistema de IA de alto riesgo ya introducido en el mercado o puesto en servicio, sin perjuicio de los acuerdos contractuales que estipulen otra distribución de las obligaciones. Este también sería el caso si dicha parte modifica sustancialmente un sistema de IA de alto riesgo que ya se haya introducido en el mercado o puesto en servicio de tal manera que el sistema modificado siga siendo un sistema de IA de alto riesgo de conformidad con el presente Reglamento, o si modifica la finalidad prevista de un sistema de IA, como un sistema de IA de uso general, que ya se haya introducido en el mercado o puesto en servicio y que no esté clasificado como sistema de alto riesgo, de tal manera que el sistema modificado pase a ser un sistema de IA de alto riesgo de conformidad con el presente Reglamento». A lo largo del Reglamento aparecen las figuras de proveedor, responsable del despliegue, representante autorizado, importador, distribuidor y operador (que puede ser el fabricante del producto, responsable del despliegue, representante autorizado, importador o distribuidor, según lo que dispone el artículo 3.8 del Reglamento. Si tomamos en cuenta que las obligaciones de registro deben ser adoptadas en todo el ciclo de vida del sistema de alto riesgo, claramente puede interpretarse que sería aplicable a todos estos actores dicha obligación.

cual también atañe a los registros). Sin embargo, en la versión votada en marzo de 2024 (se mantiene en la de mayo) el Considerando se convierte en el 71 y especifica que dicha información debe «incluir las características generales, las capacidades y las limitaciones del sistema y los algoritmos, datos y procesos de entrenamiento, prueba y validación empleados, así como documentación sobre el sistema de gestión de riesgos pertinente, elaborada de manera clara y completa».

Pero, además, se pone énfasis en que los sistemas de IA de alto riesgo deben «permitir técnicamente el registro automático de eventos, *mediante archivos de registro*, durante toda la vida útil del sistema» (las palabras en negritas fueron añadidas en el reenumerado Considerando (ahora 71) en la versión votada en marzo; acentuando así la importancia de los registros, pero omitiendo la referencia al período de tiempo que aparecía en las primeras versiones del RIA (aunque tampoco estaba estipulado en plazos concretos, pero sí se hacía alusión a éste como el período de tiempo «adecuado»).

Aunado a esto, el nuevo texto de marzo y mayo de 2024 integra en el Considerando 73 unas disposiciones que no aparecían previamente de este modo concreto, disponiendo que «Los sistemas de IA de alto riesgo deben diseñarse y desarrollarse de tal modo que las personas físicas puedan supervisar su funcionamiento, así como asegurarse de que se usan según lo previsto y de que sus repercusiones se abordan a lo largo del ciclo de vida del sistema». Es oportuno mencionar que se refiere a las personas físicas, no a los usuarios (como anteriormente se les denominaba en versiones previas) ni a los responsables del despliegue (como se les llama ahora), por lo que se refiere a otra figura⁷. Es decir, con esto se establece un nivel de transparencia y de supervisión humana que deberían cumplir los sistemas de alto riesgo que no estaba especificada de este modo en los otros textos del Reglamento.

Además, en este nuevo Considerando 73 se agrega un amplio párrafo en donde se recogen otros elementos que implican a los registros o *logs*. Destaca lo siguiente: se menciona que las personas físicas que deban encargarse de la supervisión humana han de inscribir a su vez las verificaciones que hagan cada una de ellas por separado en los registros generados por el sistema, esto es, precisamente los que nos ocupan en este trabajo.

Complementando lo que disponen los otros Considerandos, en el 82 (en las versiones anteriores aparecía en el 56 o en el 56 bis) se habla de los casos en que los proveedores de sistemas de IA de alto riesgo estén fuera del territorio de la Unión, en cuyo caso deberán actuar a través del representante autorizado, con los consecuentes problemas técnicos y prácticos que esto pueda plantear para el resguardo y el «en

7. Según se lee en la reciente redacción del Considerando, se refiere por un lado a la cuestión de la supervisión humana, pero en otras partes se refiere a las personas (sin el adjetivo de «físicas»), por lo que se entiende que se enfoca en los individuos, sujetos de derechos fundamentales (por ejemplo, cuando se puntualiza sobre «las enormes consecuencias para las personas en caso de una correspondencia incorrecta efectuada por determinados sistemas de identificación biométrica»). En el concepto de personas físicas, se centra en aquellos a los que «se haya asignado la supervisión humana para que tomen decisiones con conocimiento de causa acerca de si intervenir, cuándo hacerlo y de qué manera, a fin de evitar consecuencias negativas o riesgos, o de detener el sistema si no funciona según lo previsto».

manos de quién/ quiénes» se encuentren los registros. El problema radicaba, a nuestro entender, en que en la versión de 2021 se le confería al representante autorizado la categoría de «responsable solidario con el proveedor» cuando un producto fuese defectuoso (sin menoscabo de la aplicabilidad de la normativa en materia productos defectuosos de la propia UE)⁸. Desde el punto de vista técnico (y jurídico), esto planteaba retos importantes puesto que la carga que se imponía al representante autorizado podía interpretarse como excesiva, como así lo hicimos constar en diversos foros.

En la versión del RIA de 2024, el representante autorizado ya no aparece como responsable solidario (como si lo hacía en los textos previos hasta antes de la votación de diciembre de 2023), sino solamente como persona de contacto en el territorio comunitario, a efectos prácticamente de notificación para los proveedores de sistemas de alto riesgo con sede fuera de la Unión Europea. Eso sí, el representante autorizado deberá ser designado mediante mandato escrito y se prevé en el actual Considerando 82 del Reglamento que ejercerá un cometido primordial al momento de asegurar la conformidad de los sistemas de IA de alto riesgo comercializados o puestos en servicio en la Unión. De este modo, la obligación de generar y guardar registros no aparece aquí para ser aplicable al representante autorizado, como sí podría desprenderse de la interpretación de la redacción anterior del Considerando, en donde, siendo (como era) responsable solidario, la conservación de los registros sería fundamental.

No obstante, de conformidad con el artículo 3.5 de la última versión del Reglamento, los representantes autorizados son aquella «persona física o jurídica ubicada [“situada” en la versión anterior] o establecida en la Unión que haya recibido y aceptado el mandato por escrito [en la versión anterior sólo “escrito” y no “por escrito”] de un proveedor de un sistema de IA o de un modelo de IA de uso general [de “propósito general” en la versión previa] para cumplir las obligaciones y llevar a cabo los procedimientos establecidos en el presente Reglamento en representación de dicho proveedor». En la versión anterior a esta de mayo de 2024 se explicaba que el mandato escrito lo sería para «respectivamente, ejecutar y llevar a cabo en su nombre las obligaciones y procedimientos establecidos por el presente Reglamento».

Puede observarse, entonces, que hay un matiz importante en la nueva redacción del artículo 3.5, ya que ahora habla de «cumplir las obligaciones» y antes expresaba que «para ejecutar y llevar a cabo en su nombre». De esta manera, el representante autorizado aparece obligado nuevamente y aquí sí (no como en el Considerando 82 que no lo desarrolla) a cumplir las obligaciones del sistema de Inteligencia Artificial de alto riesgo, con lo que se puede interpretar que se incluyen claramente las obligaciones de conservación de *logs*.

8. Directiva del Consejo, de 25 de julio de 1985, relativa a la aproximación de las disposiciones legales, reglamentarias y administrativas de los Estados miembros en materia de responsabilidad por los daños causados por productos defectuosos. Actualmente se discute la Propuesta de Directiva del Parlamento Europeo y del Consejo sobre responsabilidad por los daños causados por productos defectuosos (COM/2022/495 final).

Respecto de las obligaciones para los responsables del despliegue (antes usuario⁹), el Considerando 58 (versiones de 2021 y 2023) venía a decir que teniendo en cuenta «la naturaleza de los sistemas de IA y de los riesgos para la seguridad y los derechos fundamentales», los usuarios deben utilizar los sistemas de IA de acuerdo con las instrucciones de uso y deben asumir otras responsabilidades, entre las cuales están las del mantenimiento de registros, «según proceda». Encontramos aquí cierto vacío, puesto que la interpretación de este precepto también podía plantear diversos problemas técnicos y de tipo material, por mencionar sólo algunas de las implicaciones que su cumplimiento puede traer consigo para los usuarios, ahora responsables del despliegue.

Sin embargo, en la versión votada en marzo de 2024 (y en la corrección de errores de mayo) esta disposición aparece en el Considerando 91, en donde se manifiesta que los responsables del despliegue deben cumplir con las obligaciones de supervisión humana y de conservación de registros y, con este fin, deben «adoptar las medidas técnicas y organizativas adecuadas para garantizar que utilizan los sistemas de IA de alto riesgo conforme a las instrucciones de uso».

Con ello que se abre una ventana adicional en cuanto a las obligaciones de dichos responsables del despliegue, pues tendrán que implementar dichas medidas y en especial las de tipo técnico, con todo lo que ello conlleva en la práctica, puesto que, además, deben asegurarse de que las personas a las que encomienden las obligaciones antedichas cuenten con las capacidades y competencias necesarias para llevarlas a cabo. Se habla incluso de alfabetización, formación y autoridad en materia de Inteligencia Artificial, lo que se antoja cada vez más complejo en relación con la conservación de registros.

Por otra parte, el Considerando 133 (que no tenía precedente en la versión original del Reglamento) se refiere a la denominada Inteligencia Artificial generativa, aunque no la llama específicamente así¹⁰. Allí se plantea que es conveniente que se defina con la mayor claridad posible, cuando un contenido ha sido creado por dicha IA y no por un humano. Para distinguirlo y a efectos de controlar las consecuencias jurídicas, éticas y técnicas que podrían derivarse de un uso no adecuado e incluso lesivo de dicha Inteligencia generativa, se exigirá a los proveedores de estos sistemas que implementen una serie de técnicas y medidas para «marcar y detectar» que un contenido proviene de ese tipo de IA y no de un humano. Entre las técnicas que allí se mencionan se incluye la de «métodos de registro», todo ello según proceda y de acuerdo con el estado de la técnica; lo cual también plantea una serie de dudas de tipo práctico y de cuán viable es que se cumpla en los hechos con esa exigencia.

9. Que, de acuerdo con el artículo 3.4 de definiciones, pueden ser entidades privadas, Administraciones Públicas o incluso personas físicas, siempre que «utilice [n] un sistema de IA bajo su propia autoridad», salvo en el caso de usos meramente personales.

10. Se refiere a que «Una diversidad de sistemas de IA puede generar grandes cantidades de contenidos sintéticos que para las personas cada vez es más difícil distinguir del contenido auténtico generado por seres humanos. La amplia disponibilidad y las crecientes capacidades de dichos sistemas tienen importantes repercusiones en la integridad del ecosistema de la información y en la confianza en este, haciendo surgir nuevos riesgos de desinformación y manipulación a escala, fraude, suplantación de identidad y engaño a los consumidores».

Finalmente, el Considerando 165 que en versiones previas era el 81 se ocupaba de especificar que los sistemas de Inteligencia Artificial que no son de alto riesgo podrían asumir códigos de conducta destinados a fomentar la adopción voluntaria de los requisitos obligatorios aplicables a los sistemas de alto riesgo para que haya un uso «de confianza» de la IA en la Unión Europea. Es decir, se trataría de una medida de autorregulación que no es vinculante para aquellos sistemas que no entren en la categoría de aquellos que estamos abordando. Sin embargo, esta consideración que se desplaza ahora al citado Considerando 165 se matiza cuando se expresa que se alentará a los sistemas que no son de alto riesgo a que cumplan mediante códigos de conducta «la totalidad o parte» de las obligaciones aplicables a los que sí lo son y entre las que se encuentran las de conservación de registros y las de gobernanza.

Además de esto, en dicho Considerando se «anima» tanto a los proveedores como a los responsables del despliegue de todo tipo de sistemas (los de alto riesgo y lo que no lo son) a que adopten los principios establecidos las Directrices éticas de la Unión para una IA fiable.

III. LAS OBLIGACIONES EN MATERIA DE REGISTROS EN EL REGLAMENTO: EVOLUCIÓN, TRAMITACIÓN Y CONTENIDO FINAL

En este apartado explicaremos cómo ha sido la evolución del articulado referente a las obligaciones que tienen los sistemas de alto riesgo en materia de registros, aunque debemos adelantar que para este asunto concreto de los logs, las variaciones entre las tres principales versiones del RIA no habían demasado sustanciales, puesto que los artículos 12 (el único que no ha cambiado de número) y 19 (antes 20) han recibido pocas modificaciones que pudieran alterar el fondo de la cuestión y el artículo 26 (antes el 29) ha tenido numerosos apartados añadidos, pero aunque algunos de los conceptos allí tratados pueden tener vínculo con la obligación de conservación de registros, lo cierto es que están más bien orientados a temáticas que se tratan en otras partes de esta obra (*vid. supra e infra*).

Como hemos empezado diciendo en el párrafo anterior, el Reglamento Inteligencia Artificial aborda las obligaciones para los sistemas de alto riesgo de conservar registros, primero en el artículo 12 (**que aparece en la Sección 2 –antes Capítulo 2– de «Requisitos de los sistemas de IA de alto riesgo»**), que es el que se refiere específicamente a los registros de datos¹¹. En el apartado 1 dispone que «Los sistemas de IA de alto riesgo [anteriormente se escribía “se diseñarán y desarrollarán con capacidades que permitan”] permitirán técnicamente el registro automático de eventos (archivos de registro) [en las versiones previas se decía “logs”] a lo largo del ciclo de vida [antes se añadía “del sistema mientras los sistemas de IA de alto riesgo estén en funcionamiento”]». Aunque pareciera que se trata de matices muy breves, las implicaciones técnicas pueden ser un poco más sofisticadas que lo que la redacción actual permite suponer en un primer momento, ya que antes se hacía alusión al tiempo en que los sistemas de alto riesgo estuviesen en funcionamiento

11. Aunque el artículo 11, referente a la documentación técnica se tratará en profundidad en otro de los capítulos de esta obra (*vid. supra*), tiene plena relación con la obligación de conservación de registros, puesto que dicha documentación debe incluir la información sobre éstos.

y ahora se dispone que será a lo largo del ciclo de vida, lo cual implica también a la etapa del seguimiento post-comercialización.

Es oportuno explicar en este punto de la discusión, que los registros o *logs* son archivos donde se custodia la información relevante que generan el *kernel* o núcleo del sistema y cualquier programa que dicho sistema pueda tener embebido. Cuando la palabra que se utiliza es *log*¹², se trata de un anglicismo que podría traducirse como una bitácora. Como en toda bitácora, en el registro se guardan datos sobre todos los procesos o sucesos del sistema, en este caso, del sistema de IA de alto riesgo.

La mencionada bitácora puede contener datos muy puntuales sobre lo que sucede con el sistema y lo que arroja *kernel*, tales como la «marca temporal» en donde se puede registrar la fecha y hora en que ocurrió un determinado evento, pero la hora definida en minutos y en segundos, es decir, con mucha precisión. Por lo tanto, la información que arrojan todos estos y otros muchos datos, es altamente preciada y de inestimable valor, dado que certifica y prueba la conducta del sistema.

En esencia, los «archivos de registro (*logs*) proporcionan información muy importante sobre las actividades implícitas y explícitas de cualquier sistema de hardware y *software* informático. Este tipo de registro contiene toda la información sobre el funcionamiento normal de una máquina o programa, ayudando a interceptar anomalías y problemas, apoyando la seguridad».¹³

Por su parte y para que este asunto sea más comprensible para el lector que no tiene un perfil técnico, *kernel* es el núcleo del sistema y es aquel que hace posible el acceso en «modo privilegiado» al control de un sistema, por lo tanto, es el controlador principal y centro del sistema operativo. Incluso, además de la palabra núcleo, también se emplea la palabra «corazón» para definirlo y esto nos da cuenta de su importancia. Es por eso que «se encarga principalmente de mediar entre los procesos de usuario y el hardware disponible en la máquina, es decir, concede el acceso al hardware, al software que lo solicite, de una manera segura; y el procesamiento paralelo de varias tareas».¹⁴

Ahora bien, continuando con aquello que dispone el artículo 12 del Reglamento de Inteligencia Artificial, su apartado 2 se refiere a las capacidades de registro, que deben garantizar un nivel de trazabilidad del funcionamiento del sistema (ya hemos visto cómo es posible conseguir este objetivo, tomando en cuenta todo lo que se guarda en un *log* y la información que puede brindar), pero dicha trazabilidad debe poder realizarse a lo largo de todo el ciclo de vida¹⁵ del sistema de IA de alto riesgo. Es decir, se trata de un nivel de trazabilidad «que resulte

12. En todas las versiones del Reglamento previas a la votada en marzo de 2024 (que es la última de la que disponemos), había partes en donde se escribía el anglicismo *logs*, pero en la última se habla siempre en castellano (es decir, se les menciona como registros o archivos de registro) en la versión en este idioma del RIA. En la versión en inglés, siempre se usa el término *logs*.

13. Abonyi, J. y Bántay, L. «Frequent pattern mining-based log file partition for process mining», *Engineering Applications of Artificial Intelligence*, August, n.º 123, 2023.

14. Bach, F., «Information theory with kernel methods», *IEEE Transactions on Information Theory*, 69, vol. 2, 2022 (Disponible en <https://acortar.link/8pfUEk>).

15. **Aunque en esta última versión no se detalla con el término «ciclo de vida» sí se deriva esto de sus disposiciones pues menciona incluso el seguimiento postcomercialización y las etapas de supervisión humana.**

adecuado para la finalidad prevista del sistema» (en las versiones anteriores rezaba el «ciclo de vida que sea apropiado para la finalidad prevista del sistema»).

De este modo, las capacidades de registro de eventos permitirán: i) detectar aquellas situaciones de riesgo como las descritas en el artículo 79 sobre «Procedimiento aplicable a escala nacional a los sistemas de IA que presenten un riesgo» (que a su vez se refiere a lo dispuesto en el artículo 3.19 de definiciones del Reglamento 2019/1020¹⁶); ii) que sea factible llevar a cabo el control ulterior a la fase de comercialización, con arreglo a lo dispuesto en el artículo 72 sobre la «Vigilancia poscomercialización por parte de los proveedores y plan de vigilancia poscomercialización para sistemas de IA de alto riesgo» (*vid. infra*); y, iii) que se pueda llevar a cabo la inspección sobre el funcionamiento de los sistemas a los que se refiere el artículo 26 apartado 6 que trataremos más abajo, pero que en esencia alude a los responsables del despliegue que sean entidades financieras.

Todo lo hasta aquí expuesto sobre el artículo 12 plantea varias cuestiones, puesto que habrá que ver cómo se consolida en cada sistema de IA el ciclo de vida o cómo se le define y configura. También es interpretable la frase de aquel ciclo de vida «que sea apropiado para la finalidad prevista», puesto que cada agente del ciclo podría entenderlo de manera distinta.

El artículo 12 ha sufrido las modificaciones más relevantes en su apartado 3 pues en las versiones anteriores se exponía que las capacidades de registro debían permitir «el registro de eventos relevantes para la supervisión del funcionamiento del sistema de IA de alto riesgo con respecto a: (i) la identificación de situaciones que puedan dar lugar a que el sistema de IA presente un riesgo en el sentido del apartado 1 del artículo 65 [que ahora es el antes citado 79] (o que conduzcan a una modificación sustancial; y (ii) facilitar el seguimiento posterior a la comercialización a que se refiere el artículo 61 [ahora el antes mencionado 72]; (iii) la supervisión del funcionamiento de los sistemas de IA de alto riesgo a que se refiere el apartado 4 del artículo 29 [ahora el 26]» (referente a la supervisión humana, que ahora aparece en el 26, apartado 2). Como vemos, la mayor parte de este contenido ahora se concentra en el apartado 2 del artículo 12 en comentario.

Sin embargo, ahora el artículo 12 en su tercer apartado restringe las obligaciones de registro sólo a los sistemas de alto riesgo mencionados en el punto 1, letra a) del Anexo III (que son los sistemas de identificación biométrica remota). Pero además de la restricción únicamente a dichos sistemas, esta disposición se acoge a un sistema de «mínimos» cuando indica que los sistemas de identificación biométrica remota

16. Reglamento (UE) 2019/1020 del Parlamento Europeo y del Consejo de 20 de junio de 2019 relativo a la vigilancia del mercado y la conformidad de los productos y por el que se modifican la Directiva 2004/42/CE y los Reglamentos (CE) n.º 765/2008 y (UE) n.º 305/2011; en cuyo artículo 3.19 describe que se considerará un «producto que presenta un riesgo» como aquel que «puede afectar negativamente a la salud y la seguridad de las personas en general, a la salud y la seguridad en el trabajo, a la protección de los consumidores, al medio ambiente, a la seguridad pública o a otros intereses públicos protegidos por la legislación de armonización de la Unión aplicable, en un grado que vaya más allá de lo que se considere razonable y aceptable en relación con su finalidad prevista o en las condiciones de uso normales o razonablemente previsibles del producto en cuestión, incluida la duración de su utilización y, en su caso, los requisitos de su puesta en servicio, instalación y mantenimiento».

deberán asegurar que las capacidades de registro incluirán «como mínimo»: i) un registro del período de cada uso del sistema¹⁷, ii) la base de datos de referencia con la que el sistema ha cotejado los datos de entrada, iii) los datos de entrada con los que la búsqueda ha arrojado una correspondencia, y, iv) la identificación de las personas físicas involucradas en la comprobación de los resultados que se citan en el artículo 14, apartado 5¹⁸.

El artículo 16 (apartado e), antes d) y en las versiones anteriores en el Capítulo 3, ahora en la llamada Sección 3) sobre las «Obligaciones de los proveedores de sistemas de IA de alto riesgo» dispone que éstos deberán cumplir con la obligación de conservar «los archivos de registro generados automáticamente por sus sistemas de IA de alto riesgo a que se refiere el artículo 19 [antes el 20]» cuando estén bajo su control. Por supuesto que estas últimas cuatro palabras dan mucho margen de acción y una posibilidad alta de confusión entre los proveedores de los sistemas respecto de sus obligaciones de registro.

El artículo 16 actual también dispone en su apartado i), que los proveedores de los sistemas de Inteligencia Artificial de alto riesgo cumplirán las obligaciones de registro a que se señalan en el artículo 49, apartado 1, concerniente a las obligaciones de alta en la base de datos que la UE manejará a efectos de control de estos sistemas.

Por otra parte, el artículo 19 (antes el 20) alude expresamente a los archivos de registro generados automáticamente (antes sólo rezaba «registros generados automáticamente») y se mantiene con sólo dos párrafos y los cambios no han sido sumamente notorios, como veremos a continuación.

En las versiones que se han ido modificando desde 2019 y hasta la última de 2024, el anterior artículo 20, establecía que se conservarán los registros (ahora archivos de registro) «generados automáticamente por sus sistemas de IA de alto riesgo». En la actual versión (ya como artículo 19) se dispone que se conservarán aquellos «que los sistemas de IA de alto riesgo generen automáticamente», siempre y cuando éstos se encuentren bajo su control¹⁹. Como se observa, es sólo un matiz en la frase cambiando el orden de las palabras. También es importante destacar que en este artículo se hace referencia a que son los proveedores de los sistemas de alto riesgo los que deben conservar esos registros y no se hace alusión a los responsables del despliegue u otros actores intervinientes en la cadena de valor o ciclo de vida del sistema de IA.

En cuanto a los periodos en los que han de conservarse estos archivos de registro, en versiones previas y en la final, se menciona un período mínimo de seis meses, «salvo que el Derecho de la Unión o nacional aplicable, en particular el Derecho de la Unión en materia de protección de datos personales, disponga otra cosa».

17. Refiriéndose a la fecha y la hora de inicio y la fecha y la hora de finalización de cada uso.

18. Es decir, nuevamente se refiere a los sistemas de identificación biométrica remota, pues el artículo 14 (sobre vigilancia humana –en otras versiones e incluso en otros apartados del Reglamento, denominada supervisión humana–) en su apartado 5 remite nuevamente al punto 1, letra a), del anexo III que hemos citado antes.

19. En las versiones anteriores, se estipulaba que los registros se conservarían, en la medida que estuviesen bajo su control «en virtud de un acuerdo contractual con el usuario o de otro modo por ley». En la última versión de este artículo se elimina toda referencia al acuerdo contractual.

Es decir, este artículo 19 está marcando una condición para los periodos de conservación que podría verse modificado si los multicitados archivos de registro incluyen datos personales, en cuyo caso les sería aplicable el Reglamento General de Protección de Datos y las legislaciones locales, además de cualquier otro instrumento del Derecho de la Unión o nacional que pueda tener relación con esos periodos de conservación de los *logs*.

El segundo párrafo del artículo 19 se refiere específicamente a los proveedores que sean entidades financieras, en cuyo caso, la conservación de los archivos de registro se llevará a cabo con arreglo a lo dispuesto por la legislación de ese sector en concreto. La modificación que ha tenido lugar en la última versión de 2024 respecto de otras redacciones previas, es que en éstas se mencionaba específicamente el artículo 74 de la Directiva 2013/36/UE, mientras que en la versión final, no se hace referencia a ningún ordenamiento concreto, sino que se especifica que todo se hará en virtud de lo que disponga el «Derecho pertinente en materia de servicios financieros».

El artículo 26, intitulado «Obligaciones de los responsables del despliegue de sistemas de IA de alto riesgo», antes artículo 29 «Obligaciones de los usuarios de sistemas de IA de alto riesgo», fue objeto de muchos cambios en la versión de febrero de 2024, respecto a las anteriores, hasta el punto de haberle dotado de una casi nueva redacción, en algunos aspectos, realizándose cambios puntuales posteriormente, tal y como observaremos.

En este contexto, el citado artículo viene a delimitar, en su apartado 1, las medidas técnicas y organizativas²⁰ adecuadas para garantizar que utilizan dichos sistemas con arreglo a las instrucciones de uso que los acompañen, de acuerdo con los apartados 3 y 6, antes apartados 2 y 5, en las otras versiones de trabajo.

Estas obligaciones deben entenderse, como se indica en el Considerando 120, destinadas a permitir que se detecte y divulgue que los resultados de dichos sistemas han sido generados o manipulados de manera artificial resultan especialmente pertinentes para facilitar la aplicación efectiva del Reglamento (UE) 2022/2065²¹, considerando que sólo había sido recogido en la versión anterior, donde se ubicaba en el considerando 70 quiquies.

En cuanto a las instrucciones, de uso de las medidas técnicas y organizativas²², se refieren, en primer lugar, a la supervisión humana a personas físicas que tengan

20. Debe tenerse en cuenta que hablamos de medidas que garanticen la interoperabilidad, en sentido puramente informático. Esto supone una gran importancia; pues, hablamos de medidas de interoperabilidad, en este caso: a) técnicas: conexión de sistemas sean eficientes sin fallos en materia de ciberseguridad (por ejemplo, intercambio de datos) y b) organizativas, referidas a los procesos de negocio y estructuras internas, por ejemplo, que los sistemas puedan intercambiar datos más allá de su contenido técnico, es decir, hablamos de la estandarización a través del uso de las normas ISO, hablamos de *documentos que contienen reglas, instrucciones o características que pueden ser utilizadas para asegurar qué materiales, productos, procesos y servicios son adecuados para su propósito*.

21. Reglamento (UE) 2022/2065, relativo a un mercado único de servicios digitales y por el que se modifica la Directiva 2000/31/CE (Reglamento de Servicios Digitales).

22. Debe tenerse en cuenta, con respecto a estas medidas organizativas que en éstas se van a establecer, además, habilidades que están contemplados en el seno de las diferentes organizaciones; es decir, que actores participan y por tanto van a ser suje-

la competencia, la formación y la autoridad necesarias, tal y como se aclara en el apartado 2 del artículo, agregado a la última versión de marzo. Ahora bien, debe observarse que en esta versión se suprime una referencia «al apoyo necesario»²³, lo que guarda relación con las «competencias necesarias», en particular un nivel adecuado de alfabetización, formación y autoridad en materia de IA para desempeñar adecuadamente dichas tareas, tal y como se indica en el artículo 4 de referencia a la alfabetización en materia de inteligencia artificial.

En este sentido, en los apartados 3 y 4 del artículo, antes 2 y 3, respectivamente, tienen idéntica redacción en los textos anteriores en los apartados 2 y 3, salvando la referencia al responsable del despliegue.

Además, en referencia al apartado 5, antes apartado 4, se introduce como indica el Considerando 91, que no aparece con anterioridad, que no tienen correlación con ningún otro considerando en las versiones anteriores, donde se indica la necesidad de garantizar la correcta vigilancia del funcionamiento de un sistema de IA en un entorno real. En este contexto, determina la necesidad de definir las responsabilidades específicas de los responsables del despliegue. En particular, los responsables del despliegue deben adoptar las medidas técnicas y organizativas adecuadas para garantizar que utilizan los sistemas de IA de alto riesgo conforme a las instrucciones de uso. Además, es preciso definir otras obligaciones en relación con la vigilancia del funcionamiento de los sistemas de IA y la conservación de registros, según proceda.

En este sentido, el apartado 5 indica que los responsables del despliegue vigilarán el funcionamiento del sistema de IA de alto riesgo basándose en las instrucciones de uso y, cuando proceda, informarán a los proveedores con arreglo al artículo 72 (en referencia a la vigilancia postcomercialización). Cuando los responsables del despliegue tengan motivos para considerar que utilizar el sistema de IA de alto riesgo conforme a sus instrucciones podría presentar un riesgo en el sentido del artículo 79, apartado 1, informarán, sin demora indebida, al proveedor o distribuidor y a la autoridad de vigilancia del mercado pertinente y suspenderán el uso del sistema, hablamos del procedimiento aplicable a escala nacional a los sistemas de IA que presenten un riesgo.

Del mismo modo, si los responsables del despliegue detectan un incidente grave, informarán inmediatamente de dicho incidente, en primer lugar, al proveedor y, a continuación, al importador o distribuidor y a la autoridad de vigilancia del mercado pertinente. En el caso de que el responsable del despliegue no consiga contactar con el proveedor, el artículo 73 (en referencia a la notificación de incidentes graves) se aplicará *mutatis mutandis*. Esta obligación no abarcará los datos operativos sensibles de los responsables del despliegue de sistemas de IA que sean autoridades encargadas de la aplicación de la ley.

En el caso de los responsables del despliegue que sean entidades financieras, porque como indica el citado Considerando 91 es preciso definir otras obligaciones en relación con la vigilancia del funcionamiento de los sistemas de IA y la conservación

tos responsables, determinándose, en un sentido jurídico-informático, un «ámbito de aplicación personal» con arreglo a las normas de responsabilidad que se puedan establecer en otros instrumentos.

23. Esta omisión se observa al comparar el texto del proyecto de acuerdo emitido por la Comisión Europea, cuyo origen lo situamos en febrero de 2024.

de registros, según proceda, al estar estas sujetas, además, a requisitos relativos a su gobernanza, sus sistemas o sus procesos internos en virtud de la legislación de la Unión en materia de servicios financieros, se considerará que se ha cumplido la obligación de vigilancia prevista en el párrafo primero cuando se respeten las normas sobre gobernanza, sistemas, procesos y mecanismos internos de acuerdo con la legislación pertinente en materia de servicios financieros.

El apartado 6, antes incluido en el apartado 4, hace referencia a la conservación de registros, como indica el Considerando 91 y el artículo 19 (*vid. supra*), según proceda. En este sentido, en la medida en que dichos archivos estén bajo su control²⁴ durante un período de tiempo adecuado, de al menos seis meses, salvo disposición en contrario en el Derecho de la Unión o nacional aplicable, en particular en el Derecho de la Unión en materia de protección de datos personales. Con respecto a las entidades financieras, los archivos de registro como parte de la documentación conservada en virtud del Derecho de la Unión en materia de servicios financieros.

Los apartados 7 y 8, antes apartados 5-a) y b) - c), respectivamente, respecto de la anterior versión y sin que tengan equivalencia con ninguna otra, guardan relación con la conservación de los registros del apartado 6, refiriéndose a los responsables del despliegue que sean empleadores informarán a los representantes de los trabajadores y a los trabajadores afectados de que estarán expuestos a la utilización del sistema de IA de alto riesgo y a los responsables del despliegue de sistemas de IA de alto riesgo que sean autoridades públicas o instituciones, órganos y organismos de la Unión cumplirán las obligaciones de registro a que se refiere el artículo 49 (sobre Registro).

El apartado 9, antes apartado 6 y sin cambios en todas las versiones anteriores, se refiere a la información facilitada en virtud del artículo 13 para cumplir con su obligación de realizar una evaluación de impacto relativa a la protección de datos con arreglo al artículo 35 del Reglamento (UE) 2016/679²⁵ o al artículo 27 de la Directiva (UE) 2016/680²⁶. El apartado 10, antes apartado 6, bis-a), no tiene cambios respecto a la anterior y ni precede a las anteriores.

24. En este punto, en el texto realizado en la versión por mandato del Parlamento Europeo, previo al acuerdo, se eliminó del texto una referencia que puede resultar particularmente interesante en tanto en cuanto hacía referencia a que estén bajo su control «y son necesarios para garantizar y demostrar el cumplimiento del presente Reglamento, para auditorías a posteriori de cualquier mal funcionamiento, incidente o uso indebido del sistema razonablemente previsible, o para garantizar y supervisar el correcto funcionamiento del sistema a lo largo de su ciclo de vida».

25. Reglamento General de Protección de Datos, ya citado antes.

26. El artículo 27 de la Directiva protección de las personas físicas en lo que respecta al tratamiento de datos personales por parte de las autoridades competentes para fines de prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de sanciones penales, y a la libre circulación de dichos datos hace referencia a cuando sea probable que un tipo de tratamiento, en particular utiliza nuevas tecnologías, por su naturaleza, alcance, contexto o fines, suponga un alto riesgo para los derechos y libertades de las personas físicas, «los Estados miembros dispondrán que el responsable del tratamiento lleve a cabo, con carácter previo, una evaluación del impacto de las operaciones de tratamiento previstas en la protección de datos personales». En este sentido, se eliminó la referencia a que «los responsables del despliegue podrán recurrir en parte a dichas evaluaciones de impacto sobre la protección de datos para cumplir algunas de las obligaciones establecidas en el presente artículo, en la me-

El apartado 11, antes apartado 6, ter-b), se produce un cambio en relación con el artículo de referencia con respecto a lo dispuesto en el artículo 50²⁷, antes artículo 52, los responsables del despliegue de los sistemas de IA de alto riesgo a que se refiere el anexo III que tomen las personas físicas de que están expuestas a la utilización de los sistemas de IA de alto riesgo. Asimismo, si bien en la versión inmediatamente anterior se hacía referencia a los sistemas de IA de alto riesgo, utilizados con fines policiales, se abunda en los sistemas que se utilicen a los efectos de decisiones o ayuden a tomar decisiones relacionadas con personas físicas informarán a de la aplicación de la ley, aplicándose con ello el artículo 13 de la Directiva (UE) 2016/680.

Finalmente, el apartado 12, apartado 6 quarter, c, no tiene cambios respecto a la versión anterior, ni precedente en las versiones anteriores.

IV. REFLEXIONES Y ANÁLISIS FINAL

Tras el análisis de toda la evolución normativa de los Considerandos y articulado referentes a la obligación de conservar los registros generados automáticamente por los sistemas de Inteligencia Artificial de alto riesgo en el Reglamento Europeo de Inteligencia Artificial se puede concluir que el debate entre las autoridades de la UE ha sido intenso y dado paso a una serie de modificaciones profundas en las distintas versiones, pero especialmente entre la primera y la última (original de 2021 y la de la primera votación de marzo de 2024 y la corrección de errores de mayo del mismo año).

Respecto de las obligaciones de conservación de registros explicadas hasta el momento, cabe preguntarse todas estas obligaciones de registro pueden traer consigo problemas, por ejemplo, cuando haya datos personales de por medio (ya hemos visto en qué casos la legislación en la materia es plenamente aplicable). Y nos referimos no solamente datos personales e incluso sensibles o categorías especiales de datos (como podrían ser los neurodatos²⁸), sólo datos personales.

Los *logs* pueden contener información confidencial que no debe ser revelada por privacidad o incluso porque su revelación hace vulnerable la seguridad del sistema. En estos casos es necesaria proteger la confidencialidad de la información. Para solucionar estos problemas se usan en los *logs* técnicas de cifrado.

Las obligaciones de los responsables del despliegue de sistemas de IA de alto riesgo vienen a indicar cuestiones de relevancia, prueba de ello son los numerosos cambios que se han producido desde su primera redacción. Todo con el fin de poder garantizar que, en el uso de la IA, todo se desarrolle de una manera transparente, responsable y con un perfil ético. En este sentido, puede observarse como, con el objetivo de que se cumpla la norma, se busca reducir los riesgos potenciales y garantizar un uso responsable de la IA estableciendo mecanismos de supervisión y rendición de cuentas, para el establecimiento de sistemas de IA seguros y justos, que sean.

dida en que las evaluaciones de impacto sobre la protección de datos cumplan dichas obligaciones», en el texto realizado en la versión por mandato del Parlamento Europeo.

27. En referencia a las obligaciones de transparencia de los proveedores y usuarios de determinados sistemas de IA.

28. Ver, Arellano Toledo, W., «Los neuroderechos y su regulación» *Inteligencia Artificial. Iberoamerican Journal of Artificial Intelligence*, vol. 27, n.º 73 (2024).

Transparencia y comunicación de información a los responsables del despliegue en el artículo 13 Reglamento de inteligencia artificial

MARÍA ESTRELLA GUTIÉRREZ DAVID

Profesora Contratada Doctor de Derecho Constitucional
Universidad Complutense de Madrid

I. INTRODUCCIÓN: APROXIMACIÓN GENERAL AL ARTÍCULO 13 DEL REGLAMENTO

Si bien la «transparencia» viene siendo uno de los principios que con más frecuencia se menciona en la doctrina jurídica, en la *soft law* y en las incipientes legislaciones sectoriales nacionales sobre IA, la interpretación de su contenido y alcance varía en cuanto a qué debe ser transparente (e.g., el uso de datos, el código fuente, la interacción entre el ser humano y la IA, las decisiones automatizadas, la finalidad del uso de datos o la aplicación del sistema de IA), quiénes son los sujetos obligados por la transparencia y quiénes las partes interesadas destinatarias de la transparencia y, en su caso, la finalidad de la transparencia (e.g., minimización del daño, mejora de la IA, razones jurídicas, fomento de la confianza, principio de democracia).¹ En este sentido, existe un consenso amplio en que el nivel o grado de transparencia, cualitativa y cuantitativa puede variar en función de quienes sean las partes interesadas o destinatarias (implementadores de los sistemas de IA y sus usuarios, el público en general, personas individuales o colectivos afectados por las decisiones de los sistemas, analistas de incidentes, reguladores, autoridades de certificación y auditores, operadores jurídicos en el ámbito administrativo o judicial).²

1. Schneeberger, D. *et al.*, «The Tower of Babel in Explainable Artificial Intelligence (XAI)», en Holzinger, A., Kieseberg, P., Cabitza, F., Campagner, A., Tjoa, A. M., Weippl, E. (eds.), *Machine Learning and Knowledge Extraction. CD-MAKE 2023. Lecture Notes in Computer Science*, Springer, Cham, vol. 14065 (2023), p. 67. https://doi.org/10.1007/978-3-031-40837-3_5
2. IEEE Standards Association, «IEEE Standard for Transparency of Autonomous Systems», en *IEEE Std 7001-2021*, pp. 18-30, 4 de marzo 2022, doi: 10.1109/IEEESTD.2022.9726144; Information Commissioner's Office and Alan Turing Institute, *Explaining decisions made with AI*, v. 1.0.14, 17 October 2022, p. 38. <https://ico.org.uk/media/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/explaining-decisions-made-with-artificial-intelligence-1-0.pdf>; Government of Cana-

El artículo 8.1 RIA (en adelante, «RIA») dispone que los sistemas de IA de alto riesgo deberán cumplir una serie de requisitos «teniendo en cuenta sus finalidades previstas, así como el estado actual de la técnica generalmente reconocido en materia de IA y tecnologías relacionadas con la IA». Dichos requisitos se encuentran previstos en la Sección 2ª del Capítulo 3º e incluyen áreas relacionadas con la gestión de riesgos (Artículo 9), la calidad de los datos y la gobernanza (Artículo 10), la documentación técnica (Artículo 11), la conservación de registros (Artículo 12), la transparencia y comunicación de información a los responsables del despliegue (artículo 13), las medidas de supervisión humana (Artículo 14) y la precisión, solidez y ciberseguridad (Artículo 15). El presente Capítulo tiene por objeto sistematizar y analizar el contenido y alcance del requisito de transparencia y comunicación de información previsto en el artículo 13 RIA.³

1. EL PROCESO LEGISLATIVO DE CONFORMACIÓN DEL ARTÍCULO 13 Y SUS RETOS INTERPRETATIVOS

A diferencia del texto finalmente publicado el 14 de mayo de 2024 por el Parlamento Europeo y el Consejo, la propuesta inicial del Reglamento presentada por la Comisión Europea⁴ no incluía ninguna definición de la «transparencia». Entre las más de setecientas enmiendas presentadas por el Parlamento Europeo al texto propuesto por la Comisión, se incluyó un párrafo adicional al apartado (1) del Art. 13 con una mención explícita a la transparencia. En dicha enmienda, la transparencia se identificaba con la adopción de aquellas medidas técnicas que garantizaran la interpretabilidad de las decisiones adoptadas por los sistemas de alto riesgo tanto por el propio proveedor como por el usuario del sistema (en la terminología de la versión final del RIA, el «responsable del despliegue»). Además, dicha definición se conectaba con el derecho a una explicación a las personas individuales afectadas por las decisiones adoptadas por los sistemas de IA con un contenido similar al del vigente Artículo 86 del Reglamento.⁵

da, *Directive on Automated Decision-Making* (2019). Apéndice B. Impact Assessment Levels, última actualización, 25 de abril de 2023.

3. Véase Parlamento Europeo y Consejo, REGLAMENTO DEL PARLAMENTO EUROPEO Y DEL CONSEJO por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial) quedando aún pendiente en el momento de cierre del mismo la publicación en el Diario Oficial de la Unión Europea (PE-CONS 24/24), de 14 de mayo de 2024. En el momento de cierre de este Capítulo aún no se ha publicado en el Diario Oficial de la Unión Europea el texto final definitivo (en adelante, «PE-CONS 24/24»).
4. Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión (COM/2021/206 final).
5. Vid. Enmiendas aprobadas por el Parlamento Europeo el 14 de junio de 2023 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión (COM(2021)0206

Ya en el Acuerdo provisional del 2 de febrero de 2024⁶ adoptado por los legisladores europeos durante el proceso de trilogos, se introdujo un Considerando (14a) donde se definía la transparencia en la línea propuesta por las Directrices Éticas del Grupo de Expertos de la Comisión, y cuyo contenido fue recogido de manera literal por el Considerando (27) del texto finalmente aprobado del Reglamento.⁷

Por su parte, frente a la propuesta de la Comisión, el texto de Reglamento presentado por el Consejo incluía dos incisos nuevos en los apartados (3)(b)(i) y (3)(e), y dos nuevos apartados (3)(b)(va) y (3)(ea), al texto del artículo 13 previsto en la propuesta de la Comisión.⁸

A su vez, el apartado (3)(iv) del Artículo 13, añadido a la propuesta de la Comisión, tiene su origen en el Acuerdo provisional del 2 de febrero de 2024⁹ que, con alguna leve variación final, ha sido recogido en el texto definitivo aprobado en mayo de 2024.

Al abordar el contenido y alcance del requisito de transparencia previsto en el artículo 13 RIA deben hacerse las siguientes consideraciones. En primer lugar, la «transparencia» referida a los sistemas de IA es un concepto polisémico, variable y modulable en función del dominio (técnico, ético, soft law, jurídico), el contexto de aplicación del sistema (sector público o privado), sus impactos (individuales o colectivos), la finalidad de la transparencia (interna o externa) y los sujetos destinatarios de la misma.¹⁰ En segundo lugar, aunque la transparencia es un requisito esencial

— C9-0146/2021 — 2021/0106(COD)), enmienda n.º 300. La enmienda establecía que: «[...] la transparencia significará que, en el momento de la introducción en el mercado del sistema de IA de alto riesgo, se utilizarán todos los medios técnicos disponibles de conformidad con el estado actual de la técnica generalmente reconocido para garantizar que el proveedor y el usuario puedan *interpretar la información de salida* del sistema de IA de alto riesgo. El usuario estará capacitado para comprender y utilizar adecuadamente el sistema de IA sabiendo, en general, cómo funciona el sistema de IA y qué datos trata, lo que le permitirá explicar las decisiones adoptadas por el sistema de IA a la persona afectada de conformidad con el artículo 68, letra c).»

6. Vid. PE758.862v01-00.

7. PE-CONS 24/24.

8. Véase, Consejo, *Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Reglamento de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión (CONSIL_ST_15698_2022_INIT_6 DIC 22)*, de 6 de diciembre de 2022. El inciso en el apartado (3)(b)(i) incluía, junto a la finalidad prevista, el siguiente inciso: «incluido el entorno geográfico, funcional o de comportamiento específico en el que se pretende utilizar el sistema de IA de alto riesgo.» El nuevo apartado (3)(b)(va) pretendía adicionar en las instrucciones de uso una nueva categoría de información: «cuando proceda, descripción de la información de salida esperada del sistema». Ninguna de estas dos propuestas del Consejo prosperaron, pero en cambio sí se incorporaron al texto final del Artículo 13 un inciso en el apartado 3(e), «los recursos informáticos y de hardware necesarios», y un nuevo apartado 3(ea), referido a los archivos de registro (actual, apartado 3(f)).

9. El nuevo apartado introducido por el Acuerdo provisional de 2 de febrero de 2024 establecía como información a incluir en las instrucciones de uso: «(iii) en su caso, las capacidades y características técnicas del sistema de IA para proporcionar información pertinente que explique sus resultados.»

10. Sobre los significados de la transparencia en el ámbito de la IA, véase Cotino Hueso, L., «Transparencia y explicabilidad de la inteligencia artificial y “compañía” (comunicación, interpretabilidad, inteligibilidad, auditabilidad, testabilidad, comprobabili-

de los sistemas de alto riesgo, su ausencia es uno de los retos fundamentales de la IA. Se dice, no sin razón que el concepto de transparencia podría ser incluso más opaco que el de la propia IA.¹¹ Por eso, desde distintos ámbitos científicos, se han intentado aportar soluciones para mejorar la transparencia de la IA, al tiempo que se han articulado conceptos diferentes pero afines que incluyen, la explicabilidad y la interpretabilidad. Sin embargo, no existe una taxonomía común ni dentro de un mismo campo científico (Ciencias de la Computación, Ciencia de Datos) ni entre campos diferentes (Derecho y Ciencia de Datos). En tercer lugar, esta falta de consenso en el ámbito científico se ha trasladado a otros dominios, como el de la Ética, la soft-law o el Derecho, generándose un efecto de «Torre de Babel» de terminologías y conceptos confusos que hacen difícil identificar un fundamento científico común.¹² El RIA ha acabado recogiendo parte de esta confusión conceptual y terminológica.

2. OBJETIVOS Y CONSIDERACIONES METODOLÓGICAS

Teniendo en cuenta las consideraciones anteriores, el Capítulo pretende abordar un análisis sistemático del artículo 13 RIA a partir de la identificación de las distintas dimensiones de la transparencia, así como el contenido y alcance del precepto. La estructura de este Capítulo aborda, en primer lugar, una descripción sistemática del artículo 13 y la correlación de sus apartados con otras disposiciones del Reglamento, así como necesarias referencias a la normalización a efectos de completar la interpretación de los aspectos más técnicos del precepto. En segundo lugar, se analiza cómo se ha producido la integración de los requisitos de transparencia, interpretabilidad y explicabilidad en el texto del artículo 13. Dicho análisis pasa por una necesaria referencia a la conceptualización técnica, a los antecedentes pre-legislativos del texto y a la interrelación entre los tres requisitos, al tiempo que se incorpora un estudio exhaustivo del enfoque (insuficiente y ambiguo) que el RIA incorpora de la interpretabilidad y la explicabilidad. En tercer lugar, la exégesis del artículo 13 pasa por abordar las tres dimensiones de la «transparencia» que incorpora el precepto: una dimensión subjetiva (quiénes son los sujetos obligados y destinatarios de la transparencia), una dimensión formal (el cómo o modo de cumplimiento de la obligación), y una dimensión sustantiva o material (qué información debe comunicarse a los sujetos destinatarios de la transparencia). Finalmente, y dado el marcado carácter de reglamentación técnica que tiene el Reglamento, se ha incorporado un apartado específico sobre el papel de la normalización en el desarrollo del artículo 13.

A la hora de establecer una metodología adecuada para abordar el análisis y exégesis del artículo 13, se han tenido en cuenta los siguientes aspectos. En primer lugar, el artículo 13, al igual que buena parte del texto de la norma, tiene una

dad, simulabilidad...). Para qué, para quién y cuánta», Cotino Hueso, L., Castellanos, J. (coords.), *Transparencia y explicabilidad de la inteligencia artificial*, Tirant lo Blanch, Valencia, 2022, pp. 25-65.

11. Kiseleva, A., Kotzinos, D., De Hert, P., «Transparency of AI in Healthcare as a Multilayered System of Accountabilities: Between Legal Requirements and Technical Limitations», *Frontiers in Artificial Intelligence*, vol. 5 (2022), p. 1. <https://doi.org/10.3389/frai.2022.879603>
12. Schneeberger, D. *et al.*, «The Tower of Babel...», p. 66.

vocación clara de reglamentación técnica. Por tanto, para una correcta interpretación teleológica del precepto, será necesaria su integración no sólo con otros preceptos del Reglamento (con idéntica *vis técnica*) y algunos apartados específicos del Anexo IV.¹³ En segundo lugar, para asegurar una adecuada selección de aquellos términos de claro significado técnico incluidos en el artículo 13 (e.g. precisión, métricas, rendimiento, transparencia, interpretación, explicación, entre otros), a falta de definición en el Reglamento, ha resultado necesario acudir a las normas técnicas y documentos de normalización que algunos organismos internacionales y nacionales han empezado a publicar. Es el caso de la Organización Internacional de Normalización («ISO»)¹⁴, el European Telecommunications Standards Institute («ETSI»), la Institute of Electrical and Electronics Engineers Standards Association («IEEE»), el National Institute of Standards and Technology del Departamento Norteamericano de Comercio («NIST»).

A efectos metodológicos, se incorpora a continuación el listado de normas técnicas aprobadas o pendientes de aprobar por los distintos organismos de normalización que se han tenido en cuenta a la hora de elaborar este Capítulo.

Tabla 1
Relación de Normas Técnicas referenciadas

Norma Técnica [Estatus]
ISO/IEC DIS 12792: 2024(en). Information technology — Artificial intelligence — Transparency taxonomy of AI systems [En trámite].
ISO/IEC 25059:2023(E). Software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI systems [Aprobada: 26/06/2023].
ISO/IEC 25010:2023(en). Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Product quality model [Aprobada: 15/11/2023].
ISO/IEC 22989:2022. Information technology — Artificial intelligence — Artificial intelligence concepts and terminology [Aprobada: 19/07/2022]. Corrigenda: ISO/IEC 22989:2022/ AWI Amd 1.

13. Aunque el Anexo IV se refiere al contenido básico de la documentación técnica prevista en el Artículo 11 del Reglamento, debe indicarse que algunas de las categorías de información ahí indicadas coinciden con las categorías de información previstas en el Artículo 13.(3). Dado que el Anexo IV aborda con mayor detalle las categorías de información, su referencia permite integrar, en muchos casos, el significado de los distintos apartados del Artículo 13.
14. En el momento de redactar este Capítulo, el Comité ISO/IEC JTC 1/SC 42, sobre inteligencia artificial ha aprobado unas 28 normas técnicas sobre inteligencia artificial y se encuentra en un proceso de elaboración de otras 30 normas. Entre las normas técnicas en proceso de elaboración que se han tenido en cuenta a efectos de interpretación del contenido y alcance del Artículo 13, entre otras: la ISO/IEC DIS 12792 — Transparency taxonomy of AI systems y la ISO/IEC CD TS 6254— Objectives and approaches for explainability and interpretability of ML models and AI systems.

Norma Técnica [Estatus]
ISO/IEC TS 5723:2022(en). Trustworthiness — Vocabulary [Aprobada 22/07/2022].
ISO/IEC 23053:2022. Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML) [Aprobada: 20/06/2022]. Corrigenda: ISO/IEC 23053:2022/AWI Amd 1.
ISO/IEC TR 24027:2021. Information technology — Artificial intelligence (AI) — Bias in AI systems and AI aided decision making [Aprobada: 05/11/2021].
ISO/IEC TR 24029-1:2021. Artificial Intelligence (AI) — Assessment of the robustness of neural networks [Aprobada: 10/03/2021].
IEEE 7001-2021. Standard for Transparency of Autonomous Systems [Aprobada 08/12/2021].
ETSI TR 104 032 (V1.1.1) (2024-02). Securing Artificial Intelligence (SAI); Traceability of AI Models [Aprobada: 2024-02].
ETSI GR SAI 007 (V1.1.1) (2023-03). Securing Artificial Intelligence (SAI). Explicability and transparency of AI processing [Aprobada 07/03/2023].
ETSI GR SAI 009 V1.1.1 (2023-02) Securing Artificial Intelligence (SAI); Artificial Intelligence Computing Platform Security Framework [Aprobada 16/02/2023].
ETSI GR SAI 004 V1.1.1 (2020-12). Securing Artificial Intelligence (SAI); Problem Statement. [Aprobada: 12/2021].
NISTIR 8312 (2021). Four Principles of Explainable Artificial Intelligence.
NISTIR 8269 (2019). A Taxonomy and Terminology of Adversarial Machine Learning.

Fuente: Elaboración propia

La consulta y remisión a las normas técnicas indicadas en la Tabla 3 *infra* ha permitido completar el significado de los aspectos más técnicos del contenido del artículo 13, así como introducir un análisis comparado desde la perspectiva de la normalización en cuanto a la categorización de los tipos y niveles de transparencia (Tabla 2) o de las categorías de interesados y de información relevante (Tabla 7).

En particular, han sido de especial utilidad la ISO/IEC DIS 12792:2024(en), la ISO/IEC 25059:2023, sobre modelo de calidad para sistemas de IA, o la ISO/IEC 22989:2022, donde se incluyen conceptos básicos sobre IA.

En el caso de la ISO/IEC DIS 12792:2024(en), aunque se trate de una norma técnica en trámite el momento de cierre de este Capítulo, resulta relevante su consulta a los efectos de este Capítulo por los niveles de transparencia e información técnica que establece (nivel del contexto del sistema de IA, nivel del sistema de IA, nivel del modelo de IA, y nivel del conjunto de datos utilizados por el sistema).

Tabla 2
Niveles de transparencia en la ISO/IEC DIS 12792:2024(en)

Taxonomía del nivel de transparencia	Categorías de información
Nivel de contexto [7]	Contexto general [7.1] Contexto social [7.2]; Generalidades [7.2.1]; Prácticas laborales [7.2.2]; Necesidades de los consumidores [7.2.3] Contexto medioambiental [7.3]
Taxonomía a nivel de sistema [8]	Generalidades [8.1] Información básica [8.2] Procesos organizativos [8.3]: Generalidades [8.3.1]; Gobernanza [8.3.2]; Sistema de gestión [8.3.3]; Gestión de riesgos [8.3.4]; Gestión de la calidad [8.3.5] Aplicabilidad [8.4]: Generalidades [8.4.1]; Fines previstos [8.4.2]; Capacidades [8.4.3]; Limitaciones funcionales [8.4.4]; Usos recomendados [8.4.5] Usos excluidos [8.4.6] Resumen de las características técnicas [8.5]: Generalidades [8.5.1]; Entradas y salidas previstas [8.5.2]; Datos de producción [8.5.3]; Registro y almacenamiento [8.5.4]; Descomposición del sistema [8.5.5]; Interfaz de programación de aplicaciones [8.5.6]; Factores humanos [8.5.7]; Métodos de despliegue [8.5.8]; Gestión de la configuración [8.5.9] Acceso a los elementos internos [8.6] Calidad y rendimiento [8.7]: Generalidades [8.7.1]; Procesos de verificación y validación [8.7.2]; Mediciones en tiempo de ejecución [8.7.3]; Comparación con sistemas alternativos [8.7.4]
Taxonomía a nivel de modelo [9]	Generalidades [9.1] Información básica [9.2] Utilización [9.3]: Procesamiento realizado por el modelo [9.3.1]; Dependencia de otros modelos [9.3.2]; Coherencia con fines previstos del sistema de IA [9.3.3] Características técnicas [9.4]: Tipo de tecnología utilizada [9.4.1]; Atributos extraídos de los datos de entrada [9.4.2]; Algoritmo utilizado para el procesamiento [9.4.3]; Procedimiento de construcción del modelo [9.4.4]; Hiperparámetros [9.4.5]; Formatos de entrada y salida [9.4.6]; Hardware de procesamiento [9.4.7]; Costes computacionales [9.4.8]; Modelos en sistemas evolutivos [9.4.9] Datos utilizados [9.5] Corrección funcional [9.6]

Taxonomía del nivel de transparencia	Categorías de información
Taxonomía a nivel de conjunto de datos [10]	Generalidades [10.1] Información básica [10.2] Procedencia de los datos [10.3] Propiedades de los datos [10.4] Dominio y propósitos del conjunto de datos [10.5] Sesgos y limitaciones de los datos [10.6] Consideraciones sociales [10.7] Preparación de los datos realizada [10.8] Mantenimiento del conjunto de datos [10.9]

Fuente: elaboración propia a partir de ISO/IEC DIS 12792:2024(en)

Asimismo, es importante señalar que la Comisión Europea ha mandado al Comité Europeo de Normalización (CEN) y al Comité Europeo de Normalización Electrotécnica (CENELEC) para que elaboren diez normas técnicas que concreten las obligaciones principales previstas en el RIA para los sistemas de alto riesgo, entre las que se incluye una norma técnica específica para la transparencia.¹⁵

II. DESCRIPCIÓN SISTEMÁTICA DEL ARTÍCULO 13: DIMENSIONES DE LA TRANSPARENCIA

El requisito de transparencia previsto en el artículo 13 RIA tiene distintas dimensiones, subjetiva, formal y material.¹⁶ Estas dimensiones de la transparencia siguen un enfoque similar al de otras legislaciones generales o sectoriales que establecen obligaciones de transparencia formal y sustantiva a determinados sujetos incluidos en el ámbito de aplicación de la norma, y más concretamente el Reglamento

15. Comisión Europea, *Commission Implementing Decision of 22.5.2023 on a standardisation request to the European Committee for Standardisation and the European Committee for Electrotechnical Standardisation in support of Union policy on artificial intelligence*, C(2023) 3215 final. Anexo I. En el listado de nuevas normas técnicas europeas y documentos europeos de normalización que la Comisión ha encargado al CEN y al CENELEC, se incluyen las siguientes. Norma(s) europea(s) y/o documento(s) europeo(s) de normalización sobre (1) sistemas de gestión de riesgos para sistemas de IA; (2) gobernanza y calidad de los conjuntos de datos utilizados para construir sistemas de IA; (3) mantenimiento de registros mediante capacidades de registro de eventos (logs) de los sistemas de IA; (4) transparencia e información a los usuarios de los sistemas de IA; (5) supervisión humana de los sistemas de IA; (6) especificaciones de precisión para los sistemas de IA; (7) especificaciones de robustez de los sistemas de IA; (8) especificaciones de ciberseguridad para sistemas de IA; (9) sistemas de gestión de la calidad para proveedores de sistemas de inteligencia artificial, incluidos los procesos de seguimiento post-comercialización; (10) evaluación de la conformidad de los sistemas de IA.

16. Cfr. Kiseleva, A. *et al*, «Transparency of AI in Healthcare...», pp. 4-5.

General de Protección de Datos («RGPD»¹⁷). Sin embargo, respecto de los sistemas de alto riesgo que tengan un impacto individual en los derechos fundamentales de las personas físicas o efectos jurídicos similares, la doctrina considera que el artículo 13, en particular, y el Reglamento, en general, no ampliarían de manera significativa el contenido de las obligaciones de información de los Artículos 13-15 o de las garantías del 22 del RGPD.¹⁸

17. Cfr. Sovrano, F., Sapienza, S., Palmirani, M., Vitali, F., *Metrics, Explainability and the European AI Act Proposal*, J — *Multidisciplinary Scientific Journal*, vol. 5, n.º 1 (2022), p. 131. <https://doi.org/10.3390/j5010010> Nótese que la sistemática seguida por el Artículo 13 del RIA es muy similar a la de los Artículos 12 a 14 del Reglamento General de Protección de Datos («RGPD»). Mientras que el Artículo 12(1) del RGPD se refiere a la obligación de transparencia formal del responsable con relación al cómo debe facilitarse a los interesados la información prevista en los Artículos 13-15, según proceda (de forma concisa, transparente, inteligible y de fácil acceso, con un lenguaje claro y sencillo, por escrito o por otros medios); los Artículos 13, 14 y 15 determinan el *qué*, es decir, el contenido material de la información que debe proporcionarse a los interesados cuando los datos personales hayan sido recabados del propio interesado o de un tercero distinto al interesado. Precisamente, entre la información a facilitar por el responsable a los interesados, se incluye la existencia de decisiones automatizadas, incluida la elaboración de perfiles, previstas en el Artículo 22 del RGPD, y, como mínimo, la «información significativa sobre la lógica aplicada, así como la importancia y las consecuencias previstas de dicho tratamiento para el interesado.»
18. Hacker, P., Passoth, JH., «Varieties of AI Explanations Under the Law. From the GDPR to the AIA, and Beyond», Holzinger, A., Goebel, R., et al. (eds) *xxAI — Beyond Explainable AI. xxAI 2020. Lecture Notes in Computer Science*, vol 13200 (2022). Springer, Cham, p. 361. https://doi.org/10.1007/978-3-031-04083-2_17 Frente al planteamiento de los autores, sí que deben introducirse dos matices respecto de las relaciones entre el Reglamento de Inteligencia Artificial y la normativa europea de protección de datos. En primer lugar, debe tenerse en cuenta que, según el Artículo 18(9) del Reglamento, la información facilitada por el proveedor del sistema de alto riesgo conforme al Artículo 13 se utilizará por el responsable del despliegue para cumplir con la obligación de llevar a cabo una evaluación de impacto relativa a la protección de datos que le imponen el Artículo 35 del Reglamento (UE) 2016/679 o el Artículo 27 de la Directiva (UE) 2016/680. En segundo lugar, el ámbito de aplicación del derecho a una explicación que reconoce el Considerando (71) del RGPD con relación a las decisiones individuales totalmente automatizadas, incluida la elaboración de perfiles, con efectos jurídicos o similares del Artículo 22 sí que se vería ampliado, en la medida en que el derecho a una explicación con relación a las decisiones individuales adoptadas por los sistemas de alto riesgo del Anexo III, previsto en el Artículo 86 RIA, será de aplicación en aquellos casos en que este derecho «no esté previsto de otro modo en el Derecho de la Unión» (apartado 3). Una interpretación sistemática del Artículo 86(3) RIA lleva a concluir que el derecho a una explicación se ampliaría a las decisiones individuales, incluida la elaboración de perfiles, que no estén totalmente automatizadas, al existir supervisión humana, siempre que dichos tratamientos de datos personales tengan un impacto en la salud, la seguridad o los derechos fundamentales de las personas individuales. Precisamente, las que estaría excluyendo el Reglamento del alcance del Artículo 22 y del Considerando (71). Cfr. Laux, J. «Institutionalised distrust and human oversight of artificial intelligence: towards a democratic design of AI governance under the European Union AI Act», *AI & Society. Knowledge, culture and Communication* (2023), p. 5. <https://doi.org/10.1007/s00146-023-01777-z> El autor sostiene que, a diferencia del enfoque del GPDR con relación a los sistemas totalmente automatizados,

El ámbito subjetivo de aplicación del requisito de transparencia está contenido en el artículo 13, apartado (1), y comprende la identificación de los sujetos obligados (proveedores) y los sujetos destinatarios de la transparencia (responsables del despliegue). Asimismo, las manifestaciones de la «transparencia formal» se refieren a la forma en cómo debe presentarse el contenido material de las obligaciones de información previstas en el artículo 13 y se encuentran recogidas a lo largo de los apartados (1) y (2) del artículo 13. Finalmente, la «transparencia material o sustantiva» comprendería las categorías de información a incluir en las instrucciones de uso por parte del proveedor del sistema de IA de alto riesgo y descritas en el apartado (3) del artículo 13.

A partir de este esquema, la determinación del contenido y alcance del artículo 13 exige su integración con otros preceptos del RIA, fundamentalmente, con las definiciones contenidas en el Artículo 3, los preceptos de la Sección 2ª del Capítulo III del Reglamento, su Anexo IV. A efectos hermenéuticos, en la Tabla 2 siguiente se incorporan las principales correspondencias entre el artículo 13, con otras disposiciones contenidas en el RIA, y una relación de normas técnicas a efectos de concretar el contenido y alcance del requisito de transparencia.

el Reglamento de Inteligencia Artificial sería de aplicación también a los sistemas parcialmente automatizados, donde existe intervención humana.

Tabla 3
Correspondencia entre artículo 13 RIA, Articulado RIA y Normalización

Transparencia subjetiva y formal	Transparencia y comunicación de información a los responsables del despliegue	artículo 13	Articulado RIA	Normas Técnicas de referencia
	<p>Ámbito subjetivo de aplicación y nivel de transparencia que garantice interpretabilidad del sistema y cumplimiento de obligaciones</p>	<p>1. Los sistemas de IA de alto riesgo se diseñarán y desarrollarán de un modo que garantice que funcionan con un <i>nivel de transparencia suficiente</i> para que los responsables del despliegue <i>interpretan</i> y usen correctamente su información de salida. Se garantizará un <i>tipo y un nivel de transparencia adecuados para que el proveedor y el responsable</i> del despliegue cumplan las obligaciones oportunas previstas en la sección 3.</p>	<p>Artículo 3(2) (definición de «Proveedor»); Artículo 3(4) (definición de «Responsable del despliegue»).</p>	<p>Concepto de «transparencia» [ISO/IEC 22989:2022, 3.4.14, 3.5.15, 5.15.8; ISO/IEC DIS 12792:2024(en), 5.3; ISO/IEC TS 5723:2022(en), 3.2.19, 3.2.20; ETSI GR SAI 007 V1.1.1 (2023-03), 3, 4; IEEE Std 7001-2021, 4.1; NISTIR 8269, 3.99]. Niveles de transparencia según rol de partes interesadas [ISO/IEC 22989:2023, 5.19; ISO/IEC DIS 12792:2024(en), 6.2, IEEE Std 7001-2021, 5]. Concepto de «interpretabilidad» [ISO/IEC CD TS 6254, en desarrollo]</p>

Transparencia formal	Transparencia y comunicación de información a los responsables del despliegue	artículo 13	Articulado RIA	Normas Técnicas de referencia
Instrucciones de uso y características de la información proporcionada	<p>2. Los sistemas de IA de alto riesgo irán acompañados de las instrucciones de uso correspondientes en un <i>formato digital o de otro tipo adecuado</i>, las cuales incluirán <i>información concisa, completa, correcta y clara</i> que sea <i>pertinente, accesible y comprensible</i> para los responsables del despliegue.</p>		<p>Artículo 3(15) (definición de «instrucciones de uso») Artículo 11 (documentación técnica)</p>	<p>Presentación y adecuación de la información y ejemplos de transparencia en sistemas de IA [ISO/IEC DIS 12792:2024(en), 5.3, Anexo A; ETSI GR SAI 007 V1.1.1 (2023-03), 4, 5, 6]</p>
Transparencia material	Contenido de las instrucciones de uso	<p>3. Las instrucciones de uso contendrán <i>al menos</i> la siguiente información:</p> <p>(a) la <i>identidad y los datos de contacto del proveedor</i> y, en su caso, de su representante autorizado;</p>	<p>Artículo 3.3, 3.4., 3.3 (conceptos de proveedor, responsable del despliegue, representante autorizado).</p>	<p>Definición de categorías o roles de partes interesadas [ISO/IEC 22989:2023 [en], 5.19]</p>
Información sobre el proveedor del sistema				

	<p>artículo 13 Transparencia y comunicación de información a los responsables del despliegue</p>	Articulado RIA	Normas Técnicas de referencia
<p>Idoneidad funcional</p>	<p>(b) las características, capacidades y limitaciones del <i>funcionamiento</i> del sistema de IA de alto riesgo, y en particular, cuando proceda:</p>	<p>Artículo 3(18) (definición de «funcionamiento de un sistema de IA»).</p>	<p>Concepto de «idoneidad funcional» (en software y en sistemas de IA) [ISO/IEC 25010:2023(en), 3.1; ISO/IEC 25059:2023(en), 5.1]. Taxonomías de niveles de transparencia [ISO/IEC DIS 12792:2024(en)]; nivel-contexto, nivel-sistema, nivel-modelo, nivel datos].</p>
<p>Finalidad</p>	<p>(i) su <i>finalidad</i> prevista;</p>	<p>Artículo 3(12) (definición de «finalidad prevista»).</p>	<p>Taxonomía «nivel-sistema» (aplicabilidad) [ISO/IEC DIS 12792:2024(en), 8.4]. Taxonomía nivel-modelo [ISO/IEC DIS 12792:2024(en), 9.3.3].</p>

	<p>Transparencia y comunicación de información a los responsables del despliegue</p> <p>artículo 13</p>	<p>Articulado RIA</p>	<p>Normas Técnicas de referencia</p>
<p>Rendimiento predictivo (y sus métricas), solidez y ciberseguridad</p>	<p>(ii) el <i>nivel de precisión</i> (incluidos los <i>parámetros</i> para evaluarla), solidez y ciberseguridad mencionado en el artículo 15 con respecto al cual se haya probado y validado el sistema de IA de alto riesgo y que puede esperarse, así como cualquier circunstancia conocida y previsible que pueda afectar al nivel de precisión, solidez y ciberseguridad esperado;</p>	<p>Artículo 15 (precisión, solidez y ciberseguridad).</p>	<p>Concepto de «corrección funcional» en sistemas de IA [ISO/IEC 25059:2023(EN), 3.2.3, 5.4], ISO/IEC DIS 12792:2024(en), 9.6]. Métricas de rendimiento en ML [ISO/IEC 23053:2022(en), 6.5.5.]. Concepto de «solidez» en sistemas de IA [ISO/IEC 25059:2023(en), 3.2.5, 5.5]; en redes neuronales y métodos para su medición [ISO/IEC TR 24029-1:2021(en), 3.6, 4.1.1, 5, 6, 7].</p> <p>Asegurar la Inteligencia Artificial (SAI): planteamiento del problema [ETSI GR SAI 004 V1.1.1 (2020-12)]; plataformas de computación de IA [ETSI GR SAI 009 V1.1.1 (2023-02)].</p>

	<p>Transparencia y comunicación de información a los responsables del despliegue</p>	<p>artículo 13</p>	<p>Articulado RIA</p>	<p>Normas Técnicas de referencia</p>
<p>Circunstancias que den lugar a riesgos para la salud, la seguridad o los derechos fundamentales</p>	<p>(iii) cualquier <i>circunstancia conocida o previsible</i>, asociada a la utilización del sistema de IA de alto riesgo conforme a su finalidad prevista o a un uso indebido razonablemente previsible, que pueda dar lugar a <i>riesgos para la salud y la seguridad o los derechos fundamentales</i> a que se refiere el artículo 9, apartado 2;</p>	<p>(iii) cualquier <i>circunstancia conocida o previsible</i>, asociada a la utilización del sistema de IA de alto riesgo conforme a su finalidad prevista o a un uso indebido razonablemente previsible, que pueda dar lugar a <i>riesgos para la salud y la seguridad o los derechos fundamentales</i> a que se refiere el artículo 9, apartado 2;</p>	<p>Artículo 3(13) (definición de «uso indebido razonablemente previsible»).</p>	<p>Taxonomía del «nivel-contexto» [ISO/IEC DIS 12792:2024(EN), 7] y del nivel del conjunto de datos [SO/IEC DIS 12792:2024(en), 10.6, 10.7].</p>
<p>Explicabilidad</p>	<p>(iv) en su caso, las <i>capacidades y características técnicas</i> del sistema de IA de alto riesgo para proporcionar <i>información pertinente para explicar</i> su información de salida;</p>	<p>(iv) en su caso, las <i>capacidades y características técnicas</i> del sistema de IA de alto riesgo para proporcionar <i>información pertinente para explicar</i> su información de salida;</p>	<p>Anexo IV 2(a), con relación a métodos y medidas. Anexo IV 2(b), con relación a lógica, hipótesis, parámetros, compensaciones.</p>	<p>Explicabilidad y transparencia: ETSI GR SAI 007 V1.1.1 (2023-03).</p>
<p>Idoneidad funcional con relación a personas/grupos afectados</p>	<p>(v) cuando proceda, su <i>funcionamiento con respecto a personas o grupos de personas específicos</i> en relación con los que esté previsto utilizar el sistema;</p>	<p>(v) cuando proceda, su <i>funcionamiento con respecto a personas o grupos de personas específicos</i> en relación con los que esté previsto utilizar el sistema;</p>	<p>Anexo IV. 3</p>	<p>Taxonomía del nivel-contexto [ISO/IEC DIS 12792:2024(en), 7].</p>

	<p>artículo 13</p> <p>Transparencia y comunicación de información a los responsables del despliegue</p>	Artículo RIA	Normas Técnicas de referencia
<p>Especificaciones relativas a datos de entrada y datasets de entrenamiento, validación y prueba</p>	<p>(vi) cuando proceda, especificaciones relativas a los <i>datos de entrada</i>, o cualquier otra información pertinente en relación con los <i>conjuntos de datos de entrenamiento, validación y prueba usados</i>, teniendo en cuenta la finalidad prevista del sistema de IA;</p>	<p>Artículo 3(29) (definición de «Datos de entrenamiento»); Artículo 3(30) (definición de «Datos de validación»); Artículo 3(31) (definición de «Datos de prueba»); Artículo 3(32) (definición de «Datos de entrada»); Artículo 10 (datos y gobernanza de datos) Anexo IV. 2(d), con relación a datos entrada. Anexo IV. 2(g), con relación datos de validación y prueba.</p>	<p>Taxonomía del «nivel-conjunto de datos» [ISO/IEC DIS 12792:2024(en), 10]. Segos en sistemas de IA [ISO/IEC/TR 24027:2021].</p>
<p>Interpretabilidad</p>	<p>(vii) en su caso, <i>información que permita a los responsables del despliegue interpretar</i> la información de salida del sistema de IA de alto riesgo y utilizarla adecuadamente.</p>	<p>Anexo IV. 2</p>	<p>Transparencia y explicabilidad [IEEE Std 7001-2021, 3.1, 4.1; ETSI GR SAI 007 V1.1.1 (2023-03); NISTIR 8312 (2021)].</p>

	<p>artículo 13 Transparencia y comunicación de información a los responsables del despliegue</p>	Articulado RIA	Normas Técnicas de referencia
<p>Modificaciones e idoneidad funcional por defecto</p>	<p>(c) los cambios en el sistema de IA de alto riesgo y su funcionamiento <i>predefinidos</i> por el proveedor en el momento de efectuar la evaluación de la conformidad inicial, en su caso;</p>	<p>Artículo 3(23) (definición de «Modificación sustancial»); Anexo IV. 2(f). Anexo IV. 6. Anexo IV. 8.</p>	<p>ISO /IEC DIS 12792:2024(en), 8.4, 8.7, 9.4.9.</p>
<p>Medidas de vigilancia humana para facilitar la interpretabilidad/explicabilidad</p>	<p>(d) las <i>medidas de vigilancia humana</i> a que se hace referencia en el artículo 14, incluidas <i>las medidas técnicas establecidas para facilitar la interpretación</i> de la información de salida de los sistemas de IA de alto riesgo por parte de los responsables del despliegue;</p>	<p>Artículo 14 (vigilancia humana). Anexo IV.3, con relación a medidas técnicas establecidas para facilitar la interpretación de la información de salida. Anexo IV. 2(e) Anexo IV. 3.</p>	<p>ISO /IEC DIS 12792:2024(en), 8.5.7.</p>

	<p>Transparencia y comunicación de información a los responsables del despliegue</p> <p>artículo 13</p>	<p>Articulado RIA</p>	<p>Normas Técnicas de referencia</p>
<p>Eficiencia del rendimiento</p>	<p>(e) los recursos informáticos y de hardware necesarios, la vida útil prevista del sistema de IA de alto riesgo y las medidas de mantenimiento y cuidado necesarias (incluida su frecuencia) para garantizar el correcto funcionamiento de dicho sistema, también en lo que respecta a las actualizaciones del software;</p>	<p>Anexo IV. 1(c), respecto de actualizaciones. Anexo IV. 2(c), con relación a recursos computacionales.</p>	<p>ISO/IEC DIS 12792:2024(en), 9.4.7, 9.4.8, 9.4.9.</p>
<p>Trazabilidad</p>	<p>(f) cuando proceda, una descripción de los mecanismos incluidos en el sistema de IA de alto riesgo que permitir a los responsables del despliegue recabar, almacenar e interpretar correctamente los archivos de registro de conformidad con el artículo 12.</p>	<p>Artículo 12 (Registros log). Anexo IV 2(g) (test logs).</p>	<p>Registro de eventos [ISO/IEC DIS 12792:2024(en), 8.5.4]</p>

Fuente: Elaboración propia.

Como puede apreciarse en la Tabla 3, dado el carácter altamente técnico del artículo 13, y en espera de que el CEN y el CENELEC desarrollen la correspondiente norma técnica en aplicación de la Decisión de implementación de la Comisión Europea de 22 de mayo de 2023 (C(2023) 3215 final), la delimitación del contenido y alcance del artículo 13 se ha completado a partir de otras normas técnicas publicadas por otros organismos de normalización relativas a sistemas de IA.¹⁹

La doctrina ha sido crítica con la tendencia de la Comisión Europea, cada vez más generalizada y discutible, de delegar el proceso de concreción de las normas jurídicas en organismos de derecho privado (modelo de suscripción y retención de propiedad intelectual, mayor exposición al lobismo, ausencia de control democrático, perjuicio de los consumidores europeos y de pequeños desarrolladores para el acceso a las normas).²⁰

En todo caso, las posibles interpretaciones del artículo 13 del Reglamento que se incluyen en este Capítulo podrían verse ampliadas, matizadas e incluso corregidas por el contenido de la norma técnica sobre «transparencia y comunicación de información» que elaboren el CEN y el CENELEC.

III. TRANSPARENCIA, INTERPRETABILIDAD Y EXPLICABILIDAD: SU TRATAMIENTO (A)SISTEMÁTICO EN EL ARTÍCULO 13

Aunque desde el ámbito científico-técnico se han producido intentos para establecer una independencia conceptual entre los conceptos de «transparencia»,

-
19. En particular, se han tenido como referencia las siguientes normas técnicas: ISO/IEC DIS 12792:2024(en) (taxonomía de transparencia de sistemas de IA); ISO/IEC 25059:2023(E) (Calidad y Evaluación de sistemas de IA); ISO/IEC 25010:2023(en) (Calidad y Evaluación de sistemas y software); ISO/IEC 22989:2022 (conceptos y terminología de IA); ISO/IEC TS 5723:2022(en) (concepto de transparencia); ISO/IEC 23053:2022(E) (marco de sistemas basados en ML); ISO/IEC/TR 24027:2021 (sesgos) ISO/IEC TR 24029-1:2021(E) (solidez de sistemas de IA); ETSI GR SAI 007 V1.1.1 (2023-03) (explicabilidad y transparencia); ETSI GR SAI 009 V1.1.1 (2023-02) (seguridad en sistemas de IA); ETSI GR SAI 004 V1.1.1 (2020-12) (seguridad en plataformas de computación de IA); IEEE Std 7001-2021 (transparencia); NISTIR 8312 (2021) (explicabilidad) NISTIR 8269 (2019) (conceptos ML).
20. La doctrina ha señalado que será en la normalización donde va a tener lugar la verdadera elaboración de normas que concreten la aplicación del RIA. Se critica, sin embargo, la tendencia de la Comisión Europea, cada vez más generalizada y discutible, de delegar el proceso de concreción de las normas jurídicas en organismos de derecho privado (modelo de suscripción y retención de propiedad intelectual, mayor exposición al lobismo, ausencia de control democrático, perjuicio de los consumidores europeos y de pequeños desarrolladores para el acceso a las normas). Véanse, entre otros, Schneeberger, R. Röttger, F. Cabitza *et al.*, «The Tower of Babel...», *Op. cit.*, 75; Smuha, N. A., Ahmed-Rengers, Emma *et al.*, *How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act*, 5 de agosto de 2021, p. 54. <http://dx.doi.org/10.2139/ssrn.3899991>; Veale, M., Zuiderveen, B. F., «Demystifying the Draft EU Artificial Intelligence Act — Analysing the good, the bad, and the unclear elements of the proposed approach», *Computer Law Review International*, vol. 22 (2021), pp. 97-112. <http://dx.doi.org/10.9785/cr-2021-220402> Los autores critican la tendencia de la Comisión Europea, cada vez más generalizada y discutible, de delegar el proceso de concreción de las normas jurídicas en organismos de derecho privado, en claro perjuicio de los consumidores europeos. De hecho, señalan que será en la normalización donde va a tener lugar la verdadera elaboración de normas que concreten la aplicación del RIA.

«interpretabilidad» y «explicabilidad», no existe un consenso general en cuanto a su significado.²¹ Así, por ejemplo, la transparencia a menudo se solapa con otras propiedades técnicas, como la reproducibilidad²², la trazabilidad, la verificabilidad, la usabilidad, la explicabilidad y la interpretabilidad, la rendición de cuentas, la calidad o la fiabilidad del sistema.²³

Especialmente, en el dominio de la XAI resulta frecuente que los conceptos de «transparencia» e «interpretabilidad»²⁴ o de «interpretabilidad» y «explicabilidad»²⁵ se utilicen de manera indistinta respectivamente. También hay quienes consideran la «explicabilidad» como una parte integrante de la «transparencia»²⁶. Mientras que, en otros casos, se busca una diferenciación conceptual, identificando las correlaciones existentes en los mismos.²⁷

21. UK Parliament POST, «Interpretable machine learning», *Postnote*, n.º 633, The Parliamentary Office of Science and Technology, Westminster, Londres, octubre de 2020. <https://post.parliament.uk/research-briefings/post-pn-0633/>
22. Deben diferenciarse los conceptos de «reproducibilidad» y «replicabilidad». En el área del machine learning, el proceso de entrenamiento es «reproducibile», si bajo la misma configuración de entrenamiento (por ejemplo, el mismo conjunto de datos de entrenamiento, código, entorno), el modelo entrenado produce los mismos resultados bajo los mismos criterios de evaluación. Los criterios de evaluación pueden definirse para una muestra de datos (por ejemplo, resultados de inferencia) o sobre una distribución de datos (por ejemplo, métricas de rendimiento). La reproducibilidad del proceso de entrenamiento del modelo para la eliminación de errores, la evaluación y la trazabilidad del modelo, así como la auditoría y la verificación de las reclamaciones. La «reproducibilidad» debe distinguirse de la «replicabilidad», que significa que bajo una muestra de datos diferente (con la misma distribución que la muestra de datos original) combinada con el código y el análisis originales se obtienen resultados similares. Véase, ETSI TR 104 032 V1.1.1 (2024-02), p. 26.
23. Cfr. ISO/IEC DIS 12792:2024(en). Information technology —Artificial intelligence— Transparency taxonomy of AI systems [en tramitación actualmente], 5.3.
24. Barredo Arrieta, A., Díaz-Rodríguez, N., *et al.*, «Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI», en «Information Fusion», vol. 58, 2020, p. 84. <https://doi.org/10.1016/j.inffus.2019.12.012>
25. Véase, HLEG, *The assessment list for trustworthy Artificial Intelligence (ALTAI) for self assessment*, Comisión Europea, 17 de julio de 2020, p. 27; Molnar, Ch., *Interpretable machine learning: A guide for making black box models explainable*, 2ª ed., 2020. <https://christophm.github.io/interpretable-ml-book/>; Carvalho, D. V.; Pereira, E. M.; Cardoso, J. S. «Machine Learning Interpretability: A Survey on Methods and Metrics», *Electronics*, vol. 8, n.º 8, 832 (2019), p. 7. <https://doi.org/10.3390/electronics8080832>; Adadi, Amina y Berrada, Mohammed, «Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)», *IEEE Access*, vol. 6 (2018), pp. 52141-52142. DOI: 10.1109/ACCESS.2018.2870052.
26. Winfield, Alan F. T., Booth, Serena, Dennis, Louise A., *et al.* «IEEE P7001: A Proposed Standard on Transparency», *Frontiers in Robotics and AI*, vol. 8, 2021, p. 3. DOI: 10.3389/frobt.2021.665729.
27. Cfr. Doshi-Velez, Finale y Kim, Been, *Towards A Rigorous Science of Interpretable Machine Learning*, 2 de marzo de 2017, p. 1. <https://arxiv.org/abs/1702.08608>; Lepri, Bruno, Oliver, Nuria, Letouzé, Emmanuel, *et al.* «Fair, Transparent, and Accountable Algorithmic Decision-making Processes», *Philosophy & Technology*, vol. 31, 2018, pp. 619-622. <https://doi.org/10.1007/s13347-017-0279-x>; Rudin, C., *Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead*, p. 2, 2019. <https://arxiv.org/abs/1811.10154>; Mittelstadt, B., Russell, C. y Wachter, S., «Ex-

Finalmente, parte de la literatura científica identifica como objetivo primordial el desarrollo de modelos de IA interpretables (entendibles o inteligibles para un observador humano) a partir de la información que proporciona el modelo en sí mismo (transparencia)²⁸ o de técnicas complementarias (normalmente, otros modelos algorítmicos) que permiten extraer explicaciones (explicabilidad) en aquellos casos en que el modelo no es interpretable por un humano (cajas negras)²⁹. De hecho, esta es la interpretación que se acoge en este capítulo.

Desde el punto de vista del *soft law*, el trabajo de distinción conceptual tampoco ha corrido mucha mejor suerte. A nivel europeo, por ejemplo, la confusión terminológica descrita se ha hecho patente en algunos de los documentos elaborados por el Grupo de Expertos de Alto Nivel de la Comisión Europea (en adelante, por sus siglas en inglés «AI HLEG»), como las «Directrices Éticas para una IA Fiable» (en adelante, «Directrices Éticas»)³⁰ o del Listado de Auto-Evaluación («ALTAI», por sus siglas en inglés)³¹. Pues bien, toda esa

-
- plaining explanations in AI», *Proceedings of Fairness, Accountability, and Transparency (FAT*)*, ACM Digital Library, 2019, p. 280. <https://doi.org/10.1145/3287560.3287574>; Longo, L.; Brcic, M.; Cabitza, F. *et al.* «Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions», *Information Fusion*, vol. 106 (2024). <https://doi.org/10.1016/j.inffus.2024.102301>
28. Rudin, C., Chen, C., Chen, Z., *at al.* «Interpretable Machine Learning: Fundamental Principles and 10 Grand Challenges», *Statistic Surveys*, vol. 16 (2022), pp. 1-16. DOI: 10.1214/21-SS133.
 29. Information Commissioner's Office, Alan Turing Institute, *Explaining decisions...*, *Op. cit.*, p. 69.
 30. Cfr. HLEG, *Directrices Éticas para una IA Fiable*, Bruselas, Comisión Europea, 8 de abril de 2019. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> Debe tenerse en cuenta que las Directrices Éticas no incluyen una definición clara de estos conceptos ni tampoco establecen una correlación adecuada entre los mismos. Así, por ejemplo, las Directrices incurren en una aproximación cuasi-tautológica a los conceptos de «explicabilidad» y «transparencia». Y así, mientras el principio ético de explicabilidad («explicability», en su versión en inglés) comprendería, entre otros aspectos, la transparencia de los procesos y la capacidad de poder explicar las decisiones a las partes afectadas; el requisito de transparencia incluiría, a su vez, la «explicabilidad» («explainability», en su versión en inglés). No está claro, en todo caso, el sentido y alcance de la diferente terminología empleada en la versión inglesa, «explicability» (principio ético) y la «explainability» (elemento de la transparencia), que en la versión en español aparecen traducidos de la misma manera. De hecho, los términos «explicability» y «explainability» no aparecen recogidos en los distintos diccionarios de referencia en inglés (Cambridge, Oxford, Merriam-Webster), aunque sí los términos «explication» y «explanation», cuyos respectivos significados incluyen sutiles diferencias. Por su parte, el concepto de «interpretabilidad» está prácticamente ausente en los principios y requisitos de las Directrices (ni se vincula con el principio ético de explicabilidad, ni con el requisito de transparencia), si bien se introduce un principio de interpretabilidad desde el diseño (desde la concepción del sistema) y por defecto (adopción de los modelos más simples e interpretables posibles) que aparece asociado al listado de verificación de la explicabilidad (como elemento del requisito de transparencia).
 31. HLEG, *The assessment list for trustworthy Artificial Intelligence (ALTAI) for self assessment*, Comisión Europea, 17 de julio de 2020, pp. 26 y 27. El documento del Grupo de Expertos no define la «transparencia», pero sí la «explicabilidad» y la «interpretabilidad». La «explicabilidad» sería entonces aquella «propiedad» o «característica de un

indefinición y confusión terminológica ha acabado recogida en buena medida en el RIA³², que ha incorporado como parte de su acervo interpretativo las Directrices Éticas.³³

1. SIGNIFICADO Y TIPOS DE TRANSPARENCIA EN EL REGLAMENTO Y EN EL ARTÍCULO 13

Desde el punto de vista técnico, no existe un concepto uniforme de la «transparencia» de los modelos de IA. En el ámbito de la XAI, la transparencia se ha venido identificando con el grado de interpretabilidad intrínseca que tiene un modelo de IA.³⁴ Desde el ámbito de la normalización el concepto se refiere bien a la comunicación o puesta a disposición de las partes interesadas de información relevante sobre el sistema³⁵, bien al grado de apertura de un sistema para que pueda ser inspeccionado y no tener partes ocultas.³⁶

Asimismo, la ISO/IEC DIS 12792:2024(en) distingue entre la transparencia como propiedad técnica del sistema de IA («transparencia del sistema») y la transparencia como propiedad de la organización («transparencia organizacional»). En cuanto a la transparencia del sistema de IA, ésta significa que se pone a disposición de las partes interesadas la información relevante sobre el sistema. Dicha información relevante puede incluir: los objetivos del sistema, sus limitaciones conocidas, elecciones de diseño y premisas previas, características, modelos, algoritmos, métodos de entrenamiento, detalles de los datos utilizados y procesos de garantía de calidad. Sin embargo, las necesidades de transparencia pueden ser diferentes para las distintas partes interesadas. Respecto de la transparencia organizacional, está relacionada con la forma en que las actividades y decisiones se comunican a las partes interesadas pertinentes y su relevancia viene determinada porque los principios

sistema de IA que es inteligible para los no expertos», de manera que un sistema de IA es inteligible «si su funcionalidad y operaciones pueden explicarse de forma no técnica a una persona no experta en la materia.» El concepto de «interpretabilidad», por su parte, parece asimilarse al de «explicabilidad», al identificarse con la «comprensibilidad, explicabilidad o entendibilidad», de manera que «cuando un elemento de un sistema de IA es interpretable, significa que es posible, al menos para un observador externo, entenderlo y encontrar su significado».

32. Kiseleva, A. *et al*, «Transparency of AI in Healthcare...», pp. 2-3.

33. Cfr. Considerandos (7), (27) y (165) del RIA.

34. Un modelo de IA se considera transparente si el funcionamiento global del modelo («simulabilidad»), de sus componentes individuales («descomponibilidad») y de su algoritmo de aprendizaje («transparencia algorítmica») resultan inteligibles o comprensibles para un humano. Por tanto, la transparencia general de un modelo dependerá de un adecuado equilibrio entre estos tres niveles. Cfr. Lipton, Z. C. «The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery». *ACM Queue*, vol. 16, n.º 3 (2018), p. 12; Lepri *et al.*, «Fair, transparent...», *Op. cit.*, p. 619; Barredo *et al.*, *Explainable Artificial Intelligence (XAI)*... *Op. cit.*, pp. 88-100; Information Commissioner's Office y Alan Turing Institute, *Explaining decisions...*, pp. 61-63, 115-118; Mittelstadt, B.; Russell, Chris; Wachter, S., «Explaining Explanations...», *Op. cit.*, p. 280.

35. ISO/IEC 22989:2022, 3.5.15, 5.15.8.

36. ETSI GR SAI 007 V1.1.1 (2023-03), p. 6.

y procesos organizativos subyacentes afectan al sistema de IA a lo largo de su ciclo de vida.³⁷

En conexión con este concepto de «transparencia del sistema», es importante señalar que la reciente norma ISO/IEC 25059:2023(en), sobre el modelo calidad de los sistemas de IA, diferencia los sistemas transparentes de los opacos en función de cómo los sistemas documenten, registren o muestren información sobre sus procesos técnicamente.³⁸

A diferencia de lo que ocurre con la interpretabilidad y la explicabilidad, el Reglamento acoge una definición de «transparencia» en su Considerando (27) que tiene su fundamento en las Directrices Éticas del Grupo de Expertos de la Comisión.³⁹

«De acuerdo con las directrices del Grupo independiente de expertos de alto nivel sobre IA [...] [p]or “transparencia” se entiende que los sistemas de IA se desarrollan y utilizan de un modo que permita una trazabilidad y explicabilidad adecuadas, y que, al mismo tiempo, haga que las personas sean conscientes de que se comunican o interactúan con un sistema de IA e informe debidamente a los responsables del despliegue acerca de las capacidades y limitaciones de dicho sistema de IA y a las personas afectadas acerca de sus derechos».

En realidad, este planteamiento no contiene una definición propiamente dicha de la «transparencia» sino más bien una identificación los elementos integrantes de la misma, a saber: la trazabilidad, la explicabilidad y la comunicación de información relevante a los responsables del despliegue y a las personas expuestas a los sistemas de IA. Además de una falta de definición propiamente dicha de la «transparencia», deben destacarse dos limitaciones importantes en la definición incorporada en la parte expositiva del Reglamento.

En primer lugar, sorprende la ausencia del requisito de interpretabilidad de los sistemas de alto riesgo en el Considerando (27), a menos que, para el legislador europeo, transparencia e interpretabilidad sean la misma cosa.⁴⁰ En segundo lugar, la identificación de los elementos integrantes de la transparencia en la parte expositiva

37. ISO/IEC DIS 12792:2024(en), 5.3.

38. ISO/IEC 25059:2023(en), 5.6. Según la norma, los sistemas de IA transparentes documentan, registran o muestran sus procesos internos mediante herramientas de introspección y archivos de datos. El flujo de datos puede ser rastreable en cada paso, con decisiones aplicadas, excepciones y reglas documentadas. Es posible rastrear y registrar los procesos secuenciales a medida que varían los datos, así como los errores. Los sistemas de IA altamente transparentes pueden construirse a partir de subcomponentes bien documentados cuyas interfaces se describen explícitamente, lo que en última instancia facilita la investigación de los fallos del sistema. Por el contrario, un sistema poco transparente tiene un funcionamiento interno difícil de inspeccionar externamente. La falta de disponibilidad de registros de procesamiento detallados puede perjudicar la verificación y la evaluación del impacto social y ético y el tratamiento de los riesgos.

39. Cfr. Directrices Éticas, apartados 75-78.

40. Parece obvio que esta limitación tiene su origen en el enfoque que hacen las Directrices Éticas del Grupo de Expertos de la Comisión donde la interpretabilidad se encuentra ausente a la hora de delimitar los elementos de la transparencia (trazabilidad, explicabilidad y comunicación).

no se compadece, sin embargo, con el enfoque del Art. 13 del Reglamento, donde en cambio sí que aparece mencionada la interpretabilidad en varias ocasiones, pero no así la explicabilidad. Bien porque este último requisito aparece mencionado de manera indirecta (apartado 3.b iv), bien porque se encuentra diluido o confundido con el de interpretabilidad (apartado 3.d).

En lo que respecta a otros elementos integrantes de la definición de «transparencia» –trazabilidad y comunicación, incluidos en el Considerando (27), debe indicarse que la trazabilidad⁴¹ de los sistemas de IA se manifiesta a través de distintas obligaciones de documentación y de registro establecidas a lo largo del Reglamento⁴² previstas, entre otros, en los Artículos 11 (documentación técnica), 12 (conservación de registro de eventos logs), 13.3 (instrucciones de uso), 18 (deber de conservación por el proveedor de la documentación técnica por un periodo de 10 años).

Por su parte, las obligaciones de comunicación de información relevante previstas a lo largo del Reglamento permiten identificar dos tipos de transparencia, una transparencia interna de contenido y alcance muy técnico ([1], [2], [6]), y otra transparencia externa dirigida al público en general ([3], [4], [5]). En particular, la comunicación de información relevante está presente, entre otras disposiciones del Reglamento, en:

— La obligación del proveedor de poner a disposición al responsable del despliegue las instrucciones de uso con el contenido prescrito en el Art. 13(3) [1].

— La obligación de cooperación del proveedor con las autoridades competentes prevista en el Artículo 21 del Reglamento de proporcionar a dichas autoridades de toda la información y documentación necesarias para demostrar la conformidad del sistema de IA de alto riesgo con los requisitos establecidos en la sección 2ª del Capítulo III, así como el acceso a los archivos de registro generados automáticamente por el sistema de IA, en la medida en que dichos registros estén bajo el control del proveedor [2].

— La obligación del responsable del despliegue de informar a las personas físicas que están expuestas a los sistemas de alto riesgo del Anexo III contenida en el Artículo 26(11) RIA. En el caso particular de los sistemas de IA de alto riesgo que se utilicen por las autoridades competentes con fines de prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de sanciones penales, esta obligación de información habrá de modularse conforme a las previsiones contenidas en el artículo 13 de la Directiva (UE) 2016/680 [3].

— La obligación de información de proveedores y responsables del despliegue a las personas físicas que se relacionan con determinados sistemas de IA (sistemas que interactúan directamente con personas físicas, sistemas de IA generativa, sistemas de reconocimiento de emociones o categorización biométrica, sistemas de ultrafalsificación) previstas en el Artículo 50 [4].

41. Vid. ETSI TR 104 032 V1.1.1 (2024-02), pp. 13-14. La trazabilidad puede entenderse como el seguimiento de todo el ciclo de vida de un modelo, que incluye no sólo el rastreo de un modelo, sino también de sus datos y metadatos, así como los detalles del proceso de formación. La trazabilidad puede aplicarse al conjunto de datos de entrenamiento y prueba, los parámetros de entrenamiento, el modelo, y a la arquitectura subyacente del sistema.

42. Cfr. Considerando (71).

— El acceso público a la información contenida en la base de datos de la Comisión Europea para los sistemas de IA de alto riesgo enumerados en el ANEXO III prevista en el Artículo 71 [5], a excepción de la sección segura no pública a que se refieren los Artículos 49(4), y el Artículo 60(4)(c) [6].⁴³ Por su parte, el Anexo VIII del Reglamento enumera las categorías de información a incluir en la base de datos en función de la clasificación realizada por los Artículos 49 y 71 del Reglamento.⁴⁴

La transparencia prevista en el artículo 13 es una transparencia interna y de tipo técnico. Los fines de la transparencia que consigna el precepto vendrían determinados por los roles del proveedor y responsables del despliegue. En el caso del proveedor, el fin de la obligación de transparencia (activa) sería el cumplimiento normativo; mientras que en el caso del responsable del despliegue, la transparencia (pasiva) tiene por objeto no sólo el cumplimiento normativo sino una finalidad capacitadora en los términos que se explicará más adelante.

2. LA INTERPRETABILIDAD EN EL ARTÍCULO 13: ¿DESINCENTIVACIÓN DE LOS MODELOS DE CAJA NEGRA?

El concepto de «interpretabilidad» no está exento de confusión terminológica, pues en algunos casos se identifica o confunde con la explicabilidad. De hecho, hay quienes incluyen la explicabilidad como un elemento de la interpretabilidad.⁴⁵ La «interpretabilidad» es una característica *pasiva* del modelo de IA que se refiere al grado en que el comportamiento y los resultados de un modelo concreto son comprensibles o inteligibles para el observador humano. La interpretabilidad de un modelo es mayor si resulta fácil para una persona razonar y rastrear de una forma coherente por qué el

43. Fundamentalmente se excepcionan del acceso público, los sistemas de alto riesgo en el ámbito de la garantía del cumplimiento del Derecho y de la gestión de la migración, el asilo y el control fronterizo.

44. La base de datos se divide en tres Secciones (A, B, y C) accesibles. Así, por ejemplo, con relación a los sistemas del Anexo III, incluidos en la Sección A de la base de datos, con excepción de los sistemas relativos a infraestructuras críticas, el proveedor o su representante facilitarán, entre otra, la siguiente información: identificación y datos de contacto y localización del proveedor, o en su caso, del representante autorizado; el nombre comercial del sistema de IA y toda referencia inequívoca adicional que permita su identificación y trazabilidad; la descripción de la finalidad prevista del sistema de IA y de los componentes y funciones que se apoyan a través del mismo; una descripción sencilla y concisa de la información que utiliza el sistema (datos, entradas) y su lógica de funcionamiento; la situación del sistema de IA (comercializado o puesto en servicio, ha dejado de comercializarse o de estar en servicio, recuperado); una copia de la declaración UE de conformidad; las instrucciones de uso electrónicas; y, con carácter opcional, una URL para obtener información adicional.

45. Cfr. ALTAI, *Op. cit.*, p. 27. «La interpretabilidad se refiere al concepto de comprensibilidad, *explicabilidad* o entendibilidad. Cuando un elemento de un sistema de IA es interpretable, significa que es posible al menos para un observador externo entenderlo y encontrar su significado [cursiva nuestra]». Véase también, UK Parliament POST, «Interpretable machine learning», *Op. cit.*, p. 2, donde el Parlamento Británico señala que el concepto de «interpretabilidad» se suele utilizar para «describir la *capacidad de presentar o explicar el proceso de toma de decisiones* de un sistema de inteligencia de IA en términos comprensibles para los seres humanos (incluidos los desarrolladores de IA, usuarios, compradores, reguladores o afectados por las decisiones del sistema) [cursiva nuestra]».

modelo hizo, por ejemplo, una predicción concreta. En términos comparativos, dado un modelo A, éste será más interpretable que otro modelo B si las predicciones de A son más fáciles de entender que las realizadas por B.⁴⁶ Frente a la transparencia, la opacidad de un modelo (ya sea intencional o buscada por el diseñador, por falta de capacitación y habilidades técnicas o intrínseca al propio modelo) es conocida en la comunidad *machine learning* como el «problema de la interpretabilidad».⁴⁷

Desde el dominio técnico de la XAI, se diferencia así entre los modelos que son «interpretables por diseño» («modelos transparentes») de aquellos otros que, no siendo interpretables, *prima facie*, sin embargo, pueden ser explicables (y, por tanto, interpretables) mediante distintas técnicas consistentes en la extracción de información relevante del modelo y que generan explicaciones.⁴⁸ Las explicaciones generadas a partir del modelo pueden ser, a su vez, de distintos tipos (matemáticas, estadísticas, en lenguaje natural) dependiendo del destinatario de la explicación (autoridad, auditor externo, usuarios del sistema, afectados, público en general).

La relevancia de la explicabilidad se justifica por una aceptación generalizada de la relación inversa entre interpretabilidad y el rendimiento de los modelos de IA. Con ello se quiere significar que los modelos más simples suelen ser más interpretables, pero tienen una capacidad predictiva menor; y, al contrario.⁴⁹ Para resolver el problema de la

-
46. Barredo *et al.*, «Explainable Artificial Intelligence (XAI)...». *Op. cit.*, p. 84; Carvalho, D. V.; Pereira, E. M.; Cardoso, J. S., «Machine Learning Interpretability: A Survey on Methods and Metrics». *Electronics*, vol. 8, n.º 8: 832 (2019), p. 10; Molnar, C., *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*, Leanpub, 2019, sp.
47. Un modelo de IA se considera transparente si el funcionamiento global del modelo («simulabilidad»), de sus componentes individuales («descomponibilidad») y de su algoritmo de aprendizaje («transparencia algorítmica») resultan inteligibles o comprensibles para un humano. Por tanto, la transparencia general de un modelo dependerá de un adecuado equilibrio entre estos tres niveles. Cfr. Lipton, Z. C. «The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery». *ACM Queue*, vol. 16, n.º 3 (2018), p. 12; Lepri *et al.*, «Fair, transparent...», *Op. cit.*, p. 619; Barredo *et al.*, *Explainable Artificial Intelligence (XAI)...*. *Op. cit.*, pp. 88-100; Information Commissioner's Office y Alan Turing Institute, *Explaining decisions...*, pp. 61-63, 115-118; Mittelstadt, B.; Russell, Chris; Wachter, S., «Explaining Explanations...», *Op. cit.*, p. 280.
48. Mittelstadt, B., Russell, C., Wachter, S. (2019). «Explaining Explanations in AI». *FAT*19: Proceedings of the Conference on Fairness, Accountability, and Transparency*, p. 280; DEEKS (2019). «The judicial demand...» pp. 1832; Barredo *et al.*, «Explainable Artificial Intelligence (XAI)...», *Op. cit.*, p. 83.
49. Barredo, A.; Díaz-Rodríguez, N., *et al.* «Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI». *Information Fusion*, vol. 58 (2020), p. 100; Marcinkevics, R., Vogt, J. E., «Interpretability and Explainability: A Machine Learning Zoo Mini-tour», *ArXiv abs/2012.01805* (2020), p. 2; Cátedra iDANAE. *Interpretabilidad de los Modelos de Inteligencia Artificial*, Universidad Politécnica de Madrid, Management Solutions, Newsletter Trimestral, 2019, p. 4; Kroll, J. A., Huey, J., Barocas, S. *et al.*, «Accountable Algorithms», *University of Pennsylvania Law Review*, vol. 165, n.º 3 (2017), pp. 658-660; Edwards, L. y Veale, M., «Slave to the algorithm? Why a «right to an explanation» is probably not the remedy you are looking for», *Duke Law & Technology Review*, vol. 16, n.º 18 (2017), p. 8; Ananny, M. y Crawford, K., «Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability», *New Media & Society* (2016), p. 983.

interpretabilidad/rendimiento predictivo de los modelos de IA y, en particular, de los «modelos de caja negra», la XAI viene desarrollando un conjunto de técnicas de cuya finalidad es generar modelos más explicables manteniendo niveles altos de rendimiento.⁵⁰

No obstante, se trata de un planteamiento criticado por algunos autores, en la medida en que este enfoque no hace sino fomentar el desarrollo e implementación de modelos de caja negra propietarios en lugar de «modelos interpretables por defecto» en sectores altamente críticos como la justicia criminal o la asistencia sanitaria.⁵¹ De hecho, el Information Commissioner's Office y el Instituto Alan Turing recomiendan a las organizaciones que den prioridad al uso de sistemas basados en modelos interpretables por defecto, en la medida de lo posible, especialmente cuando los sistemas de IA tengan un impacto adverso potencialmente alto en las personas o sean críticos para la seguridad.⁵²

Como ya se señaló *supra*, mientras que en la definición de la «transparencia» incluida en el Considerando (27) la interpretabilidad se encuentra ausente en favor de la explicabilidad; en el artículo 13, el enfoque es precisamente el inverso. Asimismo, se ha criticado el enfoque que el Reglamento hace de la interpretabilidad en el artículo 13, en la medida en que dicho precepto no aclara si resulta suficiente que el proveedor garantice técnicamente la interpretabilidad del sistema de alto riesgo para cumplir con el requisito de transparencia.⁵³

50. Guning, D., *Explainable Artificial Intelligence (XAI)*, [7] D. Technical Report, Defense Advanced Research Projects Agency (DARPA), 2017.

51. Rudin, C., «Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead». *Nature Machine Intelligence*, vol. 1 (2019), pp. 206-215. Con relación a las estrategias de aprendizaje automático, la autora cuestiona la creencia generalizada de que deba existir necesariamente una compensación entre el rendimiento predictivo y la interpretabilidad y, por tanto, la necesidad de implementar modelos complejos de caja negra para obtener el máximo rendimiento predictivo. Según Rudin esto no sería así cuando se utilizan datos estructurados con atributos representativos y relevantes, pues en tales casos, no suele haber, por ejemplo, diferencias significativas en el rendimiento entre clasificadores más complejos (redes neuronales profundas, árboles de decisión potenciados, bosques aleatorios) y clasificadores mucho más sencillos (regresión logística, listas de decisión). La autora considera que, en lugar de implementar modelos intrínsecamente interpretables, existe una creciente tendencia a fomentar enfoques de «ML explicable», generando un segundo modelo (*post-hoc*) para explicar el primer modelo de caja negra. Para la autora este planteamiento resulta problemático por dos motivos. En primer lugar, porque las explicaciones resultantes en los modelos *post-hoc* pueden resultar engañosas por varios motivos: (i) no siempre son fieles a lo que el modelo original ha computado; (ii) no proporcionan suficiente detalle de lo que el modelo de caja negra está haciendo; (iii) pueden no ser compatibles con situaciones en las que información externa al sistema requiera combinarse con una evaluación de riesgos; (iv) pueden conducir a una decisiones excesivamente complejas y propicias para el error humano. En segundo lugar, porque incentivan la proliferación de modelos propietarios sujetos a derechos de propiedad intelectual e industrial oponible frente a las necesidades de «apertura» del modelo para el ejercicio de derechos por parte de los sujetos afectados por decisiones adversas en contextos críticos.

52. Cfr. Information Commissioner's Office, Alan Turing Institute, «*Explaining decisions...*», *Op. cit.*, pp. 68-69.

53. Vid. Kiseleva, A., «Making AI's transparency transparent: notes on the EU Proposal for the AI Act», *European Law Blog*, 20 julio 2021. <https://europeanlawblog>.

Pero aun siendo cierto lo anterior, el artículo 13 tampoco aclararía el nivel de interpretabilidad exigible. Esto último es esencial, pues incluso los modelos más interpretables (e.g. árboles de decisión, regresión lineal/logística, bayesianos), tienen distinto nivel de transparencia en función del grado de simulabilidad, descomponibilidad o transparencia algorítmica que permite cada modelo.⁵⁴ En la tabla siguiente se incluye una relación de modelos interpretables y los niveles de transparencia con relación a los elementos integrantes de la interpretabilidad.

Tabla 4
Relación entre transparencia y modelos interpretables

Modelo	Simulabilidad	Descomponibilidad	Transparencia algorítmica
Regresión lineal/logística	Los predictores son legibles y las interacciones entre ellas se reducen al mínimo.	Las variables siguen siendo legibles, pero el número de interacciones y predictores implicados en ellas han crecido para forzar la descomposición.	Si las variables e interacciones son demasiado complejas, son necesarias herramientas matemáticas para analizarlas.
Árboles de decisión	Un humano puede simular y obtener la predicción del árbol de decisión por su cuenta, sin necesidad de base matemática.	El modelo comprende normas que no alteran los datos, y conserva su legibilidad.	Normas legibles por el ser humano que explicar los conocimientos aprendidos de los datos y permite para una comprensión directa del proceso de predicción.
Vecinos más próximos K	La complejidad del modelo (número de variables, su comprensibilidad y la medida de similitud) coincide con las capacidades humanas para la simulación.	Si la cantidad de variables es demasiado alta y/o la similitud medida es demasiado compleja para ser capaz de simular el modelo completamente, la similitud y el conjunto de las variables pueden descomponerse y analizarlos por separado.	Si la medida de similitud no puede descomponerse y/o el número de variables es elevado, el usuario debe acudir a las herramientas matemáticas y estadísticas para analizar el modelo.

Fuente: Barredo *et al.* (2020)

[eu/2021/07/29/making-ais-transparency-transparent-notes-on-the-eu-proposal-for-the-ai-act/#:~:text=Transparency](https://eur-lex.europa.eu/2021/07/29/making-ais-transparency-transparent-notes-on-the-eu-proposal-for-the-ai-act/#:~:text=Transparency)

54. Barredo, et al., «Explainable Artificial Intelligence (XAI)....». *Op. cit.*, p. 90.

La interpretabilidad aparece mencionada a lo largo del artículo 13: en el apartado (1), en el apartado (3)(b)(vii) y en el apartado (3)(d). Mientras que el artículo 13(1) incluiría un mandato a los proveedores para que diseñasen sistemas de IA de alto riesgo interpretables a fin de que los responsables del despliegue «*interpreten* y usen correctamente su información de salida»; los otros dos apartados (3)(b)(vii) y (3)(d) se referirían al contenido que debería incluirse en las instrucciones de uso para posibilitar al responsable del despliegue la correcta interpretación de los resultados de salida y el uso correcto del sistema.

Una posible interpretación del enfoque adoptado por el artículo 13 es que el precepto trata de fomentar los sistemas interpretables por defecto, en lugar de los sistemas de caja negra necesitados de técnicas y herramientas complementarias de la explicabilidad. Es más, si nos atenemos a los antecedentes del Reglamento, en las Directrices Éticas, como ya se explicó *supra*, la interpretabilidad está ausente a la hora de definir el requisito de la transparencia (trazabilidad, explicabilidad y comunicación). Sin embargo, en el check-list de verificación para una IA confiable buena parte de las preguntas de verificación incluidas en el apartado de explicabilidad van encaminadas a comprobar el grado de interpretabilidad que, por defecto, tendría el sistema, y en menor medida, la capacidad del sistema de generar explicaciones que posibiliten la comprensión o interpretación de los resultados del sistema.⁵⁵

Tabla 5

Lista de verificación para una IA confiable: explicabilidad e interpretabilidad

Explicabilidad	
Pregunta de verificación	Finalidad
¿Ha evaluado en qué medida son <i>comprensibles</i> las decisiones y, por tanto, el resultado producido por el sistema de IA?	Interpretabilidad por defecto
¿Se ha asegurado de que se pueda elaborar una <i>explicación</i> comprensible para todos los usuarios que puedan desearla sobre las razones por las que un sistema adoptó una decisión determinada que diera lugar a un resultado específico?	Explicabilidad
¿Diseñó el sistema de IA teniendo en mente la interpretabilidad desde el principio?	Interpretabilidad desde el diseño
¿Investigó y trató de utilizar el modelo más sencillo e interpretable posible para la aplicación en cuestión?	
¿Ha evaluado si puede analizar los datos de entrenamiento y de prueba? ¿Puede cambiar y actualizarlos con el tiempo?	
¿Ha evaluado si, tras el entrenamiento y desarrollo del modelo, o si tiene acceso al flujo de trabajo interno del modelo?	

Fuente: Elaboración propia a partir de Directrices Éticas del HLEG (2018)

55. Véanse Directrices Éticas, Op. cit., pp. 37-38.

A la vista entonces de los antecedentes y de que el propio artículo 13 parece exigir al proveedor el diseño de sistemas de IA que posibiliten a los responsables del despliegue la interpretación de la información de salida y la adecuada utilización del sistema, sería razonable plantearse si la voluntad del legislador europeo no habría sido la de fomentar los *sistemas interpretables por defecto* frente a modelos de caja negra y necesitados de técnicas complementarias de explicabilidad.

Algunos autores rechazan tal interpretación, puesto que la restricción de los modelos complejos de caja negra en aras de favorecer modelos interpretables podría limitar la innovación.⁵⁶ Razón no falta en este argumento, pues las referencias a la innovación son constantes a lo largo de todo el Reglamento.⁵⁷

3. LA EXPLICABILIDAD EN EL ARTÍCULO 13: UN ENFOQUE AMBIGUO Y LIMITADO

Desde el ámbito de la normalización, la «explicabilidad» es la propiedad técnica de un sistema de IA que se refiere a los factores relevantes que influyen en una decisión y que pueden ser expresados de una forma que los humanos puedan comprenderlo. Ahora bien, aunque la explicabilidad busca responder a la pregunta de «¿Por qué?», en realidad no proporciona una argumentación que justifique si el curso de la acción que se tomó fue necesariamente el más óptimo.⁵⁸ Por su parte, las Directrices Éticas del HLEG definen la explicabilidad como «la capacidad de explicar tanto los procesos técnicos de un sistema de IA como las decisiones humanas relacionadas (por ejemplo, las áreas de aplicación de un sistema)».⁵⁹

En su caso, las explicaciones serían el medio a través del cual pueden explicarse las decisiones de un sistema de IA de una forma clara, comprensible, transparente e interpretable para el destinatario de la explicación. Por tanto, si la interpretabilidad es el objetivo final a conseguir, las explicaciones son herramientas para conseguir la

56. Kiseleva, A., «Making AI's transparency transparent...», *Op. cit.*

57. Cfr. Considerandos (1), (2), (3), (8), (25), (68), (102), (105), (119), (138), (139), (143), (146), Artículos 1, 40.3, y Capítulo VI del Reglamento de Inteligencia Artificial.

58. UNE-EN ISO/IEC 22989: 2023, 3.5.7. En sentido similar, la norma IEEE Std 7001-2021 define la «explicabilidad» como «el grado en que la información puesta a disposición de una parte interesada de forma transparente puede ser fácilmente interpretada y comprendida por una parte interesada».

59. HLEG, *Op. cit.*, apartado 77. Las Directrices establecen una diferencia entre la explicabilidad *ad-intra* («explicabilidad técnica» frente a los usuarios o responsables del despliegue de los sistemas de IA), y la explicabilidad *ad-extra* (colectiva o de los individuos afectados). «La explicabilidad técnica –explica el HLEG– requiere que las decisiones tomadas por un sistema de IA puedan ser comprendidas y rastreadas por los seres humanos. Además, es posible que haya que elegir entre aumentar la explicabilidad de un sistema (lo que puede reducir su precisión) o aumentar su precisión (a costa de la explicabilidad). Siempre que un sistema de IA tenga un impacto significativo en la vida de las personas, debería ser posible exigir una explicación adecuada del proceso de toma de decisiones del sistema de IA. Dicha explicación debe ser oportuna y adaptarse a los conocimientos de la parte interesada (por ejemplo, legos, reguladores o investigadores). Además, debe ser posible obtener explicaciones sobre el grado en que un sistema de IA influye y configura el proceso de toma de decisiones de la organización, las opciones de diseño del sistema y la justificación de su despliegue».

interpretabilidad del modelo.⁶⁰ Por tanto, la «explicabilidad» es una característica *activa* de los modelos de IA que se refiere a su capacidad *técnica* para generar una explicación sobre su comportamiento a partir de los datos utilizados, de los resultados obtenidos y del proceso completo de la toma de decisión⁶¹, en función de la audiencia o perfil de los destinatarios a los que se dirige la explicación⁶².

En concreto, dicha explicación habrá ser oportuna y adaptarse al nivel de especialización de la parte interesada (e.g. regulador, autoridad de control, experto, investigador, afectado por la decisión o público en general) a efectos de que el sistema sea realmente explicable.⁶³ A su vez, los sistemas de IA explicables deben cumplir con una serie de principios básicos: (i) que el sistema produzca una explicación (por ser intrínsecamente interpretable, técnicamente por sí mismo, o a partir de metodologías y métricas complementarias); (ii) que la explicación sea significativa y adecuada para las partes interesadas a las que va dirigida; (iii) que la explicación refleje los procesos del sistema con precisión (distinta de la precisión de la decisión o rendimiento predictivo); (iv) y que el sistema exprese los límites de su diseño y dominio.⁶⁴ Así pues, mientras la transparencia respondería a la pregunta de *cómo funciona el modelo*, la explicabilidad respondería a *qué información adicional puede extraerse del modelo* (explicaciones) cuando resulta imposible o complejo ver y entender (interpretabilidad) cómo trabaja éste internamente (caja negra).⁶⁵ A su vez, mientras que la interpretabilidad sería el fin último, las explicaciones serían las herramientas para alcanzar la interpretabilidad cuando el modelo no es interpretable por sí mismo.⁶⁶

Tanto para el Grupo de Expertos de la Comisión Europea, como para la Autoridad Británica de Protección de Datos y el Instituto Alan Turing, los métodos que incluyen técnicas XAI de tipo *post-hoc*, locales o globales (e.g. como los modelos subrogados o proxy, Gráficos de Dependencias Parciales, LIME, Explicaciones Aditivas Shapley, Contrafácticos, entre otras) resultan esenciales, no solo para explicar a los usuarios el comportamiento de los sistemas de IA no interpretables intrínsecamente, sino también para desplegar una tecnología fiable.⁶⁷ Asimismo, las explicaciones pueden

60. Carvalho *et al.* (2019). «Machine Learning Interpretability...» *Op. cit.*, p. 15.

61. Cátedra iDANAE (2019). *Interpretabilidad...*, *Op. cit.*, 3.

62. Barredo *et al.* (2020), «Explainable Artificial Intelligence (XAI)...» *Op. cit.*, p. 84.

63. HLEG, *Op. cit.*, apartado 77.

64. Phillips, P. Jonathon, Hahn, Carina A., *et al.*, *Cuatro principios de inteligencia artificial explicable*, NISTIR 8312, 2021, <https://doi.org/10.6028/NIST.IR.8312> Los autores señalan que, mientras existen métricas establecidas para evaluar el rendimiento predictivo, las métricas específicas para medir la precisión de las explicaciones estarían aún en proceso de desarrollarse.

65. Lipton, Z. C., «The Mythos of Model Interpretability». *ACM Queue*, vol. 16, n.º 3 (2018), p. 12; Lepri, *et al.*, «Fair, transparent...», *Op. cit.*, p. 622.

66. Kiseleva *et al.* «Transparency of AI in healthcare...» *Op. cit.*, p. 6.

67. HLEG, *Directrices Éticas...*, *Op. cit.*, apartado 99; ICO y Alan Turing Institute, *Op. cit.*, pp. 120–128, <https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence/> A la fecha de elaboración de este Capítulo, la OECD tiene publicado un listado de 16 métricas específicas de explicabilidad. Vid. OCDE.AI Policy Observatory, *Catalogue of Tools & Metrics for Trustworthy AI*, 2024. <https://oecd.ai/en/catalogue/metrics?objectives=11&page=1>

incluir tanto medidas técnicas (bien de tipo *post-hoc*, bien explicaciones generadas automáticamente por el propio sistema mediante tecnologías de IA), como medidas no técnicas (por ejemplo, explicaciones escritas u orales en lenguaje natural sobre el funcionamiento del sistema de IA).⁶⁸

Al examinar con detenimiento la génesis, evolución y texto final del Reglamento en su conjunto, y del artículo 13, en particular, puede afirmarse que la explicabilidad no tiene un tratamiento adecuado en el texto. Es más, a pesar de la importancia que en las Directrices Éticas tiene este requisito, la explicabilidad sólo aparece mencionada en tres ocasiones: en la definición de «transparencia» que incorpora el Considerando (48) en los términos explicados anteriormente, en el artículo 13 y en el Artículo 86 del Reglamento.

Si nos retrotraemos a los antecedentes legislativos, la pura realidad es que, en la propuesta de la Comisión, no había ni rastro de la explicabilidad a lo largo de todo el texto del Reglamento, con la sola excepción del Considerando (48).⁶⁹ En cambio, la interpretabilidad sí que aparecía en los apartados (1) y (3.d) del artículo 13 en los mismos términos que el texto definitivo aprobado. Ahora, en su versión final, el artículo 13(3)(b)(iv) RIA prevé que las instrucciones de uso de los sistemas de alto riesgo incorporen:

«en su caso, las capacidades y características técnicas del sistema de IA de alto riesgo para proporcionar información pertinente para *explicar* sus resultados de salida [cursiva nuestra]».

La redacción actual del apartado (3.b.iv) del artículo 13, que incluye una mención muy genérica a la «explicabilidad», tiene su origen en la enmienda núm. 38 introducida por el Parlamento.⁷⁰ Además del artículo 13(3)(b)(iv) del Reglamento, la otra referencia a la explicabilidad es la contenida en el Artículo 86, donde se reconoce el derecho a una *explicación* frente al responsable del despliegue de las personas individuales afectadas por las decisiones adoptadas por un sistema de alto riesgo del Anexo III –pero no así de los sistemas de alto riesgo del Anexo I sujetos a legislación armonizada! El precepto no delimita cuál habría de ser, en todo caso, la información básica a facilitar a los afectados para garantizar su derecho a una explicación, «clara

68. Kiseleva et al. «Transparency of AI in healthcare...» *Op. cit.* pp. 6-7.

69. Cfr. Considerando (48) de la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la unión (COM/2021/206 final). El Considerando incluía una referencia mínima al impacto que, en el derecho a la tutela judicial efectiva, podían tener los sistemas de alto riesgo utilizados por las autoridades públicas con fines de aplicación de la ley (investigación, detección, prevención y sanción de delitos por autoridades competentes) que «no sean lo suficientemente transparentes y explicables ni estén bien documentados».

70. La Enmienda 308 del Parlamento Europeo introdujo un inciso iii.bis (nuevo) con el siguiente texto entre la información a incorporar en las instrucciones de uso: «el grado en que el sistema de IA pueda ofrecer una *explicación* de las decisiones que adopte [cursiva nuestra].» Vid. Parlamento Europeo, P9_TA(2023)0236, *Op. cit.* Ya, durante el proceso de trilogos, en el Acuerdo Provisional del Consejo, el Parlamento y la Comisión de 02/02/2024 (PE758.862v01-00) se incluyó una versión prácticamente idéntica a la del texto actual del Artículo 13(3)(b)(iv).

y significativa y [que pueda] servir de base para que las personas afectadas puedan ejercer sus derechos».⁷¹

Así las cosas, el enfoque que hace el RIA respecto de la explicabilidad ha sido objeto de interpretaciones distintas por la doctrina, lo cual denota la ambigüedad y la falta de claridad del legislador. Unos consideran que, frente a la interpretabilidad, la explicabilidad es efectivamente la gran ausente del Reglamento.⁷² En este sentido, se sostiene que el artículo 13 no establecería obligación general de explicabilidad para los sistemas de IA de alto riesgo, sino la transparencia del funcionamiento del sistema y de la generación de resultados. En todo caso, esta transparencia, al menos, debería garantizar que esos elementos sean interpretables, lo que no equivale necesariamente a la exigencia de una explicación, al menos en los términos explicados antes.⁷³

En cambio, la lectura que hacen otros autores del artículo 13.1 en conexión con el Artículo 14(4)(c) es que, a través de estas disposiciones específicas, el Reglamento impone al proveedor una «obligación de explicabilidad» que es, a la vez, capacitadora del responsable del despliegue y está orientada al cumplimiento normativo. Por un lado, porque dicha obligación serviría para que los responsables del despliegue del sistema de IA puedan interpretarlo y utilizarlo correctamente; por otro, porque ayudaría a verificar la adecuación del sistema a las obligaciones establecidas por el Reglamento, contribuyendo en última instancia a lograr el cumplimiento normativo.⁷⁴ Sin embargo, consideramos que este último planteamiento es erróneo pues confunde la interpretabilidad con la explicabilidad.

También se ha entendido que, al exigir el artículo 13.1 «un nivel de transparencia suficiente para que los responsables del despliegue interpreten y usen correctamente sus resultados de salida», el precepto daría cobertura tanto a distintos tipos de explicaciones (locales, globales, contrafácticas) como a información más o menos granular sobre la importancia de las variables. En todo caso, las explicaciones deberían ser fieles al modelo en el sentido de que habrán de ser, al menos de manera aproximada, una correcta reconstrucción de los parámetros de decisión interna.⁷⁵ Tampoco compartimos esta lectura del artículo 13, puesto que, vuelve a confundir la interpretabilidad y la explicabilidad y, además, como se argumentará más adelante,

71. Cfr. Considerando (171).

72. Kiseleva, A., «Making AI's transparency transparent...»; Schneeberger, *et al.*, «The Tower of Babel...», p. 70.

73. Bordt, S., Finck, M., Raidl, E., von Luxburg, U., «Post-Hoc Explanations Fail to Achieve their Purpose in Adversarial Contexts». *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT 22)*, Association for Computing Machinery, New York, p. 894. <https://doi.org/10.1145/3531146.3533153>

74. Sovrano *et al.*, «Metrics, Explainability...», *Op. cit.*, p. 132.

75. Hacker, P «Varieties of AI Explanations Under the Law...», p. 359. En sentido similar, véase Onitiu, D., «The limits of explainability & human oversight in the EU Commission's proposal for the Regulation on AI— a critical approach focusing on medical diagnostic systems», *Information & Communications Technology Law*, vol. 32, n.º 2(2022), p. 175. <https://doi.org/10.1080/13600834.2022.2116354> Sin embargo, Schneeberger, *et al.*, «The Tower of Babel...», *Op. cit.*, p. 70, son de la opinión de que el Artículo 13 deja abierta la interpretación a si el Artículo 13 RIA exige o no la aplicación de técnicas pos-hoc y, en su caso, al enfoque que debe elegirse (local, global).

la propia dicción del precepto parece inclinarse por las explicaciones locales (la «información de salida del sistema»).

A la vista de lo anterior, la opinión que aquí se sostiene es que la aproximación a los requisitos de «interpretabilidad» y de «explicabilidad» que hace el RIA, en general, y el artículo 13, en particular, es ambigua por los motivos que se indican a continuación. En primer lugar, no está clara la relación/distinción entre la interpretabilidad y la explicabilidad en el artículo 13. Para una mejor argumentación e intelección del precepto, en la tabla siguiente se incorporan las referencias a la interpretabilidad y la explicabilidad del artículo 13.

Tabla 6
Interpretabilidad y explicabilidad en el artículo 13 RIA

Apartado	Contenido	Finalidad
13(1.)	Los sistemas de IA de alto riesgo se diseñarán y desarrollarán de un modo que garantice que funcionan con un nivel de transparencia suficiente para que los responsables del despliegue <i>interpreten</i> y usen correctamente su información de salida.	Interpretabilidad por defecto [1]
13(3.b.iv)	Las instrucciones de uso incluirán, en su caso, las <i>capacidades y características técnicas</i> del sistema de IA de alto riesgo para proporcionar información pertinente <i>para explicar</i> su información de salida.	Técnicas de explicabilidad [2]
13(3.b.vii)	Las instrucciones de uso incluirán, en su caso, información que permita a los responsables del despliegue <i>interpretar la información de salida</i> del sistema de IA de alto riesgo y utilizarla adecuadamente.	Información capacitadora de la interpretabilidad [3]
13(3.d)	Las instrucciones de uso incluirán las medidas de vigilancia humana a que se hace referencia en el artículo 14, incluidas las <i>medidas técnicas establecidas para facilitar la interpretación de la información de salida</i> de los sistemas de IA de alto riesgo por parte de los responsables del despliegue.	Técnicas de interpretabilidad [4]

Fuente: Elaboración propia

Al analizar la dicción empleada por el artículo 13 en cada uno de los apartados indicados en la Tabla anterior, una de las posibles interpretaciones de cómo el precepto integra la interpretabilidad y la explicabilidad podría ser la siguiente:

— Con carácter general, el proveedor debe garantizar que el diseño y configuración de los sistemas de alto riesgo sean *interpretables por defecto* para el responsable del despliegue [1].

— En caso de que el modelo no sea intrínsecamente interpretable (e.g. modelos de caja negra), las instrucciones de uso deberán incorporar información sobre las técnicas implementadas efectivamente por el proveedor para *generar explicaciones* que posibiliten una adecuada interpretación de la información de salida (técnicas y herramientas de explicabilidad) por el responsable del despliegue [2].

— Aunque el modelo sea interpretable por defecto, las instrucciones de uso deben contener la información suficiente con vistas a *garantizar la capacitación del responsable del despliegue para realizar una adecuada interpretación* de la información de salida y un uso correcto del sistema [3]. Dicha información puede resultar útil cuando los usuarios que utilicen el sistema implementado por el responsable del despliegue no sean expertos en IA (e.g. un médico, un funcionario, o un trabajador dentro de una organización privada), e incluso teniendo cierto *expertise* el modelo seleccionado requiera de herramientas matemáticas o estadísticas para analizar la descomponibilidad y la transparencia algorítmica (cfr. Tabla 4 *supra*).

— Las instrucciones de uso deberán incorporar información sobre las medidas de vigilancia humana implementadas⁷⁶, incluyendo las técnicas establecidas para *facilitar la interpretación* de la información de salida de los sistemas de IA [4]. Esto podría incluir información sobre las técnicas concretas que se han utilizado para generar un sistema interpretable por defecto, por ejemplo, el tipo de modelo interpretable implementado por el proveedor (e.g. regresión, árboles de decisión, listas de reglas de decisión, vecinos próximos K)⁷⁷. Pero también

76. Según el Artículo 14(4)(c) RIA, las medidas de vigilancia humana que se implementen en el sistema de alto riesgo capacitarán a las personas físicas a quienes el responsable del despliegue encomiende la supervisión humana puedan para «interpretar correctamente los resultados de salida del sistema de IA de alto riesgo, teniendo en cuenta, por ejemplo, los métodos y herramientas de interpretación disponibles». No está tampoco clara la relación de la interpretabilidad/explicabilidad con las medidas de supervisión humana, sean incorporadas por el proveedor desde el diseño (e.g. mediante incorporación de herramientas de interfaz humano-máquina adecuadas), bien porque las implemente el responsable del despliegue según las recomendaciones del proveedor. Dada la redacción del Artículo 14(4)(c) RIA parece que los «métodos y herramientas de interpretación» de los resultados del sistema formarían parte de las medidas de supervisión humana. Y por la dicción empleada por el precepto, «métodos y herramientas de interpretación», el Reglamento parece referirse a técnicas de explicabilidad.

77. Cfr. Information Commissioner's Office y Alan Turing Instituto, «Explaining decisions...», *Op. cit.*, pp. 73-74. Se señala en la Guía que, «para los modelos de IA que son básicamente interpretables (como los sistemas basados en la regresión, listas de reglas/decisiones, árboles de decisión, Naïve Bayes o K vecino más cercano), el aspecto técnico de extraer una explicación significativa es relativamente sencillo. Normalmente, se recurre a la lógica intrínseca de la función de mapeo del modelo mediante la observación humana directa [...] Por ejemplo, en los árboles de decisión o en las listas de decisiones/reglas, la lógica que subyace a un resultado dependerá de las relaciones interpretables de las afirmaciones condicionales ponderadas (si-entonces). En otras palabras, cada nodo o componente de este tipo de modelos funciona, de hecho, como una razón [...] En general, resulta útil conocer la *gama de técnicas disponibles* para construir modelos de IA interpretables [...]. Estas técnicas no sólo hacen que los fundamentos de los modelos de IA sean fácilmente comprensibles, sino que también constituyen la base de muchas de las herramientas de explicación complementarias que se utilizan ampliamente para hacer que los modelos de “caja

podría hacer referencia a las técnicas complementarias para generar explicaciones que posibiliten la interpretación de los resultados del sistema (e.g. LIME, modelos interpretables subrogados, gráficos de dependencias parciales, etc).⁷⁸

Además de lo anterior, de la dicción empleada por el artículo 13, parece sugerirse que sólo sería exigible un nivel de *interpretabilidad local*, pues en sus distintos apartados la exigencia de interpretabilidad se limita exclusivamente a la «información de salida del sistema», excluyéndose así otros elementos del sistema, como el modelo y sus componentes (variables, parámetros, interacciones, algoritmo de procesamiento).⁷⁹ Esta exigencia de la interpretabilidad local podría interpretarse en el sentido de que el Reglamento estaría priorizando las técnicas de explicabilidad locales, en lugar de las técnicas explicabilidad globales.

Ahora bien, aunque las explicaciones locales son decisivas en los casos en que las decisiones del sistema tienen impacto en las personas individuales, las explicaciones globales posibilitan comprender la relación entre los componentes del sistema y su comportamiento en conjunto. En este sentido, las explicaciones globales serán a menudo fundamentales no sólo para establecer una explicación local precisa, sino para garantizar la equidad, la seguridad y el rendimiento óptimo de su sistema de IA. Asimismo, la comprensión global del sistema también puede proporcionar información esencial sobre los impactos potenciales más generales del sistema en colectivos específicos y en la sociedad en general.⁸⁰ En este sentido, las explicaciones globales podrían ser relevantes para explicar el funcionamiento del sistema de alto riesgo «con respecto a personas o grupos de personas específicos en relación con los que esté previsto utilizar el sistema» (cfr. artículo 13.3.b.v), e incluso identificar circunstancias conocidas o previsibles,

negra” sean más interpretables.» Es decir, que los modelos interpretables («medidas técnicas establecidas para facilitar la interpretación de la información de salida de los sistemas de IA de alto riesgo») se pueden utilizar, por ejemplo, como modelos subrogados para explicar los resultados obtenidos por los modelos no interpretables.

78. Cfr. *Ibidem*. La Guía del Information Commissioner’s Office y del Instituto Alan Turing señala que, si tras considerar el dominio, el impacto y los factores técnicos, se hubiera optado por utilizar un sistema de IA de «caja negra», deberían incorporarse herramientas de explicación complementarias adecuadas a la construcción de su modelo. Las estrategias de explicación complementarias disponibles según el estado de la técnica para apoyar la interpretabilidad «pueden arrojar luz sobre aspectos significativos de los procesos globales de un modelo y componentes de sus resultados locales».

79. Cfr. Con relación a los componentes de un modelo de IA, véase ISO/IEC DIS 12792: 2024 (en), 9.

80. Las interpretaciones locales tienen como objetivo interpretar predicciones o clasificaciones individuales correspondientes a instancias concretas con el fin de identificar las variables de entrada específicas que han podido ser determinantes o han tenido más peso en la generación de un predicción o clasificación particular. Por su parte, las explicaciones globales buscan ofrecer una visión amplia que abarque la importancia general de las variables y de sus interacciones en los resultados generados por el modelo, el funcionamiento interno y la lógica del comportamiento de ese modelo en su conjunto. Las interpretaciones globales se centran en explicar el conjunto del modelo, y no tanto comportamiento para un caso particular, y pueden contribuir a que el proceso de toma de decisiones sea coherente desde el punto de vista procedimental. Vid. Information Commissioner Office’s, Alan Turing, *Explaining decisions... Op. cit.*, p. 74.

asociadas «a la utilización del sistema de IA de alto riesgo conforme a su finalidad prevista o a un uso indebido razonablemente previsible, que [pudieran] dar lugar a riesgos para la salud y la seguridad o los derechos fundamentales» de las personas y para la sociedad en su conjunto (cfr. artículo 13.3.b.iii). Sin embargo, dada la indefinición del Reglamento, los proveedores podrían acogerse a una interpretación restrictiva basada en la interpretabilidad local como criterio de *minimis* de acogerse a una interpretación literal del artículo 13.

IV. ÁMBITOS SUBJETIVO Y FORMAL DEL ARTÍCULO 13

En los apartados (1) y (2) del artículo 13 se delimita el requisito de transparencia a partir de dos ámbitos. En primer lugar, se identifican los sujetos obligados y destinatarios del requisito de transparencia de los sistemas de alto riesgo. Y, en segundo lugar, se establecen los aspectos formales básicos que modulan el cumplimiento de la obligación, a saber, el nivel de transparencia exigible, así como la manera en que la información pertinente debe ser comunicada al responsable del despliegue.

1. SUJETOS Y LOS FINES DE LA TRANSPARENCIA EN EL ARTÍCULO 13: LOS GRANDES AUSENTES EN EL REGLAMENTO

Existe un común acuerdo en la doctrina en que el requisito de transparencia establecido en el artículo 13 se circunscribe a dos sujetos concretos: el proveedor, el obligado por el requisito de transparencia, y el responsable del despliegue, destinatario de la información prevista en el artículo 13(3).

Como ya se anticipó más arriba, el artículo 13(1) configura un tipo de transparencia interna. El proveedor, según el Artículo 3.3 RIA, puede ser cualquier «persona física o jurídica, autoridad pública, órgano u organismo que *desarrolle* un sistema de IA o un modelo de IA de uso general o *para el que se desarrolle* un sistema de IA o un modelo de IA de uso general y lo introduzca en el mercado o ponga en servicio el sistema de IA con su propio nombre o marca, previo pago o gratuitamente [cursiva nuestra]».

Este concepto comprendería así aquellas circunstancias en que el proveedor es el responsable directo del diseño y desarrollo del sistema, como aquellas otras, en que un tercero ha diseñado y desarrollado el sistema para el proveedor, siendo este último el responsable de su introducción en el mercado o puesta en servicio con su propio nombre comercial o marca. A su vez, el destinatario de la información contenida en el artículo 13.3 RIA es el responsable del despliegue, que según el Artículo 3.4 se identificaría con cualquier «persona física o jurídica, o autoridad pública, órgano u organismo que utilice un sistema de IA bajo su propia autoridad, salvo cuando su uso se enmarque en una actividad personal de carácter no profesional».

En el caso de los responsables del despliegue, la obligación de transparencia prevista en el artículo 13 tendría una doble finalidad (de cumplimiento normativo y capacitadora). Por un lado, la transparencia posibilita el cumplimiento normativo por parte de los responsables del despliegue de las obligaciones previstas en la Sección 3ª del Capítulo III, y en particular, de las obligaciones del

Artículo 26 (entre otras, la supervisión humana, la vigilancia del funcionamiento del sistema de IA de alto riesgo, la conservación de los archivos de registro que los sistemas de IA de alto riesgo generen automáticamente, o el cumplimiento de la evaluación de impacto). Por otro, la transparencia también capacitaría a los responsables del despliegue para que puedan interpretar la información de salida y utilizar correctamente el sistema de conformidad con las instrucciones de uso facilitadas por el proveedor (artículo 13(1)).

En el caso de los proveedores, la finalidad de la transparencia es el cumplimiento normativo (artículo 13(1)) y, en su caso, la acreditación de dicho cumplimiento ante las autoridades de control. En el primer caso, el cumplimiento normativo se refiere a las obligaciones previstas en el Artículo 18, entre otras, el cumplimiento de los requisitos definidos en la sección 2ª, la implementación del sistema de gestión de la calidad, la conservación de la documentación, la adopción de medidas correctoras y su comunicación, la conservación de los archivos de registro generados automáticamente por sus sistemas, la sujeción al procedimiento de evaluación de la conformidad, la elaboración de la declaración de conformidad, el registro del sistema de alto riesgo en la base de datos de la UE.

Junto al cumplimiento normativo, la transparencia (mediante la comunicación y la trazabilidad) también posibilita al proveedor la demostración del cumplimiento normativo ante la autoridad nacional competente de la conformidad del sistema de IA de alto riesgo con los requisitos establecidos en la Sección 2ª del Capítulo III (Artículo 18(2)(K)).

Frente a esta transparencia interna, una de las críticas más unánimes que se han hecho al artículo 13 y, en general, a todo el Reglamento, es que el enfoque de la transparencia es exclusivamente técnico y fuertemente limitado desde el punto de vista subjetivo. Y, por lo tanto, el RIA en ningún caso sería verdaderamente habilitante para el ejercicio de derechos por los afectados por los sistemas de alto riesgo, al no establecerse un marco claro que ofrezca a los particulares vías claras para impugnar las decisiones adoptadas por los sistemas de IA que les afecten.⁸¹ En su caso, el artículo 13 habría establecido una suerte de «transparencia de expertos para expertos», lo cual quedaría ejemplificado en el listado de información descrito por el artículo 13.(3) y a incluir por el proveedor en las instrucciones de uso orientadas y dirigidas exclusivamente al responsable del despliegue. En este sentido, el Reglamento habría establecido un objetivo particular y restringido de la transparencia, que se limitaría exclusivamente facilitar el cumplimiento por parte de proveedores y responsables del despliegue de las obligaciones establecidas en la Sección 3ª del Capítulo 3º (Artículo 16-27), en detrimento de una transparencia orientada al ejercicio de derechos por las personas afectadas por las decisiones de los sistemas de alto riesgo. En particular, el artículo 13 habría configurado «un novedoso tipo de transparencia instrumental, autorreferencial y orientada al cumplimiento, centrada en la aplicación eficaz y conforme de los sistemas de IA en entornos concretos».⁸²

81. Onitiu, D., «The limits of explainability...», *Op. cit.*, p. 171; Smuha, et al., «How the EU Can Achieve Legally Trustworthy AI», *Op. cit.*, p. 52.

82. Hacker, P., Passoth, JH., «Varieties of AI Explanations Under the Law...», *Op. cit.*, p. 361.

Este enfoque restrictivo de la transparencia entraría en clara contradicción con el planteamiento de la parte expositiva del RIA, donde se insiste continuamente, de una forma u otra, en que el objetivo esencial del Reglamento es «promover la adopción de una inteligencia artificial (IA) centrada en el ser humano y fiable, garantizando al mismo tiempo un elevado nivel de protección de la salud, la seguridad y los *derechos fundamentales* consagrados en la Carta de los Derechos Fundamentales de la Unión Europea (en lo sucesivo, “Carta”), incluidos la democracia, el Estado de Derecho».⁸³

2. *El ámbito formal de la transparencia: el indefinido «tipo y nivel de transparencia adecuados»*

Según el artículo 13(1) RIA, «[l]os sistemas de IA de alto riesgo se diseñarán y desarrollarán de un modo que se garantice que funcionan con un nivel de transparencia suficiente para que los responsables del despliegue interpreten y usen correctamente sus resultados de salida. Se garantizará un *tipo y un nivel de transparencia adecuados* para que el proveedor y el responsable del despliegue cumplan las obligaciones pertinentes previstas en la sección 3 [cursiva nuestra]».

El enfoque adoptado por la AIA resulta ambiguo: ni se concreta el tipo ni el nivel de transparencia que se considera adecuado.⁸⁴ La forma y el nivel adecuados de transparencia parecen ser relativos y meramente instrumentales con vistas a lograr el cumplimiento de otros requisitos del RIA.⁸⁵

Es posible que las normas técnicas que desarrollen el CEN y el CENELEC a en respuesta a la petición de normalización realizada por la Comisión Europea desarrollen esta cuestión⁸⁶, o en su defecto, las especificaciones comunes que pudiera aprobar la Comisión mediante actos de ejecución.⁸⁷

Mientras tanto, a la hora de determinar el tipo y nivel de transparencia exigible, debe tenerse en cuenta que las diferentes normas técnicas consultadas (la ISO/IEC DIS 12792:2024(en) o IEEE 7001-2021) establecen la necesidad de adecuar el tipo, nivel e incluso formato de la información relevante a los diferentes perfiles de las partes interesadas.

A los efectos de este Capítulo y de determinar el significado y alcance de la expresión «tipo y niveles de transparencia», la norma IEEE 7001-2021 para sistemas autónomos⁸⁸ resulta interesante porque precisamente gradúa los niveles

83. Vid. Considerando (1) del Reglamento. Y, en sentido similar, véanse Considerandos (8)-(10), (20), (28), (32), (43), (46), (48) entre otros muchos más.

84. Cfr. Onitiu *et al.*, «The limits of explainability...», *Op. cit.*, p. 174; Kiseleva, «Making AI's Transparency transparent...», *Op. cit.*; Boch, A., Hohma, E., Trauth, R., *Towards an Accountability Framework for AI: Ethical and Legal Considerations*, Technical University of Munich, Munich Center for Technology in Society, Institute for Ethics in Artificial Intelligence, febrero 2022, p. 5.

85. Véanse Schneeberger, *et al.*, «The Tower of Babel...», *Op. cit.*, pp. 65, 70; Hacker, P., Passoth, J.H., «Varieties of AI Explanations...», *Op. cit.*, p. 359.

86. Vid. Artículo 40 RIA y Commission Implementing Decision of 22.5.2023 on a standardisation request (C(2023) 3215 final), *Op. cit.*

87. Vid. Considerando (121) y Artículo 41 RIA.

88. El ámbito de aplicación de la norma comprende a todos los sistemas autónomos, tanto físicos como no físicos. Entre los primeros, cabe incluir los vehículos con sistemas de conducción automatizada o los robots asistenciales; y, entre los segundos,

de transparencia exigibles del 0 (más bajo) al cinco (más alto)⁸⁹ en función del rol de las partes interesadas involucradas (usuarios del sistema, público en general, organismos de certificación o regulación, investigadores de incidentes/accidentes y asesores expertos en procedimientos administrativos o judiciales).

A su vez, la norma distingue, entre distintas sub-categorías de usuarios del sistema que requerirían diferentes niveles de transparencia:

— Los usuarios no expertos, que incluyen tanto a las personas que sólo tienen una breve interacción con el sistema como a las personas que interactúan manera frecuente con el sistema.

— Los usuarios expertos en el dominio, incluye a usuarios con conocimientos y experiencia en el dominio el que se aplica el sistema (por ejemplo, un médico). Estos usuarios tienen cierta responsabilidad en el uso del sistema de IA.

— Los super-usuarios son expertos no sólo en sistemas de IA, sino también en los sistemas concretos de los que son responsables. Entre estos super-usuarios se incluyen las personas responsables del desarrollo, el diagnóstico de fallos, la reparación, el mantenimiento y la actualización, además del funcionamiento y la supervisión de sistemas autónomos concretos.

En la tabla siguiente se incorporan los distintos niveles de transparencia según el perfil de los usuarios. Nótese que, a partir del nivel 3 de transparencia, la norma establece diferentes requisitos de explicabilidad adecuados al perfil del usuario.

los sistemas de diagnóstico médico (recomendadores) o los chatbots. Los sistemas autónomos inteligentes que utilizan el aprendizaje automático también entran en el ámbito de aplicación de la norma. Asimismo, los conjuntos de datos utilizados para entrenar dichos sistemas también están dentro del ámbito de aplicación de la norma cuando se considera la transparencia del sistema en su conjunto. Vid. IEEE 7001-2021, 1.1.

89. Dentro de cada categoría de partes interesadas se establecen requisitos de niveles de transparencia medibles y comprobables. Los niveles de transparencia se definen de 0 (ninguna transparencia) a 5 (el máximo nivel de transparencia alcanzable). Cada definición es un requisito expresado como una propiedad cualitativa del sistema que debe cumplirse. Los niveles 1 a 5 se han definido para describir niveles sucesivamente mayores de transparencia. Todos los niveles se consideran técnicamente viables, mientras que cada nivel sucesivo suele ser más exigente. Cada nivel es acumulativo y se basa en los anteriores, por lo que se espera que cuando un sistema cumpla el nivel n de una categoría concreta, también cumpla los niveles $n - 1$. En cada caso, la verificación del nivel consiste simplemente en determinar si el requisito se cumple o no, es decir, si la propiedad de transparencia exigida por un nivel determinado para un grupo de interesados dado está presente de forma demostrable o no. *Idem*, 5.

Tabla 7

Niveles de transparencia y explicabilidad para los usuarios en IEEE 7001-2021

Nivel	Definición
0	No hay transparencia.
1	<p>Se facilitará al usuario información accesible en que incluya como mínimo: a) escenarios de ejemplo con el comportamiento esperado y previsto del sistema, incluidos los modos de funcionamiento degradados, y b) principios generales de su funcionamiento, es decir, si existe un componente de aprendizaje y qué datos utiliza.</p> <p>La documentación deberá explicar los principios generales de funcionamiento del sistema. En el caso de un sistema que utilice el aprendizaje automático, la documentación deberá ofrecer una explicación sencilla de qué fuentes examina/utiliza el sistema como parte del proceso de aprendizaje, incluidas las posibles fuentes de sesgo. Esta documentación consistirá, por ejemplo, en un manual escrito, una guía pictórica o una audioguía, según las necesidades del usuario, que le explique cómo se comporta el sistema en las distintas circunstancias y situaciones que sus diseñadores esperan que encuentre.</p> <p>Los usuarios y super-usuarios expertos recibirán la documentación de usuario especificada anteriormente y elaborada de conformidad con la norma IEC/IEEE 82079-1. Esta documentación detallará el funcionamiento seguro y la supervisión del sistema.</p> <p>Para los super-usuarios, la documentación detallará los procedimientos de diagnóstico de fallos del sistema, reparación, mantenimiento, actualización y desmantelamiento al final de su vida útil.</p>
2	<p>Se proporcionará al usuario material de formación interactivo que le permita ensayar sus interacciones con el sistema en situaciones virtuales específicas y pertinentes.</p> <p>Además, los usuarios y super-usuarios expertos en la materia recibirán material formativo interactivo sobre el funcionamiento seguro y la supervisión del sistema. Los super-usuarios recibirán además material de formación interactivo sobre diagnóstico de averías, reparación, mantenimiento, actualización y desmantelamiento al final de la vida útil.</p>

Nivel	Definición
3	<p>Se proporcionará al usuario no experto una funcionalidad iniciada por el usuario que produzca una <i>explicación</i> breve e inmediata de la actividad más reciente del sistema. Estas <i>explicaciones</i> se expresarán a través de medios comúnmente comprensibles, como el lenguaje natural u otro medio apropiado (por ejemplo, una imagen). Ni la realización de solicitudes ni la comprensión de las respuestas del sistema a dichas solicitudes requerirán que el usuario no experto reciba formación alguna. No obstante, se aceptarán los avisos por motivos de seguridad o legales que resulten necesarios.</p> <p>En el caso de los sistemas diseñados para ser utilizados por usuarios expertos, se proporcionará la misma funcionalidad especificada anteriormente, con la salvedad de que a) el sistema permitirá solicitar <i>explicaciones</i> de cualquiera de sus decisiones recientes y b) las <i>explicaciones</i> podrán expresarse utilizando un lenguaje apropiado para la materia. Además, se proporcionará a los expertos documentación en la que se detalle cómo deben solicitarse e interpretarse estas explicaciones. Dicha documentación también deberá incluir los subsistemas de procesamiento del lenguaje natural (NLP), si existen.</p>
4	<p>Se proporcionará al usuario no experto una funcionalidad iniciada por el usuario que produzca una explicación breve e inmediata de lo que hace el sistema en una situación dada. La conformidad con este nivel de transparencia permite al usuario explorar escenarios hipotéticos de «qué pasaría si» en una situación dada, si es aplicable al ámbito de trabajo del sistema.</p> <p>Ni la realización de solicitudes ni la comprensión de las respuestas del sistema a dichas solicitudes requerirán que el usuario no experto reciba formación alguna, aunque sí es necesario que se familiarice con la documentación de usuario del sistema.</p> <p>En el caso de los sistemas diseñados para ser utilizados por usuarios expertos, se proporcionará la misma funcionalidad aquí especificada, con la salvedad de que las <i>explicaciones</i> podrán expresarse utilizando un lenguaje apropiado para la materia. Además, se proporcionará a los usuarios expertos documentación en la que se detalle cómo deben solicitarse e interpretarse estas <i>explicaciones</i>. Dicha documentación también deberá incluir los subsistemas de NLP, en caso de que existan.</p> <p>Es importante destacar que este nivel de transparencia permite al usuario extraer <i>explicaciones contrafácticas</i>.</p>

Nivel	Definición
5	<p>Se proporcionará al usuario una explicación continua del comportamiento que adapte el contenido y la presentación de la <i>explicación</i> en función de las necesidades de información del usuario y del contexto. Esto incluirá el acceso a archivos de registro y datos de entrenamiento siempre que no contengan información sensible como datos personales. La <i>explicación</i> del funcionamiento se logrará mediante alguna presentación visual sencilla y visible, después de que el sistema realice una acción, o mediante la vocalización de frases explicativas mientras el sistema realiza una acción.</p> <p>No se exigirá a los usuarios no expertos un esfuerzo adicional para acceder a las explicaciones pertinentes. Esta interacción deberá adaptarse al historial de interacciones del usuario, ya que la confianza se pierde fácilmente si, por ejemplo, el sistema se comporta de forma inesperada.</p> <p>Se dispondrá de detalles explicativos adicionales, a petición, según lo requieran los usuarios expertos o los super-usuarios, lo que les permitirá explorar interactivamente el sistema y su funcionamiento.</p>

Fuente: IEEE 7001-2021, 5.1.1

Además de exigir tipos y niveles de transparencia adecuados al cumplimiento normativo por parte del proveedor y del responsable del despliegue, el artículo 13(2) del Reglamento prescribe ciertos requisitos formales para la información relevante a incorporar en las instrucciones de uso que acompañarán al sistema: presentación en un formato digital o de otro tipo adecuado, concisión, completitud, corrección, claridad, accesibilidad y comprensibilidad adecuadas al responsable del despliegue. En este sentido, la transparencia debería tener en cuenta la posible percepción y comprensión de las partes interesadas y, en su caso, evitar revelar información de una manera que, aunque sea técnicamente cierta, esté enmarcada de una manera que conduzca a una interpretación errónea.⁹⁰

V. EL CONTENIDO MATERIAL DE LA OBLIGACIÓN DE TRANSPARENCIA

Desde el punto de vista material, el proveedor del sistema de alto riesgo deberá incluir en las instrucciones de uso que facilite al responsable del despliegue la información referida en el artículo 13(3) RIA. Primeramente, es importante aclarar que las instrucciones de uso no pueden confundirse con la documentación técnica (Artículo 11) cuya conservación se exige al proveedor durante un período de diez años a disposición de las autoridades nacionales competentes (Artículo 18(1)(a)).⁹¹ No

90. IEEE 7001-2021, 3.1.

91. Vid. Artículo 11(1) RIA: La documentación técnica de un sistema de IA de alto riesgo se elaborará antes de su introducción en el mercado o puesta en servicio, y se mantendrá actualizada. La documentación técnica se redactará de modo que demuestre que el sistema de IA de alto riesgo cumple los requisitos establecidos en la presente sección y que proporcione de manera clara y completa a las autoridades nacionales competentes y a los organismos notificados la información necesaria para evaluar la

por obvio, está de menos hacer la aclaración. En segundo lugar, el artículo 13(3) no contiene un *numerus clausus* de categorías de información, sino que debe interpretarse como un criterio *de minimis*, ya que el Reglamento es claro al señalar que «[l]as instrucciones de uso contendrán *al menos* la siguiente información [...] [cursiva nuestra]».

Las instrucciones de uso facilitadas al responsable del despliegue deberán incluir como mínimo información relativa a:

- La identidad y los datos de contacto del proveedor y, en su caso, de su representante autorizado (artículo 13(3)(a)).
- La idoneidad funcional del sistema, incluyendo sus características, capacidades y limitaciones (artículo 13(3)(b)).
- Los cambios en el sistema e idoneidad funcional predeterminados por el proveedor (artículo 13(3)(c)).
- Las medidas de vigilancia humana previstas, incluyendo técnicas que faciliten la interpretación de la información de salida del sistema (artículo 13(3)(d)).
- Los recursos computacionales, de hardware y vida útil prevista del sistema (artículo 13(3)(e)).
- Los controles de registros log implementados (artículo 13(3)(d)).

Al margen de que nada en el artículo 13 o en otra disposición del Reglamento hace pensar que el proveedor no pueda ampliar esta información si con ello considera que contribuye a aumentar la transparencia del sistema con vistas a mejorar la capacitación del responsable del despliegue, habrá que esperar a las normas técnicas que elaboren el CEN y el CENELEC para saber si esta información *de minimis* se puede ver ampliada con otra información relevante. A este respecto, debe tenerse en cuenta que la norma ISO/IEC DIS 12792:2024(en) incluye dentro de cada nivel de taxonomías de transparencia categorías de información que no están contempladas en el artículo 13, y que podrían adaptarse al perfil de la parte interesada destinataria de la información.⁹²

conformidad del sistema de IA con dichos requisitos. Contendrá, como mínimo, los elementos contemplados en el anexo IV.

92. Así, por ejemplo, en el nivel-contexto podría incluirse información pertinente relativa al impacto ambiental del sistema, por ejemplo, las evaluaciones de impacto ambiental realizadas, el consumo energético del sistema, su huella de carbón y agua, retirada del sistema y gestión de residuos. En el nivel-sistema se incluye el acceso a los elementos internos del sistema de IA, entre los que se encuentran determinados componentes individuales; partes del código fuente; elementos específicos de los modelos (listado de reglas o elementos de conocimiento embebidos en grafos, los parámetros en los modelos de machine learning), valores internos e intermedios resultantes del procesamiento de un input concreto). En el nivel-modelo el acceso podría incluir, la dependencia del modelo de otros modelos, el algoritmo concreto de procesamiento, el procedimiento para construir el modelo, los hiperparámetros, o los formatos de los datos de entrada y de salida. Por último, en el nivel-datos, la información relevante podría abarcar la identificación del origen de los datos, sus propiedades estadísticas, sus sesgos y limitaciones, el pre-procesamiento y preparación de los datos en bruto, o la forma de etiquetamiento, identificación de desequilibrios, o medidas de anonimización y pseudoanonimización implementadas.

En tercer lugar, el artículo 13(3) no identifica el nivel detalle con el que proveedor deberá especificar las distintas categorías de información identificadas en los sub-apartados (a)-(f). De hecho, esas categorías de información incluidas en el apartado (3) del artículo 13 coinciden, a su vez, con buena parte de las categorías de información previstas en el Anexo IV que forman parte del contenido de la documentación técnica.⁹³ No está claro, sin embargo, si el nivel de detalle de la información contenida en las instrucciones de uso deberá ser menor, cualitativamente y cuantitativamente, que el de la documentación técnica, teniendo en cuenta, además, los distintos destinatarios y finalidades de comunicación de una u otra información.⁹⁴

Además de lo anterior, el apartado (3) del artículo 13 emplea locuciones modales acotadoras o limitativas («en su caso», «cuando proceda») que, en una lectura literal podrían llevar a interpretaciones restrictivas de algunos preceptos. Sería el caso de los apartados (3)(b)(vi) (con relación a los datos utilizados por el sistema) (3)(b)(vii) (interpretación adecuada de la información de salida y uso correcto del sistema), o (3)(f) (con relación a los archivos de registro).⁹⁵

Por último, si tenemos en cuenta las taxonomías de niveles de transparencia identificadas en la ISO/IEC 25059:2023(en) (nivel contexto, nivel sistema, nivel modelo, nivel datos), se aprecia que la mayoría de las categorías de información previstas en el artículo 13(3) se refieren al nivel-sistema (finalidades previstas, capacidades, limitaciones funcionales, usos recomendados y prohibidos, conservación y archivo de registros log, factores humanos), y sólo algunas de ellas al nivel-modelo

93. Cfr. Tabla 2 de correspondencia entre Artículo 13 RIA, Articulado RIA y Normalización, donde se incluyen las equivalencias entre los sub-apartados del Artículo 13(3) y las categorías de información del Anexo IV.

94. En el caso de las instrucciones de uso, los destinatarios son los responsables del despliegue y su finalidad es el cumplimiento normativo por estos últimos de las obligaciones previstas en la Sección 3ª del Capítulo III, así como su capacitación para que puedan interpretar la información de salida y utilizar correctamente el sistema (Artículo 13(1)). En el caso de la documentación técnica, los destinatarios últimos son las autoridades nacionales competentes y los organismos notificados y su finalidad es que autoridades y organismos puedan evaluar la conformidad del sistema de IA con los requisitos establecidos en la Sección 2ª del Capítulo III del Reglamento (Artículo 11(1)), incluyendo, por tanto, el requisito de transparencia.

95. La restricción gramatical introducida en los apartados indicados significa que, «en su caso» o «cuando proceda», el proveedor podría no facilitar la información concernida, lo que chocaría con las propias finalidades del Artículo 13 respecto de los responsables del despliegue. En unos casos, unos casos porque se omitiría información que podría ser relevante para la adecuada interpretación de los resultados de salida y el uso correcto del sistema. En otros, porque la información en cuestión posibilita el cumplimiento normativo del responsable con algunas de las obligaciones impuestas por el Artículo 26 RIA, entre otras, la supervisión humana (apartado 2); la pertinencia y representatividad de los datos utilizados por el sistema, especialmente si el control de los datos los tiene el proveedor (apartado 3); la vigilancia del funcionamiento del sistema de alto riesgo (apartado 5); la conservación de los archivos de registro (apartado 6), e incluso la adecuada implementación de la evaluación de impacto en materia de protección de datos cuando ésta sea preceptiva (apartado 9).

(rendimiento predictivo, recursos computacionales y de hardware), y al nivel-datos (información básica).⁹⁶

1. INFORMACIÓN SOBRE LA IDONEIDAD FUNCIONAL Y OTRAS PROPIEDADES

El artículo 13(3)(b) del Reglamento dispone que las instrucciones de uso que acompañen a los sistemas de alto riesgo deberán incluir información sobre «las características, capacidades y limitaciones del funcionamiento del sistema de IA de alto riesgo [cursiva nuestra]». La expresión «funcionamiento del sistema de IA» debe entenderse como en el sentido indicado por el propio Reglamento «la capacidad de un sistema de IA para alcanzar su finalidad prevista» (Artículo 3.18). Este concepto coincide con la noción técnica de «idoneidad funcional», es decir, «la capacidad del sistema para proporcionar funciones que faciliten la realización de tareas y objetivos especificados».⁹⁷ A su vez, la «finalidad prevista» del sistema de alto riesgo (cfr. artículo 13.3(b)(i)) sería «el uso para el que un proveedor concibe un sistema de IA, incluidos el contexto y las condiciones de uso concretas, según la información facilitada por el proveedor en las instrucciones de uso, los materiales y las declaraciones de promoción y venta, y la documentación técnica» (Artículo 3.12). Aquí, el concepto de «funcionamiento» se identifica con la finalidad del sistema (su capacidad para producir una predicción, una recomendación, una decisión), prevista, declarada, documentada y testada por el proveedor para un contexto o dominio concretos y unas condiciones de implementación específicas.

Al identificar la finalidad prevista del sistema de alto riesgo (artículo 13(3)(b)(i)), las instrucciones de uso deberían describir cuáles son los objetivos del sistema de acuerdo con las necesidades del usuario que pueden abordar y cómo la IA puede contribuir a conseguir esos objetivos.⁹⁸ Dado que la finalidad depende, entre otros factores, del contexto concreto del sistema de IA de alto riesgo, parece pertinente que las instrucciones de uso incorporen información básica sobre el contexto social (ubicación geográfica prevista para su implementación, contexto socio-laboral u organizacional del despliegue, limitaciones lingüísticas o culturales del sistema).⁹⁹

La información sobre el contexto social puede ser relevante, a su vez, para que el responsable del despliegue pueda interpretar correctamente los eventuales «riesgos para la salud y la seguridad o los derechos fundamentales» previstos o a consecuencia de un uso indebido (artículo 13(3)(b)(iii)), o el funcionamiento del sistema «respecto a determinadas personas o determinados colectivos de personas en relación con los que esté previsto utilizar el sistema» (artículo 13(3)(b)(v)).

96. Sobre los cuatro niveles de taxonomías de transparencia en la ISO/IEC 25059:2023(en), véase la Tabla 2 *supra*.

97. Con relación a la idoneidad funcional, véanse ISO/IEC 25010:2023 (en), 3.1; ISO/IEC 25059:2023 (en), 5.1, Figura 1.

98. Cfr. ISO/IEC DIS 12792:2024 (en), 8.4.2.

99. Cfr. ISO/IEC DIS 12792:2024 (en), 7.2.1. La ISO/IEC 22989:2022 (en), no sólo contempla el impacto social (5.18), sino también aspectos relacionados con la jurisdiccionalidad (5.17), pues en el país donde se ha diseñado o producido el sistema éste podría estar sujeto a distintos requerimientos legales que en la Unión Europea.

En cualquier caso, debe tenerse en cuenta que, durante el trámite legislativo del Reglamento, el texto presentado por el Consejo frente a la propuesta de la Comisión contemplaba la incorporación de un inciso al apartado (3)(b)(i), que incluyera no sólo la finalidad prevista, sino también información sobre «el entorno geográfico, funcional o de comportamiento específico en el que se pretende utilizar el sistema de IA de alto riesgo». Sin embargo, este planteamiento no fue incorporado en el texto final del Reglamento.

Además de la idoneidad funcional, el artículo 13(3) contempla diversas categorías de información que se refieren a otras propiedades técnicas de los sistemas de IA, como la solidez y la seguridad, o la eficiencia del rendimiento (o desempeño). Respecto de la solidez y la seguridad, se trata de dos requisitos que, junto con la precisión, deben cumplir los sistemas de alto riesgo en los términos previstos en el Artículo 15. Por su parte, con relación a la solidez y la seguridad, el artículo 13(3)(ii) se refiere al «nivel probado y validado del sistema de IA de alto riesgo y que puede esperarse, así como cualquier circunstancia conocida y previsible que pueda afectar en su nivel esperado». Sin embargo, no se define en el precepto cuál debería ser la información específica, cuantitativa o cualitativa, que debería facilitarse al responsable del despliegue para cumplir con la exigencia de transparencia (interpretación de los resultados de salida, uso correcto del sistema y cumplimiento de las obligaciones del Artículo 26). Por ejemplo, el artículo 13 incluye de forma explícita la información sobre las métricas de rendimiento, como se comentará más abajo, pero no así las métricas de solidez y que, por otra parte, el Artículo 15 sí menciona expresamente. Igualmente, tampoco queda clara la información básica a incluir en las instrucciones de uso sobre las medidas técnicas y organizativas de seguridad para cumplir con las finalidades de la transparencia del artículo 13.

Algunos apartados del artículo 13 también incluyen información relativa a la «eficiencia del rendimiento». Esta propiedad representa la capacidad del sistema para realizar sus funciones dentro de unos parámetros de tiempo y rendimiento especificados y ser eficiente en el uso de los recursos en unas condiciones determinadas. Los recursos pueden ser la CPU, la memoria, el almacenamiento, los dispositivos de red, otros productos de software con los que interactúa el sistema, o la energía utilizada.¹⁰⁰ En este sentido, el artículo 13(3)(e) RIA también prevé que se facilite al responsable del despliegue la información relativa a los «recursos computacionales¹⁰¹ y de hardware necesarios, la vida útil prevista del sistema de IA de alto riesgo, así como las medidas de mantenimiento y cuidado necesarias, incluida su frecuencia, para garantizar el correcto funcionamiento de dicho sistema, también en lo que respecta a las actualizaciones del software».¹⁰²

100. ISO/IEC 25010:2023(en), 3.2. Véase también, Janapa Reddi, Vijai (ed.) *Machine Learning Systems with TinyML*, Harvard University, última actualización 21 de marzo de 2024, p. 392. https://harvard-edge.github.io/cs249r_book/contents/benchmarking/benchmarking.html

101. En la más que mejorable versión en español del Reglamento se utiliza la expresión «recursos informáticos», en lugar de «recursos computacionales» («*computational and hardware resources*», en la versión en inglés).

102. Nótese que la ISO/IEC DIS 12792:2024 (en), 9.4.7 y 9.4.8 incluye en la taxonomía de transparencia correspondiente al nivel del modelo el tipo de hardware informático y

2. INFORMACIÓN SOBRE LA CORRECCIÓN FUNCIONAL O RENDIMIENTO PREDICTIVO: LA «PRECISIÓN» Y SUS «MÉTRICAS»

En los sistemas de IA, la llamada «corrección funcional» o «rendimiento predictivo»¹⁰³ es una de las propiedades técnicas que mayor repercusión puede tener en la fiabilidad de los sistemas de IA¹⁰⁴ y, por tanto, en la interpretación de los resultados de salida y en su uso correcto por parte del responsable del despliegue.

La «corrección funcional» define la capacidad del sistema de proporcionar resultados correctos con el grado de precisión necesario. Los sistemas de IA, y en particular, los que utilizan modelos de aprendizaje automático, no suelen proporcionar, sin embargo, una corrección funcional en todas las circunstancias observadas porque se espera que haya una cierta tasa de error en sus resultados. Por ello, existen numerosas métricas que evalúan la precisión funcional. Además, dependiendo del contexto de uso y la finalidad del sistema, podría ser necesario establecer compensaciones entre la corrección funcional, y otras propiedades del sistema, como la eficiencia del rendimiento o la solidez.¹⁰⁵ Es más, la corrección funcional también podría verse afectada por la ciberseguridad. Así, por ejemplo, la Agencia Europea de Ciberseguridad identifica entre las amenazas a los modelos de aprendizaje, el compromiso de la corrección de la inferencia; la reducción del nivel de precisión de los datos, modificándolos o mezclándolos con otros dataset de diferentes calidades; o la manipulación de los datos etiquetados en los modelos supervisados¹⁰⁶. A su vez, el despliegue de controles de seguridad a menudo conduce a un delicado equilibrio entre la seguridad del sistema y su rendimiento.¹⁰⁷

los costes computacionales (e.g. total de tiempo CPU y GPU por muestra de datos de entrada o por tamaño de datos de entrada).

103. ISO/IEC 25059:2023(E), Anexo C. Si bien resulta común en el ámbito de la IA el uso de la expresión de «rendimiento predictivo» para significar cuán bien desempeña sus tareas previstas un sistema de IA concreto, la norma aclara que resulta preferible referirse a esta propiedad como «corrección funcional» para diferenciarla claramente de la eficiencia del rendimiento o desempeño («performance efficiency»). El «rendimiento predictivo» se refiere a la capacidad de generalización del modelo, esto es, a su capacidad para obtener resultados precisos con nuevas y desconocidas entradas de datos, más allá de los ejemplos específicos con los que el modelo se entrenó. Vid. Martínez-Heras, Jose Antonio, IArtificial.net. Technical Report, 2023. DOI: 10.13140/RG.2.2.16587.77609M; Ministerio para la Transformación Digital y Red.es, How do I know if my prediction model is really good?, 26 de enero de 2020. <https://datos.gob.es/en/blog/how-do-i-know-if-my-prediction-model-really-good> Por su parte, con relación a los modelos de aprendizaje automatizado, la norma ISO/IEC 42001:2023 define la voz «generalisation» («generalización») como «la capacidad de un modelo entrenado para realizar predicciones correctas a partir de datos de entrada no vistos previamente» por dicho modelo.
104. Cfr. ISO/IEC 22989:2022 (en), 5.15.3. Precisamente, la norma define la fiabilidad en los sistemas de IA como la capacidad del sistema que «le permite proporcionar la predicción requerida [...], la recomendación y la decisión de *forma coherente y correcta* durante su fase de funcionamiento [cursiva nuestra].»
105. ISO/IEC 25059:2023(E), 3.2.3, 5.4.
106. ENISA, *AI Cybersecurity Challenges. Threat Landscape for Artificial Intelligence*, diciembre 2020, DOI 10.2824/238222, p. 44-47.
107. Vid. ENISA, *Securing machine learning algorithms*, diciembre 2021, pp. 3, 27. DOI: 10.2824/874249.

A la corrección funcional –además de otras propiedades, como la solidez y la ciberseguridad– parece referirse el artículo 13(3)(b)(ii), al exigir que las instrucciones de uso incorporen información específica sobre el «*nivel de precisión (incluidos los parámetros para evaluarla)*», solidez y ciberseguridad mencionado en el artículo 15 con respecto al cual se haya probado y validado el sistema de IA de alto riesgo y que puede esperarse, así como las circunstancias conocidas y previsibles que podrían afectar al nivel de precisión, solidez y ciberseguridad esperado [cursiva nuestra].» A su vez, el Artículo 15(3) del Reglamento reitera esta previsión, al establecer que «[e]n las instrucciones de uso que acompañen a los sistemas de IA de alto riesgo se indicarán los niveles de precisión de dichos sistemas, así como los *parámetros pertinentes para medirla*».¹⁰⁸

Respecto al «nivel de precisión (incluidos los parámetros para evaluarla)» tanto la expresión utilizada en la versión en castellano como en la versión en inglés («level of accuracy, including its metrics») deben hacerse dos observaciones preliminares. En primer lugar, la traducción española de «metrics» como «parámetros para evaluar [la precisión]» resulta inapropiada.¹⁰⁹

Primero porque el término «parámetro» tiene un significado técnico específico. Y es, en tal sentido técnico como es utilizado este término en los Considerandos (98, 102, 104), referidos a los modelos de uso general y, en particular, a los «pesos», o en el Artículo 3, apartados (29) y (30), con relación a los parámetros entrenables y no entrenables (o hiperparámetros). En segundo lugar, porque el término «precisión» (o «accuracy» en la versión en inglés) en sentido estricto se refiere a una métrica de rendimiento específica de modelos de clasificación basados en ML.¹¹⁰

108. Una vez más, la versión en español del Reglamento traduce el término «métricas» como «parámetros pertinentes para medir[...] [la precisión]».

109. La corrección funcional o rendimiento predictivo es una propiedad medible y evaluable, cuantitativa y cualitativamente (ISO/IEC 42001:2023, 3.11; ISO/IEC 25059:2023(en), Anexo C) mediante las llamadas «métricas de rendimiento» («performance metrics»), también denominadas «métricas de error» («error metrics») o «métricas de evaluación» («evaluation metrics»). Dichas métricas incluyen construcciones lógico-matemáticas diseñadas para medir la proximidad o cercanía entre el resultado previsto (predicción) y el resultado real. Es decir, las métricas de rendimiento o error permiten una evaluación de la calidad del modelo en términos de su capacidad predictiva. En este sentido, cuanto mayor sea la diferencia entre el resultado real «r» y el resultado previsto «p», más «alejado» estará el modelo de ser una representación precisa de la realidad. Por el contrario, cuanto más se acerquen los valores estimados «p» a la realidad «r», mejor será el rendimiento del modelo en términos predictivos. Vid. Plevris, Vagelis; Solorzano, G.; Bakas, N. P.; et al, «Investigation of performance metrics in regression analysis and machine learning-based prediction models», *ECCOMAS Congress 2022 — 8th European Congress on Computational Methods in Applied Sciences and Engineering*, 2022, https://www.scipedia.com/public/Plevris_et_al_2022a

110. La métrica de exactitud («accuracy») mide el porcentaje de casos (verdaderos positivos y verdaderos positivos) que el modelo ha acertado. La ISO/IEC 23053:2022(E), 6.5.5.4, identifica las métricas más habituales para los modelos de clasificación (*accuracy*, precision, confusion matrix, recall, F1 score) y regresión (mean absolute error, root mean squared error, relative absolute error, relative squared error, mean zero one error, coefficient of determination). Sobre la aplicación de métricas de error a soluciones de IA adquiridas por el ámbito del Sistema de Salud Nacional, véase Gutiérrez

A partir de una interpretación sistemática del artículo 13.3(b)(ii) RIA en conexión con la parte expositiva del RIA, debería concluirse que la expresión «nivel de precisión (incluidos los parámetros para evaluarla)» necesariamente se refiere a la identificación y descripción en las instrucciones de uso de la corrección funcional o rendimiento predictivo del modelo, así como de las métricas de rendimiento o error implementadas en el sistema de IA por el proveedor.

Por parte del proveedor, la medición del rendimiento predictivo de un modelo de IA puede tener distintas finalidades¹¹¹:

— La evaluación del modelo, para conocer cuán fiables son sus predicciones¹¹² o la frecuencia y el tamaño esperado de sus errores [1].

— La comparación de distintos modelos, para elegir entre aquellos que presenten mejores compensaciones entre rendimiento y eficiencia¹¹³ [2].

— Las comparaciones fuera de muestra y a lo largo del tiempo, para comprobar que el rendimiento del modelo no se ha degradado con nuevos datos de producción¹¹⁴ [3].

— La determinación, en función del caso de uso y contexto de aplicación, de la manera más óptima de compensar la relación (normalmente inversa) entre el rendimiento del modelo y su nivel de interpretabilidad [4].

— El diseño de modelos más interpretables, y en su caso, explicables, manteniendo niveles altos de rendimiento [5].

Facilitar al responsable del despliegue información relevante sobre el rendimiento predictivo del sistema contribuye a garantizar la finalidad capacitadora de la

David, M.E. y Quintana Cortés, J.L., «Public Procurement of AI for the EU Healthcare Systems. First Insights from the Spanish Experience», *European Review of Digital Administration & Law — Erdal*, Volume 4, Issue 1 (2023), pp. 131-132.

111. Véase, ISO/IEC TS 4213:2022, 6.

112. Liu, Zhenyu y Chen, Huanhua, «A predictive performance comparison of machine learning models for judicial cases», en *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, Honolulu, 2017, pp. 1-6, DOI: 10.1109/SSCI.2017.8285436. Así, por ejemplo, los autores han evaluado y comparado el rendimiento de distintos algoritmos de aprendizaje automatizado (k-NN, regresión logística, bagging, bosques aleatorios y máquinas de soporte vectorial) en la predicción de decisiones judiciales a partir de una selección de variables que representan el contexto semántico de casos procedentes de la base de datos HUDOC del Tribunal Europeo de Derechos Humanos, en los que se alega la vulneración de los Artículos 3 (Prohibición de la tortura y de los tratos inhumanos o degradantes), 6 (derecho a un juicio justo) y 8 (derecho al respeto a la vida privada y familiar, del domicilio y de la correspondencia).

113. Por ejemplo, en Deng, Fei, Huang, Jibing, Yuan, *et al.* «Performance and efficiency of machine learning algorithms for analyzing rectangular biomedical data», en *Laboratory Investigation*, vol. 101, 2021, pp. 430-441, <https://doi.org/10.1038/s41374-020-00525-x>, se analiza comparativamente el rendimiento de distintos modelos de aprendizaje automático (árboles de decisión, bosques aleatorios, máquinas de soporte vectorial y redes neuronales artificiales) para la clasificación multicategoría de causas de muerte (supervivencia, cáncer de mama, otros tipos de cáncer, enfermedades cardiovasculares, otras causas) a partir de grandes conjuntos de datos biomédicos.

114. Cfr. Janapa Reddi, Vijai (ed.) *Machine Learning Systems with TinyML*, Harvard University, última actualización 21 de marzo de 2024, p. 607. https://harvard-edge.github.io/cs249r_book/contents/benchmarking/benchmarking.html

transparencia en el artículo 13 (correcta interpretación de los resultados de salida y uso adecuado del sistema) y de cumplimiento normativo en los siguientes términos.

En primer lugar, facilitar información relevante sobre la corrección funcional del sistema, la frecuencia o el tamaño de sus errores, lo que contribuye a mejorar la fiabilidad de las predicciones. La fiabilidad de las predicciones no sólo depende de la implementación de adecuadas métricas de error según el contexto y finalidad prevista del sistema, sino también de la calidad de los datos utilizados para entrenar, validar y testar los modelos. Por eso, resulta pertinente que el artículo 13(3)(vi) en las instrucciones de uso se incorpore «especificaciones relativas a los datos de entrada, o cualquier otra información pertinente en relación con los conjuntos de datos de entrenamiento, validación y prueba usados, teniendo en cuenta la finalidad prevista del sistema de IA.» Una información adecuada sobre los datos y su preprocesamiento puede ayudar a detectar sesgos inherentes en los mismos y fuentes de posible discriminación.

En segundo lugar, dada la naturaleza evolutiva de algunos sistemas de IA, resulta imprescindible comprobar el comportamiento del sistema a lo largo del tiempo con nuevos datos de producción. Por ello resulta oportuno que el artículo 13(3)(c) incluya en las instrucciones de uso de «los cambios en el sistema de IA de alto riesgo y su funcionamiento, predeterminados por el proveedor en el momento de efectuar la evaluación de la conformidad inicial». Aquí podría considerarse la descripción de los mecanismos aplicados para que el comportamiento del modelo evolucione de la manera prevista dentro de la versión predeterminada por el proveedor¹¹⁵.

Asimismo, a lo largo del ciclo de vida del sistema de IA pueden producirse nuevos riesgos para las personas afectadas por el sistema no previstos inicialmente por el proveedor debido a determinados usos del responsable del despliegue, ya sean conformes a su finalidad o indebidos. Tales riesgos podrían deberse a la entrada de nuevos datos con una distribución y representatividad distintas a los que se utilizaron para entrenar, validar y testar el modelo. De ahí que el artículo 13(3)(b)(iii) prevea la incorporación en las instrucciones de uso de la información sobre «cualquier circunstancia conocida o previsible, asociada a la utilización del sistema de IA de alto riesgo conforme a su finalidad prevista o a un uso indebido razonablemente previsible, que pueda dar lugar a riesgos para la salud y la seguridad o los derechos fundamentales».

En tercer lugar, aunque ya explicado *supra* conocer el grado de interpretabilidad y explicabilidad del sistema manteniendo resulta pertinente en contextos críticos donde una tasa de error a partir de un cierto umbral o la ausencia de un nivel adecuado interpretabilidad y explicabilidad tengan consecuencias adversas para la salud, la seguridad o los derechos fundamentales.¹¹⁶ De ahí que el artículo 13(3)(iv) incluya la

115. ISO/IEC DIS 12792:2024(en), 9.4.9. Según la norma, dichos mecanismos podrían incluir la existencia de bases de datos que agregan nueva información para el uso del modelo (no modificado); el uso de datos de producción para modificar el modelo en tiempo real; el almacenamiento y explotación de las operaciones del lado del responsable del despliegue (por ejemplo, la corrección de la decisión del modelo) u otras formas de retroalimentación para influir o modificar el comportamiento del modelo.

116. Por ejemplo, en contextos críticos como la asistencia sanitaria o la justicia penal. Cfr. Rudin, Cynthia, «Stop Explaining Black Box...», *Op cit.*, pp. 206-207.

información relativa a «las personas o grupos de personas específicos con los que se pretenda utilizar el sistema».

VI. VALORACIÓN FINAL DEL ARTÍCULO 13 DEL REGLAMENTO

A continuación se incorpora una valoración final del artículo 13 RIA relativa a su mejorable técnica redaccional o la posible desincentivación de los modelos de caja negra, también sobre la indefinición del «tipo y nivel de transparencia adecuados». Igualmente se valora el contenido de «minimis» de la transparencia material, sin concreción de su alcance, la normalización y transparencia de los sistemas de alto riesgo y, finalmente, los «grandes olvidados» del RIA.

1. UNA MEJORABLE TÉCNICA REDACCIONAL

La sistemática seguida y contenido del artículo 13 contiene algunas limitaciones que pueden plantear complejidad interpretativa:

— Remisiones explícitas e implícitas a otros preceptos del Reglamento o a terminología técnica específica que deben ser integrados en la interpretación del artículo 13.

— Uso inapropiado de conceptos técnicos o deficiente traducción de los mismos (al menos, en la versión en español (rendimiento/funcionamiento, métricas, parámetro, precisión).

— Recurso a expresiones ciertas expresiones ambiguas y abiertas («tipo y niveles de transparencia») o la falta de determinación del grado de detalle con el que debe que se debe describir la información referida en los apartados (3) (b-f) del artículo 13, lo que deja un amplio margen de libertad interpretativa al proveedor a la hora de concretar el contenido y alcance de la obligación de transparencia en las instrucciones de uso.

En la práctica, esto se traduce en una evidente inseguridad jurídica y en un nivel de cumplimiento variable por parte de proveedores que, dependiendo de su posición y fuerza en el mercado, podrían verse claramente desincentivados en cuanto al nivel de al nivel de transparencia necesario y adecuado para cumplir con las obligaciones del artículo 13 del Reglamento.

2. UNA MEJORABLE ARTICULACIÓN DE LAS RELACIONES ENTRE LA TRANSPARENCIA, LA INTERPRETABILIDAD Y LA EXPLICABILIDAD: ¿DESINCENTIVACIÓN DE LOS MODELOS DE CAJA NEGRA?

Aunque en su Considerando (27), el Reglamento incorpora una definición del término «transparencia», sin embargo, no hace lo propio con la «interpretabilidad» y la «explicabilidad». El Considerando (27) no incorpora propiamente una definición de la «transparencia» sino que identifica sus elementos integrantes (trazabilidad, explicabilidad y comunicación de información relevante). Mientras que la definición de la «transparencia» del Considerando (27) incluye la explicabilidad y obvia la interpretabilidad, el artículo 13 parece otorgar mayor relevancia a la interpretabilidad en detrimento de la explicabilidad.

El artículo 13 establece un tipo de transparencia técnica, interna, auto-referencial y circunscrita a la relación entre el proveedor y el responsable del despliegue. Desde la perspectiva del proveedor, la transparencia tendría como fin el cumplimiento normativo y acreditar dicho cumplimiento frente a las autoridades competentes. Desde la perspectiva del responsable del despliegue, la transparencia tendría como fines no sólo el cumplimiento normativo, sino la capacitación de este último para interpretar correctamente los resultados de salida y el uso adecuado del sistema de conformidad con las instrucciones de uso del proveedor.

No está clara la relación/distinción entre la interpretabilidad y la explicabilidad (por ejemplo, en el artículo 13(3)(d)). El artículo 13 parece incentivar los modelos interpretables intrínsecamente en los apartados (1) (3)(b)(vii) (3)(d), al menos con relación a los sistemas de alto riesgo, en lugar de modelos de caja negra necesitados de técnicas y herramientas de explicabilidad complementarias. Pero no está claro, en todo caso, si ésa ha sido la intención del legislador, puesto que tal planteamiento podría frenar la innovación. La dicción empleada por el artículo 13 parece sugerir que sólo sería exigible un nivel de explicabilidad local, en detrimento de la explicabilidad global, pues en sus distintos apartados la exigencia de interpretabilidad se limita exclusivamente a la «información de salida del sistema», excluyéndose así otros elementos del sistema.

3. INDEFINICIÓN DEL «TIPO Y NIVEL DE TRANSPARENCIA ADECUADOS»

Las posibles taxonomías y niveles de transparencia exigibles para cumplir con el artículo 13 no están definidas. Existen expresiones con un grado de indeterminación elevado, como «nivel de transparencia suficiente» o «tipo y un nivel de transparencia adecuados» (Art. 13.1), que parecen modular el requisito de transparencia en función de unos criterios no definidos en el texto, lo que, en la práctica, se traducirá en una inevitable inseguridad jurídica para el proveedor y el responsable del despliegue del sistema de IA de alto riesgo.

4. UN CONTENIDO DE «MINIMIS» DE LA TRANSPARENCIA MATERIAL Y SIN CONCRECIÓN DE SU ALCANCE

Las instrucciones de uso facilitadas al responsable del despliegue deberán incluir como mínimo información relativa a:

- La identidad y los datos de contacto del proveedor y, en su caso, de su representante autorizado (artículo 13(3)(a)).
- La idoneidad funcional del sistema, incluyendo sus características, capacidades y limitaciones (artículo 13(3)(b)).
- Los cambios en el sistema e idoneidad funcional predeterminados por el proveedor (artículo 13(3)(c)).
- Las medidas de vigilancia humana previstas, incluyendo técnicas que faciliten la interpretación de la información de salida del sistema (artículo 13(3)(d)).
- Los recursos computacionales, de hardware y vida útil prevista del sistema (artículo 13(3)(e)).
- Los controles de registros log implementados (artículo 13(3)(d)).

El artículo 13 no especifica criterios cuantitativos o cualitativos respecto del contenido o alcance de las categorías de información establecidas. No está claro aún el margen de libertad que podrán tener los proveedores a la hora de determinar el contenido y alcance de esta información. Ello podría favorecer interpretaciones restrictivas de cara a proteger los derechos de propiedad intelectual, industrial y la competitividad. Desde el ámbito de la normalización, existen ya normas técnicas aprobadas que establecen distintos niveles de transparencia en función de distintas taxonomías del nivel sistema (contexto, sistema propiamente dicho, modelo y datos), y de las categorías o roles de personas interesadas a las que se dirija la información relevante (usuario o responsable de despliegue del sistema, desarrolladores, auditores, autoridades de control, personas afectadas por los sistemas de IA, o público en general).

5. NORMALIZACIÓN Y TRANSPARENCIA DE LOS SISTEMAS DE ALTO RIESGO

La Comisión Europea ha mandado al CEN y al CENELEC para que desarrollen normas técnicas y documentos de normalización que concreten el contenido y alcance de los requerimientos de los sistemas de alto riesgos establecidos en la Sección 2ª del Capítulo 3, entre los que se encuentra el requisito de transparencia y comunicación de información del artículo 13 RIA. Será en la normalización donde va a tener lugar la verdadera elaboración de normas que concreten la aplicación del RIA y donde, en teoría deberá concretarse el tipo y nivel de transparencia. No está claro aún en qué medida la normalización es el instrumento adecuado para incorporar garantías técnico-jurídicas frente a los impactos adversos en los derechos fundamentales, y en tal caso, cómo ponderará los derechos frente la innovación y los intereses económicos de los operadores en el desarrollo y comercialización de la IA.

6. LOS «GRANDES OLVIDADOS» DEL REGLAMENTO

A pesar de las obligaciones de comunicación establecidas no sólo en el artículo 13, sino en otras disposiciones del Reglamento, la norma en ningún caso sería verdaderamente habilitante para el ejercicio de derechos por los afectados por los sistemas de alto riesgo, al no establecerse un marco claro que ofrezca a los particulares vías claras para impugnar las decisiones adoptadas por los sistemas de IA que les afecten. Ni el Artículo 50 (obligaciones de transparencia de los proveedores y responsables del despliegue de determinados sistemas de IA) ni el Artículo 86 (derecho a una explicación de las decisiones tomadas individualmente) garantizan que el público en general reciba información suficiente para comprender los riesgos a los que está sometido e impugnar de manera eficaz las decisiones individuales que causen efectos adversos en la salud y la seguridad o los derechos fundamentales.

La vigilancia o supervisión humana en el artículo 14 del Reglamento de inteligencia artificial: ¿un mero requisito obligatorio para los sistemas de alto riesgo?

GUILLERMO LAZCOZ MORATINOS

Centro de Investigación Biomédica en Red (CIBERER - ISCIII)
Instituto de Investigación Sanitaria Fundación Jiménez Díaz (IIS-FJD)

I. INTRODUCCIÓN

Resultaba más que previsible que la supervisión humana pasará a formar parte esencial de la regulación europea de la inteligencia artificial.

En 2019, el Grupo de Expertos de Alto Nivel sobre IA de la Comisión incluyó la supervisión humana como uno de los siete requisitos para el desarrollo de una IA fiable¹. En febrero de 2020, la Comisión estableció la supervisión humana como uno de los requisitos obligatorios para las aplicaciones de IA de alto riesgo en su Libro Blanco sobre IA². Ese mismo año, el Parlamento hizo lo mismo en su propuesta de Reglamento sobre los principios éticos para el desarrollo, despliegue y uso de la IA, la robótica y las tecnologías conexas³. Así pues, en abril de 2021, la supervisión humana pasó a formar parte de la primera versión del RIA por parte de la Comisión y se ha mantenido así con un amplio consenso hasta el día de hoy.

La supervisión humana se ha considerado un principio ético fundamental en los debates sobre la regulación de la IA. No obstante, ni la terminología (tal

1. Comisión Europea, Dirección General de Redes de Comunicación, Contenido y Tecnologías, *Directrices éticas para una IA fiable*, Oficina de Publicaciones, 2019. Disponible en: <https://data.europa.eu/doi/10.2759/14078>
2. Comisión Europea, «Libro Blanco sobre la inteligencia artificial — un enfoque europeo orientado a la excelencia y la confianza», Bruselas, COM(2020) 65 final, 19 de febrero de 2020. Disponible en: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligen-ce-feb2020_es.pdf, pp. 25-26.
3. Dicha propuesta de reglamento se incluía en el anexo de la Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas (2020/2012(INL)). Disponible en: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_ES.html

y como atestigua el título de este trabajo⁴, ¿vigilancia? ¿supervisión? ¿control? ¿intervención?) ni su definición guardan uniformidad entre juristas o en los documentos políticos de la UE. Sin embargo, la cuestión más general de si es necesario incluir este principio en la normativa parece estar fuera de discusión. Traducir este principio a una norma concreta y hacer que funcione de acuerdo con los objetivos fijados por los legisladores (bajar a la tierra la supervisión humana) parece una tarea más complicada.

Este texto analiza en primer lugar cómo la Comisión plasmó el principio de supervisión humana en la primera versión del RIA y, posteriormente, cómo fue debatido por las instituciones europeas a lo largo del camino legislativo hasta su versión final. Esta plasmación del principio en el reglamento es crucial. En palabras de Enqvist, el diseño específico de la obligación, sobre quién recae y en qué contextos aflora, es de gran importancia para evaluar qué impacto podría y puede tener la supervisión humana en la supervisión de los procesos del sistema de IA⁵.

Como veremos, la supervisión humana solo ha incorporado revisiones menores en esta trayectoria legislativa. Por un lado, una trayectoria tan incontrovertida podría significar que la Comisión adoptó pronto una integración satisfactoria de este principio. En este sentido, argumentaré que la primera versión incorporaba ya un requisito de supervisión humana bastante flexible para los sistemas de IA de alto riesgo, que resultará útil en diferentes contextos de toma de decisiones. Por otro lado, una vía legislativa sin controversias podría significar que no se han realizado las críticas oportunas. O, al menos, que no se ha observado plenamente la complejidad de «bajar a la tierra» la supervisión humana en el plano legislativo.

Por ello, la última parte de este texto pretende acercarse a este análisis jurídico a otras consideraciones que no se han hecho explícitas en esta vía legislativa. Los méritos, las limitaciones y los defectos de la supervisión humana en el RIA se explorarán a través de tres cuestiones que quedan abiertas para un debate más profundo. A saber, si los seres humanos pueden cumplir el objetivo normativo de la supervisión humana en el RIA, si los seres humanos son necesarios *in the loop* dentro de los procesos de toma de decisiones para garantizar la supervisión eficaz que exige el reglamento, y si está centrado en el ser humano más allá de la supervisión humana. A la luz de estas reflexiones, parece razonable afirmar que queda mucho trabajo interdisciplinar —y no sólo normativo— por hacer si no queremos que el RIA se convierta en otro fracaso de las políticas de supervisión humana de sistemas automatizados⁶.

-
4. Incluso la traducción del inglés «human oversight» ha resultado confusa. De la versión de la Comisión se tradujo como «vigilancia humana», término que prácticamente no había sido utilizado en la literatura o anteriores documentos, y aunque así se ha mantenido, la versión final incorpora varias referencias a la «supervisión humana» en paralelo. En este trabajo opto por el término «supervisión» dado que es el más extendido en la literatura, y también como traducción más habitual de «oversight» en distintos documentos de las instituciones de la UE.
 5. Enqvist, L., «“Human oversight” in the EU artificial intelligence act: what, when and by whom?», *Law, Innovation and Technology*, vol. 15, n.º 2 (2023), pp. 508-535.
 6. Entre otros trabajos críticos con estas políticas, destacaría: Green, B., «The flaws of policies requiring human oversight of government algorithms», *Computer Law & Security Review*, vol. 45 (2022), 105681; Huq, A. Z., «A Right to a Human Decision», *Virginia Law Review*, vol. 106, n.º 3 (2020), pp. 611-688.

II. LA VIGILANCIA O SUPERVISIÓN HUMANA EN LA PROPUESTA DE LA COMISIÓN DE ABRIL DE 2021

En esta sección explicaré cómo se ha integrado la supervisión humana en el RIA. Primero esbozaré la propuesta inicial de la Comisión, después destacaré las cuestiones planteadas por las distintas instituciones que han participado en el proceso legislativo y, por último, abordaré la versión final del texto.

La supervisión humana desempeña un papel destacado entre los requisitos obligatorios para los sistemas de IA de alto riesgo. Desde el principio, la Comisión estableció cuál es el objetivo normativo de la supervisión humana (prevenir o minimizar los riesgos), qué tipo de supervisión humana se requiere («efectiva por diseño») y qué requisitos debe cumplir la supervisión humana para ser considerada efectiva. Como veremos, los puntos estructurales de dicha propuesta inicial han permanecido intactos.

Aunque esto es aplicable a todos los requisitos obligatorios, cabe señalar que el RIA asigna al proveedor la mayor parte de las obligaciones relativas a estos requisitos. Esto es, antes de comercializar o poner en servicio sistemas de IA de alto riesgo, los proveedores deberán asegurarse de que sus sistemas de IA de alto riesgo cumplen los requisitos obligatorios, y demostrar su cumplimiento llevando a cabo un sistema de gestión de calidad, entre otras obligaciones. Como veremos, desde la primera redacción de la Comisión, las obligaciones de los implantadores para la fase de uso han aumentado.

Según el artículo 14 de la primera versión del RIA de la Comisión Europea, los sistemas de IA de alto riesgo se diseñarán y desarrollarán de forma que puedan ser vigilados de manera efectiva por personas físicas durante su fase de uso. En otras palabras, los sistemas de IA de alto riesgo deben permitir por diseño una supervisión humana efectiva.

Sin embargo, este mandato es sólo la punta del iceberg. Como dice Enqvist, la supervisión humana no es un requisito de «talla única», y puede tener diferentes orientaciones en relación con, entre otros, los aspectos del proceso de toma de decisiones de un sistema a los que se deba dirigir la supervisión, cuándo debe llevarse a cabo o quién es el ser humano que debe realizar la supervisión⁷.

Desde este punto de vista, más allá de este primer párrafo del artículo 14, hay mucho que desentrañar de este requisito de supervisión humana. Afortunadamente, y a diferencia, por ejemplo, del artículo 22 del RGPD, que apenas proporciona información sobre la intervención humana necesaria para las decisiones automatizadas, el RIA explica en detalle cómo debe garantizarse la supervisión humana.

1. SUPERVISIÓN HUMANA PARA PREVENIR O REDUCIR RIESGOS

Artículo 14(2) propuesta de RIA: *El objetivo de la vigilancia humana será prevenir o reducir al mínimo los riesgos para la salud, la seguridad o los derechos fundamentales que pueden surgir cuando un sistema de IA de alto riesgo se utiliza conforme a su finalidad*

7. Enqvist, L., «“Human oversight” in the EU artificial intelligence act: what, when and by whom?», *Law, Innovation and Technology*, vol. 15, n.º 2 (2023), pp. 508-535.

prevista o cuando se le da un uso indebido razonablemente previsible, en particular cuando dichos riesgos persisten a pesar de aplicar otros requisitos establecidos en el presente capítulo.

En la exposición de motivos, la Comisión argumenta que la supervisión humana a lo largo del ciclo de vida de los sistemas de IA tiene por objeto minimizar el riesgo de discriminación algorítmica, complementando la legislación de la Unión vigente en materia de no discriminación. Asimismo, el memorándum explica que, en áreas críticas como la educación y la formación, el empleo, servicios importantes, la aplicación de la ley y el poder judicial, la supervisión humana también reducirá las decisiones erróneas o sesgadas asistidas por IA, lo que facilitará el respeto de otros derechos fundamentales, además de la no discriminación.

Tanto en las Directrices éticas para una IA fiable como en el Libro Blanco sobre la IA encontramos que se menciona la supervisión humana como salvaguarda para evitar efectos perjudiciales de los sistemas de IA. En la literatura científica encontramos autores que afirman que los humanos son cruciales para evitar correlaciones indebidas y garantizar así la equidad en el análisis de datos⁸, y no sólo para excluir la discriminación, sino también para reducir los falsos positivos⁹.

Sin embargo, esta hipótesis ha sido fuertemente rebatida. Entre otros, Huq argumenta que la calidad defectuosa de la decisión de una máquina no implica que un humano lo haría mejor¹⁰ y que el problema de la igualdad debe abordarse por separado de cualquier derecho a una decisión humana¹¹. Además, si los humanos fracasan sistemáticamente en esta tarea, la supervisión humana provocará una falsa sensación de seguridad¹². Y Laux nos recuerda que ordenar por imperativo legal la supervisión humana no es la panacea para prevenir y minimizar los riesgos de la IA¹³.

Este debate, que afecta a la esencia misma de la propuesta, no debe pasarse por alto. Sobre todo, si la Comisión considera a los seres humanos como una especie de última llamada cuando fallan todas las demás salvaguardas, tal y como se desprende del final del segundo párrafo de este artículo.

2. SUPERVISIÓN HUMANA EFECTIVA DESDE EL DISEÑO

Aunque la supervisión humana era un requisito esencial en todos los antecedentes político-normativos de esta propuesta de Reglamento, lo cierto es que no encontramos

8. Favaretto, M., de Clercq, E., y Elger, B. S., «Big Data and discrimination: perils, promises and solutions. A systematic review», *Journal of Big Data*, vol. 6, n.º 1 (2019), pp. 1-27.

9. Roig, A., «Safeguards for the right not to be subject to a decision based solely on automated processing (Article 22 GDPR)», *European Journal of Law and Technology*, vol. 8, n.º 3 (2017), pp. 1-17.

10. Huq, A. Z., «A Right to a Human Decision», *Virginia Law Review*, vol. 106, n.º 3 (2020), pp. 611-688.

11. *Ibid.*

12. Green, B., «The flaws of policies requiring human oversight of government algorithms», *Computer Law & Security Review*, vol. 45 (2022), 105681.

13. Laux, J., «Institutionalised distrust and human oversight of artificial intelligence: towards a democratic design of AI governance under the European Union AI Act», *AI & SOCIETY* (2023), pp. 1-14.

en ellos consenso sobre qué tipo de supervisión humana debería exigirse. Mientras que el Libro Blanco sobre la IA mencionaba que el tipo y el grado adecuados de supervisión humana pueden variar de un caso a otro¹⁴, el texto del Parlamento Europeo afirmaba que los sistemas de IA de alto riesgo deben ser objeto de revisión, evaluación, intervención y control humanos significativos¹⁵.

Como ya se ha señalado, desde su propuesta inicial, el RIA exigió que los sistemas de alto riesgo de IA se diseñaran y desarrollaran de tal forma que pudieran ser supervisados de forma efectiva por personas físicas durante su fase de uso, incluyendo entre otros dotarlos de interfaz humano-máquina adecuada (art. 14(1) propuesta RIA). Así pues, el RIA exige que los sistemas de alto riesgo puedan ser objeto de una supervisión humana *efectiva*.

Establecer esta obligación desde el diseño y desarrollo del sistema supone que el cumplimiento de este requisito de supervisión humana debe garantizarse antes de comercializar el sistema de IA. La Comisión exige al proveedor que garantice que sus sistemas de IA cumplen los requisitos de alto riesgo (Art. 16(a) RIA) y que establezca un sistema de gestión de la calidad para documentar y demostrar el cumplimiento (Art. 17(1) RIA).

De este modo, el RIA establece un mecanismo de gobernanza para el diseño de los sistemas que no determina necesariamente cómo se aplicará la supervisión humana en la fase de uso del sistema de IA de alto riesgo.

Veamos el artículo 22 del RGPD para ilustrar la diferencia entre los dos mecanismos de gobernanza. Las decisiones basadas únicamente en el tratamiento automatizado están prohibidas con carácter general por el artículo 22, apartado 1, por lo que cualquier decisión que produzca un efecto jurídico o similar sobre el interesado debe incorporar una intervención humana al circuito de decisiones de tratamiento de datos. De este modo, el artículo 22.1 crea un mecanismo de gobernanza basado en la intervención humana para el tratamiento automatizado de datos personales. Y el RGPD establece que es el responsable del tratamiento en la fase de uso de un sistema de IA quien debe garantizar esta salvaguarda. Además, el responsable del tratamiento no puede eludir la prohibición fabricando artificialmente la intervención humana y, por lo tanto, los responsables del tratamiento deben garantizar que cualquier intervención humana sea significativa para el proceso de toma de decisiones¹⁶. Así pues, el tipo de supervisión-garantía que exige el RGPD es una intervención humana significativa para la fase de uso del sistema automatizado.

La Comisión, consciente de la diferencia entre estos dos mecanismos de gobernanza, establece un vínculo entre la supervisión humana en el RIA y los mecanismos de gobernanza en fase de uso, como el artículo 22 del GDPR. El artículo 29(1) de la propuesta de RIA establece, entre las obligaciones de quienes desplieguen

14. Comisión Europea, COM(2020) 65 final, p. 19.

15. Resolución del Parlamento Europeo, de 20 de octubre de 2020, 2020/2012(INL), Considerando 10.

16. Grupo de Trabajo sobre Protección de Datos del Artículo 29, «Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679» (2019), p. 21.

sistemas de IA de alto riesgo¹⁷, que deberán utilizar dichos sistemas de acuerdo con las instrucciones de uso que se abordarán a continuación. Sin embargo, de acuerdo con el apartado segundo de este artículo, esta obligación debe entenderse *sin perjuicio de otras obligaciones que el Derecho de la Unión o nacional imponga a los usuarios* (como el artículo 22 RGPD) *y no afectarán a su discrecionalidad para organizar sus recursos y actividades con el fin de aplicar las medidas de vigilancia humana que indique el proveedor.*

3. ¿CÓMO LOGRAR UNA SUPERVISIÓN HUMANA EFECTIVA?

Un mecanismo de gobernanza de este tipo no sólo tiene que ver con el tipo o clase de supervisión humana que se requiere. Es más, la exigencia de que el sistema pueda ser supervisado «de manera efectiva», como tal, no dice mucho. Y, de hecho, lo mismo puede decirse de otros términos utilizados en mecanismos de gobernanza similares, como «significativo», o incluso de las diferencias entre «supervisión» humana y «control», «vigilancia», «intervención» o «revisión». En este sentido, parece que el RIA quiere aportar seguridad jurídica a los proveedores a la hora de cumplir con el requisito de supervisión humana efectiva desde el diseño.

En primer lugar, establece qué deben hacer los proveedores para incluir la supervisión humana en el diseño y desarrollo de sus sistemas de IA.

A este respecto, en su apartado tercero, el artículo 14 de la propuesta de Reglamento indica que el proveedor deberá aplicar medidas para garantizar la supervisión humana de dos formas distintas; (a) identificando e incorporando dichas medidas, cuando sea técnicamente viable, en el sistema de IA de alto riesgo antes de su comercialización o puesta en servicio; y/o (b) identificando las medidas de supervisión humana antes de comercializar o poner en servicio el sistema y que sean adecuadas para ser aplicadas por el responsable del despliegue.

Estas medidas se complementarán con las «instrucciones de uso»¹⁸ que el proveedor deberá elaborar de conformidad con el requisito de transparencia. Es decir, las instrucciones que recibirá el responsable del despliegue para la fase de uso del sistema de IA de alto riesgo, deberán incluir las medidas de supervisión humana implantadas por el proveedor (Art. 13(3)(d) RIA).

En segundo lugar, el texto ofrece criterios para que los proveedores entiendan qué es una supervisión «efectiva».

Con este fin, la Comisión establece una serie de capacidades que, según las circunstancias, el ser humano asignado a supervisar el sistema de IA debe ser capaz de realizar durante su uso (art. 14(4) propuesta RIA). Entre otros, entender por completo las capacidades y limitaciones del sistema, ser conscientes del sesgo de automatización, interpretar correctamente la información de salida del sistema, desestimar, invalidar o revertir dicha información o interrumpir el sistema accionando un botón específicamente destinado a tal fin.

17. «Usuarios» conforme a la primera versión de la Comisión, «implementadores» en posteriores y «responsable del despliegue» en la versión final.

18. Art. 3(15) RIA: «Instrucciones de uso»: *la información facilitada por el proveedor para informar al responsable del despliegue, en particular, de la finalidad prevista y de la correcta utilización de un sistema de IA.*

Por lo tanto, para establecer una supervisión efectiva desde el diseño, el proveedor debe implantar o identificar medidas que permitan a los humanos —según las circunstancias— comprender e interpretar correctamente los resultados del sistema, y decidir cuándo no utilizar o incluso detener el sistema de IA.

4. EL ROL DEL RESPONSABLE DEL DESPLIEGUE EN LA SUPERVISIÓN HUMANA

Ya hemos visto que la Comisión centra las obligaciones de los sistemas de IA de alto riesgo en los proveedores. Esto no quiere decir que no existan obligaciones para los responsables del despliegue en el RIA, como las que recoge la sección 3 relativa a las obligaciones de distintos agentes en relación con estos sistemas.

La obligación principal es que los responsables del despliegue utilicen dichos sistemas de acuerdo con las instrucciones de uso que acompañan a los sistemas. Es decir, seguir las instrucciones relativas a las medidas de supervisión humana. No obstante, como ya se ha señalado, el seguimiento de las instrucciones se entiende sin perjuicio de otras obligaciones del responsable del despliegue en virtud del Derecho de la Unión o de los Estados miembros y de la discrecionalidad del usuario a la hora de organizar sus propios recursos y actividades. Para dar cumplimiento a esta disposición, entra en juego la aplicación del artículo 22 GDPR, pero también el artículo 11, apartado 1, de la Directiva (UE) 2016/680¹⁹ o el artículo 7, apartado 6, de la Directiva (UE) 2016/681²⁰, entre otros. Esto dependerá del contexto de uso del sistema de alto riesgo.

Además, otra obligación que puede afectar al modo en que se lleva a cabo la supervisión humana es que los responsables del despliegue se asegurarán de que los datos de entrada sean pertinentes a la vista de la finalidad prevista del sistema de IA de alto riesgo (art. 29(3) RIA). En contextos de toma de decisiones basadas en IA, puede asignarse a operadores humanos la función de revisar los datos de entrada para dichas decisiones. Especialmente cuando dichos datos puedan ser de categorías sensibles. Además, dichos controles podrían establecerse tanto antes de la toma de decisiones (*ex ante*), como después para corregir decisiones erróneas (*ex post*).

Siguiendo lo que De Hert y yo hemos argumentado en otro artículo²¹, el artículo 29, apartado 6, del RIA establece un vínculo entre las obligaciones del proveedor

19. Directive (EU) 2016/680. Artículo 11(1): *Los Estados miembros dispondrán la prohibición de las decisiones basadas únicamente en un tratamiento automatizado, incluida la elaboración de perfiles, que produzcan efectos jurídicos negativos para el interesado o le afecten significativamente, salvo que estén autorizadas por el Derecho de la Unión o del Estado miembro a la que esté sujeto el responsable del tratamiento y que establezca medidas adecuadas para salvaguardar los derechos y libertades del interesado, al menos el derecho a obtener la intervención humana por parte del responsable del tratamiento.*

20. Directive (EU) 2016/681. Artículo 7(6): *Las autoridades competentes no adoptarán ninguna decisión que produzca efectos jurídicos adversos para una persona o que afecte significativamente a una persona únicamente en razón del tratamiento automatizado de datos PNR. Dichas decisiones no deberán basarse en la raza o el origen étnico, las opiniones políticas, las creencias religiosas o filosóficas, la pertenencia a un sindicato, la salud o la vida u orientación sexual de la persona.*

21. Lazcoz, G., y de Hert, P., «Humans in the GDPR and RIA governance of automated and algorithmic systems. Essential pre-requisites against abdicating responsibili-

en virtud de este Reglamento y las obligaciones del responsable del despliegue en virtud del RGPD (cuando resulten a su vez responsables del tratamiento de datos conforme al mismo). Los responsables del despliegue de sistemas de IA de alto riesgo harán uso de la información facilitada por los proveedores de sistemas de IA en virtud del requisito de transparencia *para cumplir la obligación de llevar a cabo una evaluación de impacto relativa a la protección de datos que les imponen el artículo 35 del Reglamento (UE) 2016/679 (...), cuando corresponda*. Como se ha señalado anteriormente, esta información incluye medidas de supervisión humana. En otras palabras, los responsables del despliegue (responsables del tratamiento de datos) recibirán información técnica y organizativa (de los proveedores de sistemas de IA) sobre los sistemas de IA que adquieran y están obligados a hacer uso de esta información —para cumplir con el artículo 35 GDPR— que permite a las personas físicas (en la organización del responsable) a las que se asigna la supervisión humana en el artículo 22 RGPD comprender las capacidades y limitaciones del sistema. Consideramos prometedor este vínculo entre el RIA y el RGPD.

Por último, para sistemas destinados a utilizarse en la identificación biométrica remota «en tiempo real» o «en diferido» de personas físicas, la Comisión estableció lo que parece ser una obligación intermedia entre proveedores y responsables del despliegue. De acuerdo con este apartado, las medidas de los proveedores para estos sistemas deben garantizar que el responsable del despliegue no actúe ni tome ninguna decisión sobre identificación generada por el sistema, salvo que un mínimo de dos personas físicas la hayan verificado y confirmado.

III. LA TRAYECTORIA DE LA SUPERVISIÓN HUMANA EN EL PROCEDIMIENTO LEGISLATIVO ORDINARIO

A pesar de que pueden encontrarse otras enmiendas y propuestas en este proceso legislativo, esta sección se centra en las dos cuestiones clave que se han debatido sobre la supervisión humana en el RIA.

Por un lado, se ha debatido el derecho a la intervención humana. Como la Comisión se centró en establecer obligaciones para los proveedores que vayan a comercializar sistemas de alto riesgo, no declaró un derecho como tal a la supervisión humana en la fase de uso de los sistemas de IA. Así pues, la versión inicial obliga a los proveedores a comercializar sistemas de IA que puedan ser supervisados de manera efectiva por personas físicas. No obstante, el tipo de supervisión deberá determinarse en función de la normativa aplicable (por ejemplo, el artículo 22 del RGPD) y a discreción del propio responsable del despliegue. En contra de este planteamiento normativo, a lo largo del proceso legislativo se ha reclamado la inclusión en el RIA de un derecho de intervención humana para las decisiones de alto riesgo basadas en IA.

Por otra parte, también se ha debatido el papel de los humanos encargados de supervisar los sistemas de alto riesgo. La participación de los humanos en la toma de decisiones basada en la IA no remediará por arte de magia los efectos nocivos de estos sistemas automatizados. En este sentido, Matsumi y Solove definen con humor la forma en la que este rol normativo se establece habitualmente para las personas: *For human involvement to be the answer, the law must set forth exactly how humans would*

ties», *Computer Law & Security Review*, vol. 50 (2023), 105833.

ameliorate the problems with algorithmic predictions in particular cases. Instead, the law just points to a human and says: «Hey, there's a human, so all is fine» even though it remains unclear what the human is to do²². Este ha sido uno de los principales problemas de los mecanismos de gobernanza basados en la intervención humana a los que se ha hecho referencia anteriormente en este texto. Aunque la Comisión explica detalladamente los requisitos del sistema para permitir una supervisión humana eficaz, no dice nada acerca de los seres humanos a los que se encomienda esta tarea.

1. VOCES QUE RECLAMAN EL DERECHO A LA INTERVENCIÓN HUMANA (Y OTRAS GARANTÍAS) PARA LA TOMA DE DECISIONES BASADA EN SISTEMAS DE ALTO RIESGO

Para muchos autores, el talón de Aquiles de la propuesta de la Comisión residía en los (inexistentes) derechos de los individuos sometidos a decisiones basadas en la IA. Y no sólo derechos en relación con el uso de los sistemas, sino también mecanismos que permitiesen a los usuarios afectados por la IA exigir responsabilidades a los distintos actores implicados en el ciclo de vida de los sistemas. Y en este sentido se pronunciaron esas primeras críticas al RIA, por ejemplo, en palabras de Veale y Borgesius: «*As only those with obligations under the Draft AI Act can challenge regulators' decisions, rather than those whose fundamental rights deployed AI systems affect, the Draft AI Act lacks a bottom-up force to hold regulators to account for weak enforcement*²³.

En cuanto a la supervisión humana, se ha venido reclamando la inclusión en el RIA de diferentes derechos para la fase de uso basados en este requisito obligatorio. En particular, el derecho a la intervención humana, o derecho a un *human in the loop*, en los procesos de toma de decisiones con sistemas de IA de alto riesgo. Sin embargo, también ha habido referencias al derecho a una explicación o al derecho a impugnar las decisiones automatizadas, que estarían también mediados por operadores/ interventores humanos que adoptarían un papel de revisores.

En primer lugar, el Comité Europeo de las Regiones reclamó un derecho a la participación humana en toda decisión adoptada por sistemas de IA de alto riesgo. En los términos utilizados por el Comité, una decisión de esta clase *estará sujeta a intervención humana y se basará en un proceso de toma de decisiones riguroso. Debe garantizarse un contacto humano con estas decisiones*.²⁴

La opinión del Comité Económico y Social Europeo (EESC) va en la misma línea, aunque con argumentos más elaborados²⁵. El Comité se pregunta si estamos

22. Matsumi, H., y Solove, D. J., «The Prediction Society: Algorithms and the Problems of Forecasting the Future», *GWU Legal Studies Research Paper*, vol. 58 (2023), pp. 1-64.

23. Veale, M., & Zuiderveen Borgesius, F., «Demystifying the Draft EU Artificial Intelligence Act — Analysing the good, the bad, and the unclear elements of the proposed approach», *Computer Law Review International*, vol. 22, n.º 4 (2021), pp. 97-112.

24. Dictamen del Comité Europeo de las Regiones — Enfoque europeo de la inteligencia artificial — Ley de inteligencia artificial (Dictamen revisado). COR 2021/02682.

25. Dictamen del Comité Económico y Social Europeo sobre la «Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión» [COM(2021) 206 final — 2021/106 (COD)] EESC 2021/02482.

preparados para que la IA asuma sustancialmente el papel de la toma de decisiones humana, incluso en procesos críticos. Entre los ámbitos críticos en los que estas decisiones tienen un notable componente moral e implicaciones jurídicas o un impacto social, el EESC menciona el poder judicial, la aplicación de la ley, los servicios sociales, la asistencia sanitaria, la vivienda, los servicios financieros, las relaciones laborales y la educación. De ahí que el EESC recomiende que en dichos ámbitos las decisiones *sigan correspondiendo a las personas*.

En cuanto a las enmiendas aprobadas por el Parlamento Europeo, el considerando 58 bis de la propuesta destaca el papel fundamental de los responsables del despliegue a la hora de garantizar la protección de los derechos fundamentales. Dado que los responsables del despliegue son los mejor situados para comprender cómo se utilizará el sistema de IA en un contexto específico, deben identificar las estructuras de gobernanza adecuadas para ese contexto. En la primera lectura del Parlamento, dichas estructuras de gobernanza apropiadas incluyen: *los procedimientos de tramitación de reclamaciones y los procedimientos de recurso, ya que las opciones en las estructuras de gobernanza pueden ser decisivas para mitigar los riesgos para los derechos fundamentales en casos de uso concretos*²⁶.

Una vez más, surge un conflicto normativo entre establecer en el RIA un conjunto universal de mecanismos de gobernanza de la supervisión humana en la fase de uso o, por el contrario, permitir una mayor discrecionalidad en función del contexto de uso específico (que también puede estar limitada por la normativa aplicable en cada contexto).

Sin embargo, el Parlamento también quiso incluir en el RIA un derecho a una explicación de la toma de decisiones individuales. Según el artículo 68 *quater* de esta versión, cualquier persona afectada que sea objeto de una decisión basada en los resultados de un sistema de IA de alto riesgo que produzca efectos jurídicos o significativos *tendrá derecho a solicitar al implementador una explicación clara y significativa (...) sobre el papel del sistema de IA en el procedimiento de toma de decisiones, los principales parámetros de la decisión adoptada y los datos de entrada correspondientes*²⁷. A primera vista, este derecho parece vinculado a una lógica de la supervisión humana (conforma un requisito de transparencia mediado por una especie de supervisión humana del proceso de decisión) por la cual se acerque la decisión del sistema de IA a su comprensión por el individuo afectado.

Por último, cronológicamente hablando, el Dictamen 44/2023 del Supervisor Europeo de Protección de Datos (SEPD) también abordó esta cuestión y solicitó la inclusión de un derecho a obtener la intervención humana del usuario (final) del sistema de IA en relación con la toma de decisiones que le afecten y a impugnar el resultado de la toma de decisiones, así como derecho a recibir explicaciones del responsable del despliegue del sistema de IA sobre la toma de decisiones que le afecten significativamente²⁸. El SEPD considera que tales derechos no afectarían, sino que complementarían los derechos establecidos por el artículo 22 del RGPD, y otros

26. Enmienda 92.

27. Enmienda 630.

28. Supervisor Europeo de Protección de Datos, Opinion 44/2023 on the Proposal for Artificial Intelligence Act in the light of legislative developments, 23 de octubre de 2023, Bruselas, p. 25. Disponible en: <https://www.edps.europa.eu/data-protection/>

derechos establecidos por la legislación aplicable en cada contexto de uso, como el crédito al consumo, los servicios de seguros, el empleo, etc²⁹.

2. HUMANOS, ¿QUÉ HUMANOS?

Quién (y en qué condiciones) se encarga de supervisar un sistema de IA también fue objeto de debate en esta vía legislativa. Koulu explica que los documentos políticos de la UE crearon grandes expectativas sobre la supervisión humana para salvaguardar la autonomía de las personas, mistificando en cierto modo las capacidades humanas como última línea de defensa contra esta avalancha de inteligencia externa³⁰. Así, en esos documentos políticos de la UE no encontramos ningún debate sobre las implicaciones de la supervisión humana, si los supervisores humanos son capaces de realizar sus tareas de supervisión y de qué manera, ni cuáles serían los criterios para la intervención humana o si un supervisor debería poseer alguna experiencia particular³¹.

Si observamos la primera versión del RIA, podemos atenernos a la opinión de Koulu sobre los documentos políticos anteriores. Únicamente el considerando 48 de la propuesta de RIA se refiere a la capacidad de los seres humanos para realizar la tarea de supervisión: *Cuando proceda, dichas medidas deben garantizar, en concreto (...) y que las personas físicas a quienes se haya encomendado la vigilancia humana posean las competencias, la formación y la autoridad necesarias para desempeñar esa función*. Estas palabras, no obstante, no se traducen en ninguna obligación para los responsables del despliegue en el articulado de la propuesta de la Comisión.

En su dictamen conjunto de 2021, el SEPD y el CEPD abogaron por una verdadera centralidad humana que debería apoyarse en una supervisión humana altamente cualificada³². Entre las diversas salvaguardas necesarias para garantizar que se respeten y garanticen los derechos de los interesados y para evitar efectos negativos sobre las personas, en particular sobre la producción de decisiones sesgadas, el SEPD y la CEPD destacaron una supervisión humana cualificada en dichos procesos de toma de decisiones. Para dicho contexto, consideran que las autoridades competentes también deberán poder proponer directrices para evaluar los sesgos en los sistemas de IA y ayudar al ejercicio de la vigilancia humana³³.

En cuanto a las relaciones laborales, y dado que la supervisión correrá a cargo de un trabajador o un grupo de trabajadores, el Comité Económico y Social Europeo subrayó que estos trabajadores deberían recibir formación sobre cómo desempeñar

our-work/publications/opinions/2023-10-23-edps-opinion-442023-artificial-intelligence-act-light-legislative-developments_en

29. *Ibid*, pp. 17-18.

30. Koulu, R., «Human control over automation: EU policy and AI ethics», *Journal of Legal Studies*, vol. 1 (2020), pp. 9-46.

31. *Ibid*.

32. CEPD-SEPD, Dictamen conjunto 5/2021 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial), 18 de junio de 2021, Bruselas, p. 6. Disponible en: https://www.edpb.europa.eu/our-work-tools/our-documents/edpb-edps-joint-opinion/edpb-edps-joint-opinion-52021-proposal_es

33. *Ibid*, p. 17.

esta tarea: *Además, dado que se espera que estos trabajadores puedan ignorar el resultado del sistema de IA o incluso no utilizarlo, deben establecerse medidas para evitar el miedo a represalias (como la degradación o el despido) en caso de que se tome esa decisión (4.18)*³⁴. Por ello, el Comité pide medidas específicas para el proceso de toma de decisiones cuando los humanos consideren que deben desobedecer al sistema de IA de alto riesgo. Conceder esta autoridad a los trabajadores puede ser una condición necesaria para evitar el sesgo de la automatización.

Asimismo, el Comité Económico y Social Europeo señala que la propuesta de RIA de la Comisión carece de una visión de futuro que subraye el potencial de la IA para aumentar, y no sustituir, la toma de decisiones humana.

Cuando el texto llega a la primera lectura del Parlamento Europeo, estas ideas se traducen en la introducción de nuevas obligaciones para los responsables del despliegue. En esta versión del RIA, entre las obligaciones tanto de los proveedores como de los responsables del despliegue que figuran en el artículo 16, es necesario garantizar que *las personas físicas a las que se asigna la supervisión humana de los sistemas de IA de alto riesgo sean, en concreto, conscientes del riesgo de sesgo de automatización o confirmación*,³⁵. Esta disposición incluye, por tanto, medidas específicas por parte del proveedor para hacer posible este control de los sesgos de automatización y confirmación y, por parte del responsable del despliegue, medidas para garantizar que las personas concretas a las que se asigna la supervisión sean conscientes de estos riesgos.

Además, el Parlamento incluyó nuevas obligaciones para los responsables del despliegue en el artículo 29, según el cual deberán aplicar (i) *la supervisión humana con arreglo a los requisitos establecidos en el presente Reglamento*; velar por (ii) *que las personas físicas encargadas de la supervisión humana de los sistemas de IA de alto riesgo sean competentes, estén debidamente cualificadas y formadas, y dispongan de los recursos necesarios (...)*; y garantizar que (iii) *las medidas de solidez y ciberseguridad pertinentes y adecuadas sean objeto de un seguimiento periódico de la eficacia y se ajusten o actualicen periódicamente*³⁶. Este último párrafo es, por tanto, prometedor, ya que incluye las nociones de competencia, cualificación, formación y recursos necesarios para los humanos que se supone deben supervisar eficazmente los sistemas de IA de alto riesgo.

IV. LA VERSIÓN FINAL DE LA SUPERVISIÓN HUMANA EN EL REGLAMENTO DE UN VISTAZO

La versión final del RIA como resultado de este proceso legislativo arroja las siguientes características fundamentales de la supervisión humana.

— Sistemas de IA de alto riesgo que puedan ser vigilados de manera efectiva por seres humanos desde el diseño (art. 14(1) RIA).

— Vigilancia humana para prevenir y reducir riesgos, particularmente cuando el resto de las salvaguardas para sistemas de alto riesgo no sean efectivas (art. 14(2) RIA).

34. Comité Económico y Social Europeo, COM(2021)0206 — C9-0146/2021 — 2021/0106(COD).

35. Enmienda 334.

36. Enmienda 401.

— Medidas proporcionales a los riesgos, al nivel de autonomía y al contexto de uso del sistema que pueden integrarse técnicamente por el proveedor y/o definirse por el mismo para que las ponga en práctica el responsable de su despliegue (art. 14(3) RIA).

— Medidas dirigidas a que la persona encargada de la vigilancia entienda adecuadamente las capacidades y limitaciones del sistema, sea consciente del sesgo de automatización, interprete correctamente la información de salida del sistema, pueda decidir en una situación concreta no hacer uso del sistema y/o incluso detenerlo en caso de que sea necesario (art. 14(4) RIA).

— Responsables del despliegue que implementen medidas técnicas y organizativas adecuadas para garantizar que los sistemas sean supervisados de acuerdo a las instrucciones y medidas implementadas por el proveedor (art. 26(1) RIA).

— Supervisión humana encomendada a personas físicas que tengan la competencia, la formación y la autoridad necesarias (art. 26(2) RIA).

— Supervisión humana por diseño que no afecte al cumplimiento de obligaciones normativas nacionales y de la UE relativas a la toma de decisiones en la fase de uso, ni a la libertad para organizar los recursos y actividades del responsable del despliegue (art. 26(3) RIA).

— Sistemas de identificación biométrica remota limitados en la toma de decisiones a que al menos dos personas físicas la hayan verificado y confirmado por separado (art. 14(5) RIA).

V. ALGUNAS REFLEXIONES ACERCA DE QUÉ PODEMOS ESPERAR DE LA SUPERVISIÓN HUMANA EN EL REGLAMENTO

La UE no es la única institución gubernamental que pretende regular la IA, pero sí una de las más avanzadas en esta tarea. El potencial simbólico de este reglamento —buscado deliberadamente por el legislador europeo— tampoco es ningún secreto. En el caso de la supervisión humana, es probable que el artículo 14 del RIA constituya la base de la primera disposición general sobre la materia, por lo que es probable que atraiga mucha atención y sirva de ensayo de los requisitos generales de supervisión³⁷.

Por lo tanto, es necesario considerar qué podemos esperar de esta novedosa disposición que marcará la pauta para otras normativas. En esta sección, concluyo explorando algunas cuestiones abiertas sobre los méritos, limitaciones y defectos de la supervisión humana en el RIA.

La primera se refiere a las exigencias que tales mecanismos de gobernanza imponen a los seres humanos. Algunos autores han sido muy críticos con otras normativas que exigen la supervisión humana de los sistemas automatizados porque han resultado ineficaces, entre otras razones, por la incapacidad de los seres humanos para cumplir los objetivos normativos que plantean.

La segunda es sobre el tipo de supervisión humana que requiere el RIA, y sobre si este tipo de supervisión humana puede funcionar en el mundo real. Diseñar qué

37. Enqvist, L., «“Human oversight” in the EU artificial intelligence act: what, when and by whom?», *Law, Innovation and Technology*, vol. 15, n.º 2 (2023), pp. 508-535.

tipo de supervisión humana es adecuada para cada contexto de toma de decisiones no es tarea fácil³⁸. Por lo tanto, la regulación debe encontrar un difícil equilibrio entre garantizar la suficiente flexibilidad para diseñar la supervisión humana para cada contexto e imponer unas normas mínimas comunes de supervisión.

La última pregunta trata de explorar la racionalidad humana de la propia supervisión humana. El llamamiento a la supervisión humana se ha vinculado al desarrollo del concepto de una IA centrada en el ser humano. De hecho, se dice que la supervisión humana es un enfoque procedimental y reactivo de la IA «centrada en el ser humano»³⁹. Ahora bien, ¿cuáles son las implicaciones de la relación entre ambos conceptos y está la supervisión humana en el RIA centrada en el ser humano?

1. ¿PUEDEN LOS SERES HUMANOS CUMPLIR LA FINALIDAD NORMATIVA DE LA SUPERVISIÓN HUMANA EN EL REGLAMENTO?

Como se ha dicho, la supervisión humana en el RIA tiene por objeto prevenir o minimizar los riesgos para la salud, la seguridad o los derechos fundamentales.

Sin embargo, este planteamiento es controvertido. Huq sostiene que la calidad defectuosa de una decisión tomada por una máquina no garantiza que un ser humano lo hiciera mejor⁴⁰, y que el problema de la igualdad y no discriminación debe abordarse separadamente de cualquier derecho a una decisión humana⁴¹. De manera análoga, Green sostiene que las políticas de supervisión humana no están respaldadas por pruebas empíricas y, por tanto, es improbable que protejan contra los prejuicios de la toma de decisiones algorítmica⁴².

Por otra parte, los documentos jurídico-políticos de la UE crearon grandes expectativas sobre la supervisión humana para salvaguardar la autonomía humana en el desarrollo y uso de la IA, lo cual no es una buena idea según Koulu⁴³. El enfoque tecnológico de esos documentos acaba asignando subjetividad a la IA al tiempo que mistifica las capacidades humanas. De este modo, los supervisores humanos —agentes humanos implicados en los procesos de toma de decisiones de la IA— se presentan como la última línea de defensa contra la IA, mientras que

38. Yurrita, M., Draws, T., Balayn, A., Murray-Rust, D., Tintarev, N., y Bozzon, A., «Disentangling Fairness Perceptions in Algorithmic Decision-Making: The Effects of Explanations, Human Oversight, and Contestability», en *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023).

39. Enqvist, L., «“Human oversight” in the EU artificial intelligence act: what, when and by whom?», *Law, Innovation and Technology*, vol. 15, n.º 2 (2023), pp. 508-535.

40. Huq, A. Z., «A Right to a Human Decision», *Virginia Law Review*, vol. 106, n.º 3 (2020), pp. 611-688. No obstante, hay que evaluar en qué se basa la afirmación de que las máquinas superan a los humanos. Muchos estudios se basan en comparaciones entre el rendimiento de la IA y el del individuo que tienen poca o ninguna importancia práctica, vid. Cabitza, F., «Many say that AI can outperform human doctors. Is it true?», LinkedIn (2018). Disponible en: <https://www.linkedin.com/pulse/many-say-ai-can-outperform-human-doctors-true-federico-cabitza/>

41. Ibid.

42. Green, B., «The flaws of policies requiring human oversight of government algorithms», *Computer Law & Security Review*, vol. 45 (2022), 105681.

43. Koulu, R., «Human control over automation: EU policy and AI ethics», *Journal of Legal Studies*, vol. 1 (2020), pp. 9-46.

ésta se antropomorfiza en un agente autónomo que podría ser malicioso hacia los humanos⁴⁴.

En mi opinión, se trata de dos caras de la misma moneda.

Estoy de acuerdo con Green y otros autores que han analizado la supervisión humana en diferentes mecanismos de gobernanza, su funcionamiento en el mundo real dista mucho de ser óptimo y, como mínimo, podemos decir que no están cumpliendo los objetivos normativos para los que están concebidos. Esto no quiere decir que los humanos no puedan (no deban) desempeñar un papel decisivo en la toma de decisiones con sistemas de IA de alto riesgo. De hecho, no parece que los humanos que controlan estos procesos de toma de decisiones lo hayan hecho tan mal hasta ahora. Y nos hemos dotado de mecanismos legales «clásicos» para los casos en que ese control humano de la toma de decisiones es inadecuado o falla. Con la IA, este paradigma de la toma de decisiones humana parece cambiar. Sin embargo, ¿significa esto que los humanos no puedan contribuir a mejorar la toma de decisiones guiada por la IA en los contextos socioculturales en los que se aplica?

Llegados a este punto, lo que parece que necesitamos son mecanismos de gobernanza de supervisión humana basados en la evidencia. Es decir, no limitarnos a afirmar en abstracto que los humanos reducen los riesgos de la IA, sino dotarnos de mecanismos que lo hagan eficazmente. Esto significaría exigir vía normativa, a lo largo de todo el ciclo del sistema de IA, que se diseñe e implemente de tal forma que los humanos sean capaces de reducir de forma demostrable los riesgos implicados en el proceso de toma de decisiones. ¿Proporciona el RIA una supervisión humana basada en la evidencia para los sistemas de IA de alto riesgo?

2. ¿ES NECESARIO, EN VIRTUD DEL REGLAMENTO, QUE HAYA SERES HUMANOS IN THE LOOP EN LA TOMA DE DECISIONES CON INTELIGENCIA ARTIFICIAL DE ALTO RIESGO PARA GARANTIZAR LA SUPERVISIÓN HUMANA EFECTIVA EXIGIDA?

En respuesta a los temores de la automatización, la supervisión humana a nivel normativo se ha asociado históricamente con un ser humano que mantenga la última palabra sobre un sistema automatizado. De este modo, se garantiza que los resultados proporcionados por un sistema automatizado no sean la única razón para la toma de decisiones, ya que el operador humano puede cambiar los criterios del sistema hasta que se tome la decisión final⁴⁵. De esta forma, el «*human in the loop*» se ha convertido en una solución normativa habitual para resolver los problemas de transparencia, sesgos, seguridad jurídica y riesgos sistémicos relacionados con la automatización⁴⁶.

Sin embargo, la complejidad de los procesos híbridos de toma de decisiones hombre-máquina aumenta debido al progreso tecnológico y social. Por ejemplo, si nos fijamos en los escenarios de casos de uso presentados por Enarsson, Enqvist

44. Ibid.

45. Wagner, B., «Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision-Making Systems», *Policy & Internet*, vol. 11, n.º 1 (2019), pp. 104-122.

46. Enarsson, T., Enqvist, L., y Naarttijärvi, M., «Approaching the human in the loop — legal perspectives on hybrid human/algorithmic decision-making in three contexts», *Information & Communications Technology Law*, vol. 31, n.º 1 (2022), pp. 123-153.

y Naarttijärvi podemos concluir que las decisiones híbridas son una amalgama de cuestiones jurídicas, sociales, técnicas y organizativas⁴⁷. De ahí que resulte difícil encontrar una solución jurídica única, como dar la última palabra a las personas implicadas en dichos procesos.

El hecho es que, aunque los modernos sistemas de IA reducen considerablemente la importancia de los humanos en los procesos de toma de decisiones, los humanos siguen participando en ellos de innumerables maneras. Matsumi y Solove lo explican así: *There are humans behind every algorithmic prediction, much like the Wizard of Oz was a man operating a machine*⁴⁸. Así pues, la clave está en determinar, para diferentes contextos y entre todos los humanos implicados, quienes están ahí para garantizar el requisito de supervisión humana y qué se espera de ellos y ellas. Su papel no tiene por qué ser el de tener la última palabra en todas las decisiones. Como señalan Fosch-Villaronga y Malgieri, en determinados contextos la intervención humana directa podría ser ineficaz o incluso perjudicial⁴⁹.

Según el RIA, los proveedores deberán diseñar sistemas de IA que permitan a los humanos interpretar correctamente sus resultados o decidir no utilizarlos, entre otros. Sin embargo, el RIA no impone que el proceso de toma de decisiones mediante el sistema de IA tenga que producirse de una forma u otra.

Si consideramos el RIA como un punto de partida focalizado en cómo diseñar un sistema para que pueda ser supervisado eficazmente y para establecer una comunicación fluida entre proveedores y responsables del despliegue, se trata de una buena noticia. Teniendo en cuenta que el RIA es una normativa general para muchos tipos de sistemas de IA, una de sus fortalezas es que no limita el tipo de supervisión humana que debe aplicarse en la fase de uso.

La mala noticia es que debemos suponer que las normas existentes en los distintos contextos de aplicación son suficientes para proporcionar seguridad jurídica —y ya hemos concluido que no es el caso—. O que los responsables del tratamiento disponen de buenos recursos y orientaciones para aplicar este requisito en cada contexto —probablemente tampoco sea el caso—. El riesgo es que el RIA caiga en la larga serie de intentos normativos fallidos de abordar la compleja interacción humano-máquina⁵⁰.

De hecho, la excepción dentro del RIA a ese margen que da para la fase de uso se establece en relación con los sistemas de identificación biométrica remota. El Considerando 73 del RIA expone la necesidad de establecer un mecanismo «reforzado» para estos sistemas⁵¹, que consiste que el responsable del despliegue no

47. Ibid.

48. Matsumi, H., y Solove, D. J., «The Prediction Society: Algorithms and the Problems of Forecasting the Future», *GWU Legal Studies Research Paper*, vol. 58 (2023), pp. 1-64.

49. Fosch-Villaronga, E., y Malgieri, G., «Queering the ethics of AI», en *Handbook on the Ethics of Artificial Intelligence*. Edward Elgar Publishing (2024), forthcoming.

50. Beck, J., y Burri, T., «From “Human Control” in International Law to “Human Oversight” in the New EU Act on Artificial Intelligence». En D. Amoroso & F. Santoni de Sio (Eds.), *Research Handbook on Meaningful Human Control of Artificial Intelligence Systems*. Elgar (2023).

51. Entre los sistemas de identificación biométrica remota considerados de alto riesgo por el Anexo III, el art. 14(5) RIA establece la salvedad en la aplicación de este requisito para los ámbitos de aplicación de la ley, de migración, de control fronterizo o de

pueda actuar ni tomar ninguna decisión basándose en la identificación generada por el sistema, salvo si al menos dos personas físicas la han verificado y confirmado por separado. No deja de ser llamativo que el RIA considere que esta doble verificación por separado es un mecanismo «reforzado», ¿hay razones para pensar que el segundo verificador no incurrirá en el mismo sesgo o error que la primera persona encargada de la supervisión? ¿acaso este mecanismo de supervisión «por separado» puede considerarse más reforzado que un proceso de toma de decisiones «conjunta» que involucre dos supervisores humanos?

Así pues, con la llegada del RIA, seguimos echando en falta mecanismos de gobernanza de supervisión humana basados en la evidencia para la fase de uso de los sistemas de alto riesgo de IA, ¿qué legisladores tomarán la iniciativa al respecto? Además, también necesitamos recursos y orientación para que los responsables del despliegue apliquen la supervisión humana, ¿qué instituciones se encargarán de ello?

3. MÁS ALLÁ DE LA SUPERVISIÓN HUMANA, ¿TENEMOS UN REGLAMENTO CENTRADO EN EL SER HUMANO?

El concepto de IA centrada en el ser humano ha estado en el núcleo de los debates políticos sobre estas tecnologías. En el Libro Blanco sobre la IA, la Comisión apoyó firmemente un enfoque centrado en el ser humano como elemento clave para el futuro marco regulador. Además, declaró que el objetivo de una IA fiable, ética y centrada en el ser humano sólo puede alcanzarse garantizando una supervisión humana adecuada de las aplicaciones de IA de alto riesgo⁵².

A pesar de que este concepto no llegó a la propuesta de RIA de la Comisión, finalmente ha ocupado un lugar destacado en la versión final. Así, en su artículo primero se declara que el objetivo de este Reglamento es, entre otros, promover la adopción de una inteligencia artificial (IA) centrada en el ser humano y fiable. Como definición de este concepto dentro del propio RIA, únicamente encontramos en el Considerando 6 que la IA debe ser una herramienta para las personas y tener por objetivo último aumentar el bienestar humano.

Si acudimos a la literatura científica, Enqvist explica con agudeza que la supervisión humana desempeña un papel procedimental y reactivo en el concepto «centrado en el ser humano» de la IA. Mientras que el objetivo de las aplicaciones de IA centradas en el ser humano es satisfacer de forma proactiva —mediante el diseño— las necesidades y preferencias humanas en diferentes contextos, las medidas de supervisión humana tratan de abordar de forma reactiva los riesgos, sesgos y perjuicios de los sistemas de IA⁵³.

Así, podemos ver cómo este enfoque procedimental y reactivo se ha llevado al RIA a través de la supervisión humana como requisito obligatorio para los sistemas de alto riesgo. Por supuesto, no hay lugar aquí para debatir si el reglamento —en su conjunto— incorpora una política de IA centrada en el ser humano. Sin embargo,

asilo, en caso de que una norma de la UE o nacional considere que este requisito es desproporcionado.

52. Comisión Europea, COM(2020) 65 final, p. 21.

53. Enqvist, L., «“Human oversight” in the EU artificial intelligence act: what, when and by whom?», *Law, Innovation and Technology*, vol. 15, n.º 2 (2023), pp. 508-535.

sí parece apropiado considerar si la propia supervisión humana, como requisito obligatorio en el RIA, está centrada en el ser humano.

Este enfoque es muy relevante porque, de hecho, es muy probable que los supervisores humanos se encuentren en una situación vulnerable.

Por un lado, la evidencia muestra cómo las personas verán mermadas sus capacidades y habilidades al ser las supervisoras de la IA. Por ejemplo, con el uso de sistemas automatizados, los humanos no desarrollan las habilidades y conocimientos que normalmente se adquieren con la experiencia —el efecto «deskilling»⁵⁴. Además, la influencia dominante que la tecnología pueda tener sobre el usuario vendrá inevitablemente acompañada de efectos secundarios, como es el caso de los sesgos de complacencia y automatización⁵⁵. Más aún, las recomendaciones algorítmicas sesgadas podrían influir de forma negativa en el comportamiento humano a largo plazo, es decir, cuando el sistema automatizado se ha retirado del proceso de toma de decisiones⁵⁶.

Por otro lado, también hay efectos secundarios de tipo jurídico que afectarán a los supervisores humanos. Green sostiene que las disposiciones sobre supervisión humana trasladan la responsabilidad de los daños causados por la IA de los responsables de las instituciones que despliegan los sistemas (y que determinan la estructura de estos) a los operadores humanos de primera línea (que son relativamente impotentes en este sentido). Por lo tanto, las políticas de supervisión humana crean una laguna que permite a las empresas adoptar sistemas de IA defectuosos y evitar asumir la responsabilidad de los daños resultantes⁵⁷.

Tenemos que pensar qué valores humanos queremos preservar en nuestra interacción con la tecnología. En lo que respecta a la supervisión humana, esto significa pensar en el individuo supervisor no sólo como un intermediario entre el sistema de IA, la persona sobre la que se toman decisiones y los riesgos de esta interacción. En otras palabras, los supervisores humanos deben tratarse como un fin en sí mismos. Los legisladores, pero también proveedores y responsables del despliegue de los sistemas, deben reflexionar sobre cómo el ejercicio de esta función afecta a las personas a las que se encomienda esta supervisión en el desempeño de su trabajo, sus capacidades o su bienestar, entre otros. Y lo que es más importante, a sus derechos como trabajadoras y como seres humanos.

El respeto de la dignidad humana y la autonomía personal debe incluir a juezas, médicos, funcionarios públicos, moderadoras de contenidos, conductores o cualquier otra persona que deba supervisar sistemas de IA de alto riesgo.

54. Sutton, S. G., Arnold, V., y Holt, M., «How Much Automation Is Too Much? Keeping the Human Relevant in Knowledge Work», *Journal of Emerging Technologies in Accounting*, vol. 15, n.º 2 (2018), pp. 15-25.

55. Cabitza, F., Campagner, A., Angius, R., Natali, C., y Reverberi, C., «AI Shall Have No Dominion: On How to Measure Technology Dominance in AI-Supported Human Decision-Making». En *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023).

56. Vicente, L., y Matute, H., «Humans inherit artificial intelligence biases», *Scientific Reports*, vol. 13, n.º 1 (2023), 15737.

57. Green, B., «The flaws of policies requiring human oversight of government algorithms», *Computer Law & Security Review*, vol. 45 (2022), 105681.

VI. CONCLUSIONES

En la introducción de este trabajo destacaba lo previsible que resultaba que la supervisión humana pasará a formar parte esencial de la regulación europea de la inteligencia artificial. Como hemos visto, lo hace como un requisito de obligado cumplimiento para los sistemas de IA de alto riesgo.

Su tramitación legislativa no ha sido polémica y, salvo algunos detalles señalados más arriba, ha mantenido la esencia de la propuesta inicial de la Comisión hasta su aprobación definitiva. Los fundamentos de la vigilancia o supervisión humana como requisito de obligado cumplimiento obligan a los proveedores a establecer medidas desde el diseño que permitan que los sistemas sean supervisados de manera efectiva con el objetivo de reducir riesgos.

En las reflexiones del apartado tercero he querido poner de manifiesto algunos de las dificultades a las que se enfrentará este modelo de gobernanza establecido por el RIA. Entre las cuales quiero destacar, por un lado, la dificultad de que los seres humanos a quienes se encomiende la supervisión sean capaces de reducir los riesgos de estos sistemas sin modelos de gobernanza basados en la evidencia. Por otro lado, la dificultad de integrar, en contextos de uso tan diversos, una supervisión humana efectiva desde el diseño cuando la normativa en dichas fases se ha mostrado infructuosa y proveedores y responsables del despliegue cuentan con pocos recursos a los que atenerse.

En definitiva, aunque soy optimista en lo que respecta al modelo de supervisión humana establecido por el RIA, creo que requeriremos de esfuerzos por parte de todos los agentes implicados en el desarrollo, uso y gobernanza de los sistemas de IA para que ese optimismo se materialice en que las personas podamos reducir los riesgos de estos sistemas de alto riesgo de manera efectiva en diversos contextos.

Precisión y solidez de los sistemas de inteligencia artificial de alto riesgo en el artículo 15 del Reglamento

ANA ABA CATOIRA

Profesora Titular de Derecho Constitucional. Universidad de A Coruña¹

I. INTRODUCCIÓN A LA PRECISIÓN Y LA SOLIDEZ EN LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL DE ALTO RIESGO

Es indiscutible que la IA tiene un gran potencial transformador y que también plantea riesgos inherentes a su uso. Por otra parte, resulta innegable que la IA no se construye en un contexto ajeno a prácticas discriminatorias o poco equitativas². Dicho esto, la IA no se puede comprender única y exclusivamente como una técnica porque tiene una dimensión social y ética que supone que una IA confiable y responsable sea algo más que un buen sistema, esto es, un sistema que haga lo que se quiere que haga³. En este orden de cosas, aparecen como elementos imprescindibles la transparencia, la calidad de los conjuntos de datos, así como la prueba, evaluación, validación y verificación⁴.

1. El presente trabajo se realiza en el marco del Proyecto «Derechos y garantías públicas frente a las decisiones automatizadas y el sesgo y discriminación algorítmicas» 2023-2025 (PID2022-136439OB-I00) financiado por MCIN/AEI/10.13039/501100011033/FEDER, UE.
2. En este sentido, la brecha o brechas digitales cada vez más profundas nos indican que una gran parte de la población mundial no participa ni en el diseño ni en el desarrollo de la tecnología y, más concretamente, en la IA. Esta carencia de oportunidades es más evidente en las mujeres y otros grupos sociales históricamente discriminados, lo que pone el foco en esta idea, es decir, todo lo relacionado con el desarrollo tecnológico tiene una dimensión ética relevante, pues las consecuencias sociales de la ausencia de mujeres y de otros grupos sociales, o al menos su menor participación, en los desarrollos de IA serán evidentes, Aba Catoira, Ana, «Discriminación a través de datos públicos sin perspectiva de género y discriminación digital», *(Des)igualdad y violencia de género: el nudo gordiano de la sociedad globalizada*, Ramos Hernández, Pablo (coord.), Aranzadi, Madrid, (2020), pp. 29-51.
3. Aba Catoira, Ana, «La garantía de los derechos como respuesta frente a los retos tecnológicos», *Derecho Público de la inteligencia artificial*, Balaguer Callejón, Francisco y Cotino Hueso, Lorenzo (dirs.), Fundación Manuel Giménez Abad, Colección Obras Colectivas 27, Zaragoza, (2022), pp.57-84.
4. Instituto Nacional de Estándares y Tecnología (NIST), *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*.

En este sentido, el Instituto Nacional de Estándares y Tecnologías ha identificado las siguientes características técnicas y sociotécnicas necesarias para cultivar la confianza en los sistemas de IA: exactitud, explicabilidad e interpretabilidad, privacidad, fiabilidad, solidez, seguridad y resistencia de la seguridad, y que se mitiguen o controlen los sesgos perjudiciales⁵.

Un sistema de IA de alto riesgo lo es, precisamente, por los riesgos potenciales que su uso presenta para la salud, seguridad y los derechos fundamentales de las personas⁶. En este sentido, estos sistemas tienen que estar preparado para minimizar y prevenir estos riesgos o lo que es lo mismo los comportamientos perjudiciales e indeseables además de ser capaz de detectarlos cuando su funcionamiento tenga lugar fuera de aquel dominio de entrada y ejecución establecido por su finalidad prevista. De igual modo tiene que estar diseñado e implementado para evitar la adopción de decisiones equivocadas o generar una información de salida errónea. Todo ello para evitar consecuencias negativas para las personas⁷.

Pues bien, en el presente estudio nos vamos a ocupar de la regulación de los requisitos de precisión y solidez exigidos a los sistemas de inteligencia artificial de alto impacto o alto riesgo en el RIA analizando la propuesta legislativa presentada por el Parlamento Europeo y la Comisión Europea en abril de 2021⁸, así como la

5. Ver *Transparencia y explicabilidad de la inteligencia artificial*, Cotino Hueso, Lorenzo/ Castellanos Claramunt, Jorge (eds), Tirant lo Blanch, Valencia, (2022); Ortiz de Zárate Alcarazo, Luis, «Explicabilidad (de la inteligencia artificial)», *Eunomía. Revista en Cultura de la Legalidad*, n.º 22, (2022), pp. 328-344.
6. Estas cuestiones han sido abordadas en otros trabajos anteriores, Aba Catoira, Ana, «Derechos de igualdad, personas con discapacidad y mayores en el entorno digital (VIII, XI y XII)», *La Carta de Derechos Digitales*, Cotino Hueso, Lorenzo (coord.), Tirant lo Blanch, Valencia, (2022), pp. 123-154; «Las garantías de los derechos en el espacio digital: La constitucionalización de lo digital», *Inteligencia artificial y democracia: garantías, límites constitucionales y perspectiva ética ante la transformación digital*, Castellanos Claramunt, Jorge (coord.), Tirant lo Blanch, Valencia, (2023), pp. 87-114; «La era de la ciudadanía conectada: digitalización y retos del futuro desde una perspectiva de género», *Un estudio sobre el Estado autonómico: propuestas de mejora para el tercer decenio del Siglo XXI*, Castellanos Claramunt, Jorge (coord.), Tirant lo Blanch, Valencia, (2023), pp. 165-190.
7. El Reglamento de Inteligencia Artificial establece tres categorías de sistemas de IA en función del riesgo que son las prácticas prohibidas (capítulo II), los sistemas de alto riesgo (capítulo III) y los modelos de uso general (capítulo IV). Se clasifican aplicando un sistema de gestión de riesgos, clara influencia del Reglamento General de Protección de Datos junto a otros elementos propios del sector privado, y, en función de ese riesgo quedan sometidos, como se verá, a diferentes requisitos y obligaciones proporcionales al riesgo que supone su utilización para la salud, la seguridad y los derechos fundamentales.
La clasificación de los sistemas y, en concreto la regulación específica de los sistemas de inteligencia artificial de alto riesgo constituyen para Gamero «la clave de bóveda de toda la norma», Gamero Casado, Eduardo, «El enfoque europeo de Inteligencia Artificial», *Revista de Derecho Administrativo*, CDA, n.º 20, (2021), pp. 268-289, en concreto p. 277.
8. COM (2021) 206 final 2021/0106 (COD), de 21 de abril de 2021. Disponible en: https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0008.02/DOC_1&format=PDF

orientación general del Consejo sobre la propuesta aprobada por el Consejo de Transporte, Telecomunicaciones y Energía en su sesión n.º 3917 celebrada el 6 de diciembre de 2022, que establece la posición provisional del Consejo sobre esta propuesta y constituyó la base para los preparativos de las negociaciones con el Parlamento Europeo⁹. En esta orientación general se introducen novedades significativas, por ejemplo, una nueva definición de IA crucial para determinar el ámbito de aplicación de la normativa sobre IA o la referencia a los sistemas de IA de uso general (hasta el momento ausentes y causa de gran preocupación en la UE). El 11 de mayo de 2023 la Comisión de Mercado Interior y la Comisión de Libertades Civiles adoptaron un proyecto de mandato de negociación sobre dichas normas que permitió que se aprobara una nueva versión que incorpora las enmiendas aprobadas por el Parlamento Europeo el 14 de junio de 2023¹⁰. El 13 de febrero de 2024 se produjo la aprobación del ansiado texto.

En este estudio nos ceñiremos a analizar parte del contenido del artículo 15 (Sección 2ª, Capítulo III) concretamente la regulación de los requisitos de precisión y solidez exigibles a los sistemas de inteligencia artificial de alto riesgo¹¹ que son así

Los Anexos de la propuesta, disponibles en: https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0008.02/DOC_2&format=PDF

9. SECRETARÍA GENERAL DEL CONSEJO, Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Reglamento de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión — Orientación general (6 de diciembre de 2022). Disponible en: https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CONSIL:ST_15698_2022_INIT-
10. Disponible en: https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_ES.html
11. Requisitos que hunden sus raíces en las *Directrices Éticas para una IA fiable* elaboradas por el Grupo Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial, creado por la Comisión Europea en junio de 2018. La IA confiable debe respetar todas las leyes y regulaciones aplicables, debe respetar los principios y valores éticos, y debe ser robusta tanto desde una perspectiva técnica como teniendo en cuenta su entorno social. Además, las directrices presentan un conjunto de 7 requisitos clave que los sistemas de IA deben cumplir para ser considerados confiables: (i) acción y supervisión humanas, (ii) solidez técnica y seguridad, (iii) gestión de la privacidad y de los datos, (iv) transparencia, (v) diversidad, no discriminación y equidad, (vi) bienestar ambiental y social, y (vii) rendición de cuentas. Es dentro del requisito de transparencia donde se integra la necesidad de que los modelos de IA sean explicables, proponiéndose además un conjunto de criterios para evaluar en qué medida un modelo cumple con estos requisitos. UNIÓN EUROPEA, *Directrices Éticas para una IA fiable*. Comisión Europea, Bruselas, 2019. En todo caso, resaltar la relevancia del principio de la explicabilidad en cuanto que sobre él se sostienen todos los demás. En efecto, parece difícil pensar que la IA pueda ser justa si no está garantizada la explicabilidad del sistema, además de tener en cuenta que resulta esencial, no sólo desde una perspectiva ética, sino también en cuanto que es imprescindible para poder ejercer un control y exigencia de responsabilidad en el ámbito tecnológico base de la IA ética y confiable, en este sentido, ver también la *Recomendación sobre la ética de la IA* UNESCO (n.º 37), donde se afirma que la transparencia y la explicabilidad de los sistemas de IA suelen ser condiciones previas fundamentales para garantizar el respeto, la protección y la promoción de los derechos humanos, las libertades fundamentales y los principios éticos. Además, estos principios permiten conocer por qué

clasificados en base a su funcionalidad, finalidad y modalidades de uso, conforme a la legislación vigente relativa a la seguridad de los productos y a las modalidades de uso.

Dicho lo anterior, la importancia de las diversas cuestiones que ocupan el análisis de este trabajo resulta meridianamente clara. En efecto, la IA es un conjunto de tecnologías transformadoras que evolucionan de forma imparable. Los sistemas de recomendación personalizados, los asistentes virtuales, los automóviles autónomos o la predicción de infecciones, entre otras muchas aplicaciones, demuestran día a día su gran capacidad para mejorar la eficiencia en muchos ámbitos de la vida, pero, también, su potencial para generar riesgos y amenazas que de materializarse causarían daños, incluso irreparables, para los intereses públicos y los derechos de las personas (Considerandos 3 y 4)¹². Esto es así, en buena medida, porque los sistemas inteligentes aprenden y se adaptan continuamente a medida que procesan grandes cantidades de datos, por lo que su comportamiento puede variar y evolucionar con el paso del tiempo. A mayor abundamiento, los algoritmos pueden proporcionar decisiones basadas en modelos matemáticos sumamente complejos que presentan enormes dificultades en relación con su análisis y comprensión.

A esta necesidad responde la iniciativa europea para aprobar la primera Ley integral sobre IA a fin de proporcionar un marco adecuado de protección de las personas que deben ostentar el protagonismo en el desarrollo tecnológico y que, en este sentido, obligan a que la tecnología cumpla con altos estándares de confianza, seguridad y libertad¹³. El Reglamento pretende ofrecer este marco jurídico que define

-
- se toman determinadas decisiones y los procesos de adopción de estos de modo que dicha información da garantías para alegar o reclamar frente a las mismas.
12. Señala también el Real Decreto 817/2023, de 8 de noviembre, que establece un entorno controlado de pruebas para el ensayo del cumplimiento de la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial, <https://www.boe.es/eli/es/rd/2023/11/08/817>; «La inteligencia artificial es una tecnología disruptiva con una alta capacidad de impacto en la economía y la sociedad. En el plano económico, y junto a otras tecnologías digitales, presenta un alto potencial para el aumento de la productividad, la apertura de nuevas líneas de negocio, el desarrollo de nuevos productos o servicios —basados, por ejemplo, en la personalización, la optimización de los procesos industriales o las cadenas de valor—, la mejora en la facilidad de realización de tareas cotidianas, la automatización de ciertas tareas rutinarias y el desarrollo de la innovación. Este potencial incide positivamente en el crecimiento económico, la creación de empleo y el progreso social. No obstante, los sistemas de inteligencia artificial también pueden suponer riesgos sobre el respeto de los derechos fundamentales de la ciudadanía, como por ejemplo los relativos a la discriminación y a la protección de datos personales, o incluso causar problemas graves sobre la salud o la seguridad de la ciudadanía».
 13. El propósito de la Unión Europea es que este Reglamento «responda a un elevado nivel de protección de los intereses públicos, como la salud y la seguridad y la protección de los derechos fundamentales, incluidos la democracia, el Estado de Derecho y la protección del medio ambiente, tal como se reconocen y protegen en el Derecho de la Unión. Para alcanzar este objetivo, deben establecerse normas que regulen la comercialización, la puesta en servicio y la utilización de determinados sistemas de IA, garantizando así el buen funcionamiento del mercado interior y permitiendo que dichos sistemas se beneficien del principio de libre circulación de mercancías y servicios. Estas normas deben ser claras y sólidas en la protección de los derechos

normas armonizadas sobre esta materia orientadas al desarrollo tecnológico y que, al mismo tiempo, ofrecen un nivel elevado de protección de los intereses públicos y de los derechos fundamentales¹⁴. En concreto, los sistemas de inteligencia artificial clasificados como de alto riesgo quedan sometidos a un conjunto de requisitos y a unas obligaciones específicas a efectos de garantizar su correcto funcionamiento desde el punto de vista técnico y así evitar daños a la seguridad y a los derechos fundamentales.

Lo anterior se traduce en que los sistemas de alto riesgo tienen que ser sólidos, robustos y precisos, es decir, eficientes, de calidad, transparentes, confiables y estar preparados para prevenir y minimizar los comportamientos que puedan causar daños, así como las decisiones equivocadas o la generación de informaciones erróneas. Lo que determina la necesidad de implementar pruebas y evaluaciones que garanticen la precisión de la tecnología empleada que asegure su robustez y solidez, así como la confiabilidad muy conectada con exigencias éticas¹⁵.

En el artículo 15, objeto de este estudio, se contienen algunos de los requisitos técnicos de carácter imperativo para los sistemas de alto riesgo que han de cumplirse y deben ser sometidos a control. Este control consiste en un sometimiento a un examen técnico, previo a su comercialización, que demuestre que estos sistemas se han desarrollado siguiendo las exigencias técnicas establecidas en normas armonizadas fijadas por los organismos europeos de normalización o en especificaciones comunes elaboradas por la Comisión o en otras soluciones técnicas equivalentes generadas por operadores informáticos. En definitiva, los sistemas de alto riesgo se diseñarán y desarrollarán de forma que alcancen un nivel adecuado de precisión, solidez y ciberseguridad, y funcionen de forma coherente en esos aspectos a lo largo de su ciclo de vida (artículo 15)¹⁶.

fundamentales, apoyar las nuevas soluciones innovadoras, permitir un ecosistema europeo de agentes públicos y privados que creen sistemas de IA en consonancia con los valores de la Unión y liberar el potencial de la transformación digital en todas las regiones de la Unión» (Considerando 5).

14. Estas normas se alinean con la Carta de los Derechos Fundamentales de la Unión Europea, los compromisos comerciales internacionales de la Unión, además de que deben tener en cuenta la Declaración Europea sobre Derechos y Principios Digitales para la Década Digital (2023/C 23/01) y las Directrices Éticas para una IA digna de confianza del Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial.

Los siete principios básicos que la Comisión Europea considera necesario establecer y regular para una IA fiable son: acción y supervisión humanas; solidez técnica y seguridad; gestión de la privacidad y de los datos; transparencia; diversidad, no discriminación y equidad; bienestar social y medioambiental; y rendición de cuentas. Estos se recogen en el *Libro blanco de la UE sobre la inteligencia artificial: un enfoque europeo orientado a la excelencia y a la confianza*, Comunicación de la Comisión Europea COM (2020) 65 final, de 19 de febrero, p. 11; así como en la Comunicación de la Comisión COM (2019)168, de 8 de abril, p. 4.

15. Podríamos pensar en distintos escenarios con efectos negativos en las personas, más allá de las cajas negras, menos relacionadas con nuestro estudio, tenemos los falsos positivos y decisiones discriminatorias que son resultado de sistemas sesgados o los desarrollos inesperados o negativos una vez que se pone en producción el sistema de inteligencia artificial.
16. En palabras de Gamero Casado «Estos sistemas no están prohibidos, pero se sujetan a una serie de restricciones y a mecanismos de control ex ante y ex post mediante los

Como se comprueba los tres requisitos mantienen entre sí una relación directa, ya que la solidez requiere de precisión y ciberseguridad, igual que éstas de la primera, ya que la solidez del sistema de IA persigue mantener la precisión alcanzada en el inicio del ciclo de vida del sistema cuando se entrena, prueba y valida; y las medidas de ciberseguridad protegen el sistema frente a posibles ataques, por lo que garantizan su solidez y precisión.

De igual modo, resulta evidente la relación con la calidad de los datos utilizados para la programación, pues este es, sin duda, un gran reto a la hora de conseguir un buen funcionamiento de los sistemas de IA. El acceso a datos de calidad es fundamental para construir sistemas inteligentes robustos y eficientes, contando ya con importantes iniciativas europeas, tales como, la Estrategia de Ciberseguridad de la UE, la Ley de Servicios Digitales y la Ley de Mercados Digitales, y la Ley de Gobernanza de Datos, que pretenden proporcionar la infraestructura adecuada para la construcción de dichos sistemas.

Por otra parte, resulta necesaria una referencia al Real Decreto 817/2023, de 8 de noviembre, que establece un entorno controlado de pruebas para el ensayo del cumplimiento de la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial. En su artículo 11, ubicado en el Capítulo 3 dedicado al «Desarrollo de las pruebas, validación del cumplimiento, seguimiento e incidencias», establece los requisitos a cumplir durante el desarrollo de pruebas que coinciden con los que nos ocupan. En efecto, su apartado 1º señala que: «La participación en el entorno controlado de pruebas tendrá como objetivo cumplir, durante el transcurso de este, con la implementación de los siguientes requisitos: h) Los sistemas de inteligencia artificial habrán sido o serán diseñados y desarrollados de manera que consigan, teniendo en cuenta su finalidad prevista, un nivel adecuado de precisión, solidez y ciberseguridad. Estas dimensiones deberán funcionar de manera consistente a lo largo de su ciclo de vida».

En resumen, de lo anterior procede señalar que en este trabajo abordaremos distintas cuestiones entre las que destacamos:

- a) si a lo largo de la tramitación del Reglamento se han producido cambios o variaciones y, en su caso, cual ha sido su finalidad y justificación.
- b) cómo se precisan estos requisitos de calidad técnica que determinarán su cumplimiento en niveles de calidad adecuados.
- c) cómo se garantiza en el Reglamento la calidad de los datos utilizados para el entrenamiento, los controles sobre el entrenamiento y la construcción del modelo, las métricas de evaluación del modelo, el mantenimiento y la verificación de que ningún cambio comprometa el objetivo o la intención original del algoritmo. El papel esencial del monitoreo continuo de las métricas de desempeño del modelo y la detección de desviaciones del concepto.

Dada la naturaleza de este trabajo la metodología seguida ha consistido en abordar un análisis comparativo de la distinta normativa aplicable en materia de

que garantizar la aplicación efectiva del Reglamento. Se trata de una luz naranja en el semáforo, puesto que estos sistemas se pueden implantar siempre que se reúnan los requisitos que el propio Reglamento establece», *supra cit.*, p. 279.

IA siguiendo las diferentes versiones de propuestas, así como la doctrina que se ha pronunciado sobre la materia con el objeto de obtener unas conclusiones válidas aplicables a la comunidad científica internacional.

II. EVOLUCIÓN, TRAMITACIÓN Y CONTENIDO FINAL DEL ARTÍCULO 15

El artículo 15 establece determinados requisitos para los sistemas de alto riesgo exigiendo precisión, solidez y ciberseguridad durante todo el ciclo de su vida. A lo largo de su tramitación este artículo no ha sido objeto de modificaciones sustanciales salvo la introducción de nuevos párrafos que, esencialmente, han integrado el desarrollo contenido en los considerandos que si han sido objeto de una profunda revisión en el texto definitivo. A lo largo de su tramitación con la Orientación general de 6 de diciembre de 2022 se mantuvo el texto prácticamente inalterado si bien el texto de 14 de junio de 2023 si introducía modificaciones. El texto definitivo ha introducido cambios en algunos preceptos incluso modificando la numeración.

En este sentido, el antiguo considerando 43 se refería a que «Deben aplicarse requisitos a los sistemas de IA de alto riesgo en lo que respecta a la gestión de riesgos, la calidad y pertinencia de los conjuntos de datos utilizados, la documentación técnica y el mantenimiento de registros, la transparencia y el suministro de información a los usuarios, la supervisión humana, y la solidez, precisión y ciberseguridad. Estos requisitos son necesarios para mitigar eficazmente los riesgos para la salud, la seguridad y los derechos fundamentales, y no se dispone razonablemente de otras medidas menos restrictivas del comercio, evitando así restricciones injustificadas al comercio». En el texto definitivo el Considerando 46 establece que «La introducción en el mercado de la Unión, la puesta en servicio o la utilización de sistemas de IA de alto riesgo debe supeditarse al cumplimiento por su parte de determinados requisitos obligatorios, los cuales deben garantizar que los sistemas de IA de alto riesgo disponibles en la Unión o cuya información de salida se utilice en la Unión no entrañen riesgos inaceptables para intereses públicos importantes de la UE, reconocidos y protegidos por el Derecho de la Unión».

Asimismo, en las versiones anteriores el considerando 49 señalaba que los sistemas de IA de alto riesgo deben funcionar de manera consistente durante todo su ciclo de vida y presentar un nivel adecuado de precisión, solidez y ciberseguridad con arreglo al estado de la técnica generalmente reconocido existiendo el deber de comunicar a los usuarios el nivel de precisión y los parámetros empleados para medirla.

La solidez técnica se define como un requisito clave para los sistemas de IA de alto riesgo de modo que deben ser resistentes frente a comportamientos nocivos o indeseables que puedan derivarse de las limitaciones de los sistemas o del entorno en el que operan (por ejemplo, errores, fallos, incoherencias, situaciones inesperadas). En este sentido, se requiere la adopción de medidas técnicas y organizativas tanto en el diseño como en el desarrollo a fin de prevenir o minimizar los comportamientos nocivos o indeseables. Entre estas medidas se incluyen mecanismos que permitan interrumpir de forma segura el funcionamiento del sistema (planes a prueba de fallos) ante la presencia de determinadas anomalías o cuando el funcionamiento tenga lugar fuera de ciertos límites predeterminados.

Por otra parte, en el último de los textos previo al definitivo, el considerando 50 al referirse a la solidez técnica del sistema como requisito clave habla de garantizar la resiliencia frente a los riesgos asociados a las limitaciones del sistema, así como a acciones maliciosas que pueden poner en peligro su seguridad y dar lugar a conductas perjudiciales o indeseables por otros motivos. De modo que la incapacidad de protegerlos frente a estos riesgos podría tener consecuencias para la seguridad o afectar de manera negativa a los derechos fundamentales, por ejemplo, debido a la adopción de decisiones equivocadas o a que el sistema de IA en cuestión genere una información de salida errónea o sesgada¹⁷.

La enmienda número 86 propuesta por el Parlamento añadía que los usuarios deben tomar medidas para garantizar que la posible compensación entre solidez y precisión no conduzca a resultados discriminatorios o negativos para subgrupos minoritarios¹⁸.

Estos requisitos esenciales de precisión y solidez para mitigar eficazmente los riesgos para la salud, la seguridad y los derechos fundamentales tienen un significado diferente pero conectado. La precisión equivale a una medida cuantitativa de la relación entre el propósito previsto del sistema y su desempeño desde el diseño hasta que el sistema se ha puesto en funcionamiento que permite conocer el comportamiento del sistema de IA atendiendo a su finalidad prevista y al conjunto de datos con los que trabaja. Por tanto, cuando hablamos de precisión referida a un modelo de IA hablamos de la proporción de predicciones correctas realizadas por el modelo en comparación con el total de predicciones realizadas. En otras palabras, la precisión nos indica qué tan exactas son las predicciones de un modelo (accuracy) fundamental dentro del sistema de gestión de calidad junto a la robustez, ciberseguridad, transparencia, gobernanza del dato y la supervisión.

La seguridad y confiabilidad del sistema de IA depende directamente de su nivel de precisión y robustez indisolublemente unidos porque la precisión requiere que los sistemas sean sólidos y robustos y que, por tanto, resistan a los errores, fallos e incoherencias que puedan aparecer en los propios sistemas o en el entorno donde operan generalmente a causa de su interacción con personas físicas u otros sistemas (artículo 15.3º anteriores versiones). En la enmienda 325 se precisaba más este inciso 1º del párrafo 3º al establecer que se deben adoptar medidas técnicas y organizativas para garantizar la resistencia de los sistemas de IA de alto riesgo importando la terminología del Reglamento General de Protección de Datos y su principio transversal de responsabilidad proactiva.

Este párrafo 3º en relación con las medidas de solidez de los sistemas se refería expresamente a la adopción de soluciones de redundancia técnica, tales como copias de seguridad o planes de prevención contra fallos que la enmienda 326 dirigía

17. Por sesgos entendemos, siguiendo ISO (2006): *Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability*, ISO, Tech. Rep. ISO 3534-1:2006. Disponible en: <https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standards/04/01/40145.html>, «el grado en que un valor de referencia se desvía de la verdad» o «una inclinación que favorece o perjudica a una persona, objeto o posición» según las *Directrices éticas ... cit.*, p. 48.

18. En la versión de junio 2023 era el considerando 50 ahora en el texto definitivo este considerando se refiere a la ciberseguridad.

expresamente al proveedor pertinente, con la aportación del usuario. Estas soluciones dan forma a la responsabilidad de adoptar medidas técnicas que garanticen la solidez del sistema conforme a la finalidad prevista y sin olvidar los resultados no deseados previsibles. Este contenido en el texto definitivo se ha trasladado al párrafo 4º que también ha modificado su redacción.

Los considerandos desarrollan el contenido del articulado incluso van mucho más allá de lo que dice la norma, de ahí su especial interés pues son de gran ayuda para conocer la motivación del legislador. Pues bien, en el texto definitivo del Reglamento se han introducido numerosas modificaciones en los considerandos, hasta el punto de que todos los que contenían referencias a nuestro tema de estudio han cambiado totalmente en su contenido y en su enumeración.

Así, el considerando 59 respecto a los sistemas de inteligencia artificial con fines de aplicación de la ley, hace una referencia expresa a los requisitos obligatorios cuando señala que «En particular, si el sistema de IA no está entrenado con datos de buena calidad, no cumple los requisitos adecuados en términos de rendimiento, de precisión o de solidez, o no se diseña y prueba debidamente antes de introducirlo en el mercado o ponerlo en servicio, es posible que señale a personas de manera discriminatoria, incorrecta o injusta». Asimismo, en el considerando 61 se dice «En particular, a fin de hacer frente al riesgo de posibles sesgos, errores y opacidades, procede clasificar como de alto riesgo aquellos sistemas de IA destinados a ser utilizados por una autoridad judicial o en su nombre para ayudar a las autoridades judiciales a investigar e interpretar los hechos y el Derecho y a aplicar la ley a unos hechos concretos».

Como novedad, el considerando 64 dice que «Con el objetivo de mitigar los riesgos que presentan los sistemas de IA de alto riesgo que se introducen en el mercado o se ponen en servicio, y para garantizar un alto nivel de fiabilidad, deben aplicarse a los sistemas de IA de alto riesgo ciertos requisitos obligatorios que tengan en cuenta la finalidad prevista y el contexto del uso del sistema de IA y estén en consonancia con el sistema de gestión de riesgos que debe establecer el proveedor». En orden a la adopción de medidas la norma es flexible estableciendo que los proveedores «para cumplir los requisitos obligatorios del presente Reglamento deben tener en cuenta el estado actual de la técnica generalmente reconocido en materia de IA, ser proporcionadas y eficaces para alcanzar los objetivos del presente Reglamento». Además, teniendo presente que la comercialización o puesta en servicio de un producto se produce solo cuando este cumple la legislación de armonización de la Unión aplicable, las previsiones del Reglamento respecto a los requisitos a cumplir por los sistemas de alto riesgo se refieren a aspectos diferentes a los previstos en los actos de armonización de la Unión viniendo a completar dicha regulación sectorial¹⁹. En este sentido, dicho considerando pone como ejemplo «las máquinas o los productos sanitarios que incorporan un sistema de IA pueden presentar riesgos de los que no se ocupan los requisitos esenciales de salud y seguridad establecidos

19. Comunicación de la Comisión titulada «“Guía azul” sobre la aplicación de la normativa europea relativa a los productos, de 2022», la norma general es que la legislación de armonización de la Unión puede ser aplicable a un producto, ya que la comercialización o la puesta en servicio solamente puede producirse cuando el producto cumple toda la legislación de armonización de la Unión aplicable.

en la legislación armonizada pertinente de la Unión, ya que esa legislación sectorial no aborda los riesgos específicos de los sistemas de IA».

Ya en el considerando 66 se encuentran más referencias a los requisitos obligatorios que han de cumplir los sistemas de IA de alto riesgo a fin de reducir efectivamente los riesgos que puedan derivar de su uso siempre que no existan otras medidas menos restrictivas para el comercio. Así, «Deben aplicarse a los sistemas de IA de alto riesgo requisitos referentes a la gestión de riesgos, la calidad y la pertinencia de los conjuntos de datos utilizados, la documentación técnica y la conservación de registros, la transparencia y la comunicación de información a los responsables del despliegue, la supervisión humana, la solidez, la precisión y la ciberseguridad».

Como se ha señalado y se explicará más adelante la calidad de los datos es crucial para desarrollar sistemas de IA fiables y seguros, calidad que ha de mantenerse durante todo el ciclo vital del sistema para que este no se vaya degradando. Al respecto, el considerando 67 en el texto definitivo haciendo hincapié en las técnicas que implican entrenamiento de modelos relaciona la calidad de los datos con la garantía de que el sistema «funcione del modo previsto y en condiciones de seguridad y no se convierta en una fuente de algún tipo de discriminación prohibida por el Derecho de la Unión».

Los requisitos que han de cumplir los datos para reunir la calidad suficiente, las prácticas de gestión y gobernanza de datos para lograr que los conjuntos de datos para el entrenamiento, la validación y la prueba sean de alta calidad se regulan en el RIA con especial atención a la mitigación de sesgos que pueden repercutir negativamente en los derechos fundamentales o conducir a discriminaciones prohibidas por el Derecho de la Unión, especialmente cuando los datos de salida influyan en la información de entrada de futuras operaciones (bucles de retroalimentación).

III. LAS EXIGENCIAS DE UN NIVEL ADECUADO DE PRECISIÓN Y SOLIDEZ

Los sistemas de IA de alto riesgo se diseñarán y desarrollarán de modo que alcancen un nivel adecuado de precisión, solidez y ciberseguridad y funcionen de manera uniforme en esos sentidos durante todo su ciclo de vida (artículo 15.1º). Este nivel adecuado se determinará a la luz de su finalidad prevista y con arreglo al estado actual de la técnica generalmente reconocido (considerando 74). En versiones anteriores se hablaba de mantenimiento de un nivel adecuado a lo largo de su ciclo vital «que se determinará “a la luz de su finalidad prevista y de conformidad con el estado de la técnica generalmente reconocido”»²⁰.

En el texto definitivo del Reglamento se introdujo un nuevo párrafo 2º sobre los aspectos técnicos de la medición de los niveles exigidos de precisión y robustez (párrafo 1º) y respecto a cualquier otra métrica de rendimiento pertinente. En este sentido, señala que «la Comisión, en cooperación con las partes interesadas y las organizaciones pertinentes, como las autoridades de metrología y evaluación

20. Considerando 49 en versión anterior que se pretendió enmendar por el Parlamento con la enmienda 312 que se refería el estado de la técnica de acuerdo con el segmento de mercado o ámbito de aplicación específico.

comparativa, fomentará, según proceda, el desarrollo de parámetros de referencia y metodologías de medición»²¹.

En el considerando 74, ya citado, se explica como el Derecho de la Unión en materia de metrología legal, incluidas las Directivas 2014/31/UE y 2014/32/UE del Parlamento Europeo y del Consejo, tiene por objeto garantizar la precisión de las mediciones y contribuir a la transparencia y la equidad de las transacciones comerciales. De modo que, en este contexto, en cooperación con las partes interesadas y las organizaciones pertinentes, como las autoridades de metrología y de evaluación comparativa, la Comisión debe fomentar, según proceda, el desarrollo de parámetros de referencia y metodologías de medición para los sistemas de IA. Al hacerlo, la Comisión debe tomar nota de los socios internacionales que trabajan en la metrología y los indicadores de medición pertinentes relacionados con la IA y colaborar con ellos²².

El proveedor como responsable del diseño, implementación, verificación y validación del sistema de IA es el principal responsable para cumplir estos requerimientos durante todo el ciclo vital. Por tanto, es quien asume la responsabilidad en la adopción de las medidas adecuadas tanto, técnicas como organizativas, para garantizar que se cumple con las exigencias de precisión y solidez del sistema. Asimismo, dentro de su ámbito de aplicación, el usuario del sistema asume responsabilidades que se materializarán en medidas concretas técnicas y organizativas.

En todo caso, el nivel previsto de los parámetros de funcionamiento debe declararse en las instrucciones de uso que acompañen a los sistemas de IA. Se insta a los proveedores a que comuniquen dicha información a los responsables del despliegue de manera clara y fácilmente comprensible, sin malentendidos ni afirmaciones engañosas (considerando 74)²³. La exigencia del principio de transparencia como requisito de la calidad del sistema obliga a que en las instrucciones de uso que acompañen a los sistemas de IA de alto riesgo se indicarán los niveles de precisión de dichos sistemas, así como los parámetros pertinentes para evaluarla (artículo 15.3º)²⁴.

21. En versiones anteriores 15.1.bis y 15.1º.a.

22. En términos similares se expresaba el considerando 49.

23. El lenguaje ha de ser claro y libre de malentendidos o de afirmaciones engañosas tal como se propuso por el Parlamento en su enmienda 85 al considerando 49 del texto anterior. El considerando 49 se refería a la obligación de comunicar a los usuarios el nivel de precisión y los parámetros empleados para medir como condición imprescindible para cumplir los requisitos en el diseño y en el desarrollo del sistema. En este orden de cosas, el nivel esperado de las métricas de rendimiento debe declararse en las instrucciones de uso adjuntas, deben estar descritas en la documentación del sistema antes de diseñar otras pruebas que se desarrollarán en la etapa de ejecución y esta información debe comunicarse de forma clara y fácilmente comprensible, sin malentendidos ni declaraciones engañosas.

24. Sobre el nivel de transparencia de los sistemas de IA de alto riesgo, en lo que respecta a los contenidos de este trabajo, reseñar que el artículo 13 se refiere a las características, capacidades y limitaciones del funcionamiento del sistema de IA de alto riesgo, y en particular: i) su finalidad prevista; ii) el nivel de precisión (incluidos los parámetros para evaluarla), solidez y ciberseguridad mencionado en el artículo 15 con respecto al cual se haya probado y validado el sistema de IA de alto riesgo y que puede esperarse, así como cualquier circunstancia conocida y previsible que pueda

El cumplimiento de estos requisitos se concibe con flexibilidad en el sentido de que cabe la adopción de las soluciones técnicas a partir de normas u otras especificaciones técnicas o con arreglo a conocimientos científicos o de ingeniería generales a discreción del proveedor del sistema de IA de que se trate. En consecuencia, los proveedores de sistemas de IA podrían elegir cómo quieren cumplir los requisitos teniendo en cuenta el estado de la técnica y los avances en ese campo concreto²⁵. Si parece absolutamente necesario que exista coordinación de evaluaciones comparativas para determinar de qué manera se deben medir los requisitos exigidos²⁶.

1. MÉTRICAS Y RENDIMIENTO DEL SISTEMA

El proveedor, como responsable del diseño, implementación, verificación y validación del sistema de IA, debe cubrir estos requerimientos durante todo el ciclo vital del sistema, pues cualquier aspecto del ciclo de vida puede tener repercusión en la precisión del sistema. De modo que asume la responsabilidad de adoptar las medidas adecuadas (tanto organizativas como técnicas) para garantizar que se cumple con los requerimientos mínimos establecidos en el artículo 15.

Ahora bien, aun cuando la precisión del sistema debe establecerse o cuantificarse a lo largo de todo el ciclo de vida del sistema, hay determinadas etapas de especial relevancia, concretamente en lo que toca a la selección de datos para el entrenamiento del sistema que deben ser datos de calidad. Por el contrario, la utilización de datos

afectar al nivel de precisión, solidez y ciberseguridad esperado; iii) cualquier circunstancia conocida o previsible, asociada a la utilización del sistema de IA de alto riesgo conforme a su finalidad prevista o a un uso indebido razonablemente previsible, que pueda dar lugar a riesgos para la salud y la seguridad o los derechos fundamentales a que se refiere el artículo 9, apartado 2; iv) en su caso, las capacidades y características técnicas del sistema de IA de alto riesgo para proporcionar información pertinente para explicar su información de salida; v) cuando proceda, su funcionamiento con respecto a personas o grupos de personas específicos en relación con los que esté previsto utilizar el sistema; vi) cuando proceda, especificaciones relativas a los datos de entrada, o cualquier otra información pertinente en relación con los conjuntos de datos de entrenamiento, validación y prueba usados, teniendo en cuenta la finalidad prevista del sistema de IA; vii) en su caso, información que permita a los responsables del despliegue interpretar la información de salida del sistema de IA de alto riesgo y utilizarla adecuadamente; c) los cambios en el sistema de IA de alto riesgo y su funcionamiento predeterminados por el proveedor en el momento de efectuar la evaluación de la conformidad inicial, en su caso; d) las medidas de vigilancia humana a que se hace referencia en el artículo 14, incluidas las medidas técnicas establecidas para facilitar la interpretación de la información de salida de los sistemas de IA de alto riesgo por parte de los responsables del despliegue; e) los recursos informáticos y de hardware necesarios, la vida útil prevista del sistema de IA de alto riesgo y las medidas de mantenimiento y cuidado necesarias (incluida su frecuencia) para garantizar el correcto funcionamiento de dicho sistema, también en lo que respecta a las actualizaciones del software.

25. Considerando 50 en su redacción anterior.

26. A pesar de la existencia de organizaciones de normalización para establecer normas, la coordinación es necesaria, correspondiéndole a la Oficina Europea de IA convocar a las autoridades nacionales e internacionales de metrología y de evaluación comparativa para facilitar orientaciones no vinculantes para abordar los aspectos técnicos de la medición de los niveles adecuados de precisión y solidez.

erróneos, incompletos o sesgados o las correlaciones falsas repercuten negativamente en el sistema en términos de confianza y fiabilidad porque impiden el objetivo de lograr «una mayor eficiencia, precisión, escala y velocidad de la IA para tomar decisiones y encontrar las mejores respuestas»²⁷. En palabras del Parlamento Europeo, dado que «los datos de capacitación a menudo son de una calidad cuestionable y no son neutrales»²⁸, la «baja calidad» de los datos o los procedimientos «podrían dar lugar a algoritmos sesgados, correlaciones falsas, errores, una subestimación de las repercusiones éticas, sociales y legales»²⁹ y, en definitiva, arrojar decisiones que afectan negativamente a las personas y que puedan incluso generar discriminación (algorítmica). Un resultado indeseado y que el Reglamento pretende evitar con la regulación de estos requisitos técnicos.

En el artículo 10 del Reglamento se regulan los requisitos que deben cumplir los datos de calidad. En este sentido, la calidad de los datos, que exige la norma, implica que los conjuntos de datos de entrenamiento, validación y prueba sean pertinentes y representativos, no contengan errores y sean completos en función de la finalidad prevista del sistema. Asimismo, deben tener las propiedades estadísticas adecuadas, también en lo que respecta a las personas o los grupos en los que en un principio se usará el sistema de IA de alto riesgo. En concreto, los conjuntos de datos de entrenamiento, validación y prueba deben tener en cuenta, en la medida necesaria en función de su finalidad prevista, los rasgos, características o elementos particulares del entorno o contexto geográfico, conductual o funcional específico en el que se pretende utilizar el sistema de IA. Con el fin de proteger los derechos de terceros frente a la discriminación que podría provocar el sesgo de los sistemas de IA, los proveedores deben ser capaces de tratar también categorías especiales de datos personales, como cuestión de interés público esencial, para garantizar que el sesgo de los sistemas de IA de alto riesgo se vigile, detecte y corrija.

A efectos de mejorar la calidad de los datos, esencial para el buen funcionamiento del sistema de IA, sería interesante seguir la recomendación de autores como Floridi³⁰, con una apuesta por cambiar el proceso de obtención de los datos. Se trataría de abandonar el Big Data optando por la calidad del dato, siendo a tal efecto más relevante dejar de trabajar con ingentes cantidades de datos para elegir conjuntos más pequeños, pero de más calidad. La mayor calidad se garantiza a través de una cuidada selección de los datos y con la fiabilidad que ello supone porque los algoritmos se entrenarían con mejores datos que ya no tenderían a ser inexactos, erróneos o contener sesgos. Esta solución puede aportarla la utilización de datos generados por sistemas de IA que cumplan con los estándares exigidos por el artículo 10 del Reglamento.

27. Foro Económico Mundial 2018, p.8.

28. PARLAMENTO EUROPEO (2017): Resolución de 14 de marzo de 2017, *sobre las implicaciones de los macrodatos en los derechos fundamentales: privacidad, protección de datos, no discriminación, seguridad y aplicación de la ley* (2016/2225(INI)), letra B. Disponible en: https://www.europarl.europa.eu/doceo/document/TA-8-2017-0076_ES.html

29. Parlamento Europeo (2017), *cit.*, Considerando m.

30. Floridi, Luciano, «The Fight for Digital Sovereignty: What It Is, and Why It Matters, Especially for the EU» en *Philos. Technol*, n.º 33, pp. 369-378 (2020). Disponible en: <https://doi.org/10.1007/s13347-020-00423-6>

En este orden de cosas, dentro de las medidas técnicas y organizativas que deben establecer los proveedores del sistema encontramos aquellas destinadas a seleccionar y evaluar las métricas de precisión desde el diseño del sistema, así como los controles de calidad del sistema de cuyos resultados depende la verificación y validación de dichas métricas³¹. La selección en todo caso debe hacerse en función de varios elementos como la finalidad y la evitación o mitigación de discriminaciones o sesgos³².

Las métricas de rendimiento permiten realizar la evaluación del rendimiento de los algoritmos de aprendizaje automático, cuantificando la calidad de las predicciones, con el fin de mitigar los riesgos potenciales que el sistema representa. Las distintas métricas proporcionan una perspectiva o visión diferente sobre el rendimiento del modelo, por lo que es importante elegir la métrica más adecuada para cada tarea.

Así lo anterior, la precisión (exactitud) del modelo es una métrica de uso común que mide la proporción de predicciones correctas realizadas por el modelo. Esta técnica puede resultar útil en determinados supuestos y en otros no tanto porque se produce un desequilibrio de clases significativo y la precisión no proporciona una representación fiel del rendimiento del modelo. En cualquier caso, la precisión es una métrica fundamental porque indica el nivel de exactitud de las predicciones realizadas por un modelo en comparación con el total de predicciones. Así las cosas, una alta precisión garantiza resultados confiables y puede marcar la diferencia en aplicaciones críticas mientras que una baja precisión puede tener consecuencias graves. Por otra parte, el cálculo de la precisión puede variar dependiendo del problema y del enfoque utilizado.

Como se ha indicado en algunos casos es conveniente recurrir a otras métricas o incluso a la combinación de varias como sería utilizar la precisión con la recuperación, concretamente en los casos de clasificación cuando se trata de asignar una etiqueta o categoría a una entrada porque la precisión mide la proporción de predicciones positivas verdaderas entre todas las predicciones positivas y la recuperación o recuerdo mide la proporción de predicciones positivas verdaderas entre todas las instancias positivas reales.

En el abanico de opciones está la métrica de puntuación F1 que combina precisión y recuperación, proporcionando un valor único que equilibra el equilibrio entre estas dos métricas, la puntuación oscila entre 0 y 1, siendo 0 el peor rendimiento y 1 el mejor; la curva ROC que traza la tasa de verdaderos positivos (sensibilidad) frente a la tasa de falsos positivos (1-especificidad) para diferentes umbrales de clasificación; o el AUC-ROC que proporciona un valor único que representa el rendimiento general del modelo en todos los umbrales de clasificación posibles y cuanto más alto sea el valor mejor será el rendimiento.

31. US National Institute of Standards and Technology (NIST). AI Measurement and Evaluation. <https://www.nist.gov/ai-measurement-and-evaluation>; OECD.AI.CatalogueofTools&MetricsforTrustworthyAI. <https://oecd.ai/en/catalogue/metrics>, 2023; IEEEStandardsAssociation. IEEEportfolioofAIStechologyandimpactstandard-sandstandardsprojects. <https://standards.ieee.org/initiatives/autonomous-intelligence-systems/standards/>

32. La garantía de la precisión del modelo depende directamente de la calidad de los datos de entrenamiento que deben ser representativos de la finalidad prevista y libres de sesgos, ver artículo 10 RIA.

Asimismo, al proveedor le corresponde la adopción de otras medidas técnicas en materia de precisión como la elaboración de la documentación técnica que contenga toda la información necesaria para la correcta puesta en funcionamiento del sistema y, en su caso, detección y comunicación de errores. La comunicación de las métricas y el rendimiento del sistema al usuario³³ es, en todo caso, responsabilidad del proveedor que debe cumplir con el principio de transparencia facilitando toda la información sobre el sistema como indicador de calidad.

Una vez que se han establecido las métricas se deben implementar otras medidas organizativas que permitan hacer un seguimiento de la precisión del modelo y, por consiguiente, si funciona con consistencia, es decir, si es un modelo sólido y robusto. En este sentido, las métricas son un indicador de calidad y un mínimo a cumplir, por lo que si no es posible garantizarlas se dará paso a la supervisión humana.

Precisión y solidez aparecen como requisitos inseparables, pues junto a la ciberseguridad son esenciales en los sistemas de inteligencia artificial de alto riesgo. De hecho, tal como se indicó en páginas anteriores, uno de los objetivos de la solidez del sistema es garantizar la precisión durante todo su ciclo de vida. El artículo 15 establece la responsabilidad de los proveedores de dotar de solidez al sistema, acorde a la finalidad prevista con el objetivo de mitigar los riesgos identificados en el plan de riesgos, que ha de mantenerse en nivel adecuado y de forma consistente a lo largo de todo su ciclo de vida.

Se trata de que los sistemas sean resilientes, resistiendo lo más posible frente a comportamientos perjudiciales o indeseables por motivos diversos como limitaciones en los sistemas o del entorno en el que funcionan, particularmente a causa de su interacción con personas físicas u otros sistemas (artículo 15.4º, considerando 75).

El Reglamento prevé que el proveedor establezca medidas como las soluciones técnicas de redundancia que pueden incluir planes de respaldo o a prueba de fallos como herramientas para garantizar la solidez y calidad del sistema. Por ejemplo, la copia de datos que asegura la redundancia de modelos, algoritmos, datos, etc; los mecanismos de fallo seguro de los componentes durante todo el ciclo vital; o la implantación de un plan de actuación cuando falla el sistema con el fin de la recuperación de los elementos o la reproducción de los datos (artículo 15.4º).

El hecho de no adoptar medidas de protección frente a estos riesgos podría tener consecuencias para la seguridad o afectar de manera negativa a los derechos fundamentales, por ejemplo, debido a decisiones equivocadas o información de salida errónea o sesgada ofrecidas por el sistema de IA. Por consiguiente, a nivel organizativo se ha de monitorear la degradación del sistema a lo largo de las diferentes etapas de su ciclo de vida, controlando la calidad constante de los datos, previendo los olvidos catastróficos y teniendo en cuenta la retroalimentación (artículo 15.4º RIA).

Estas medidas técnicas adoptadas para garantizar la solidez del sistema, a través de la prevención y minimización de los comportamientos perjudiciales o indeseables del sistema, deben estar recogidas documentalmete para facilitar al usuario las herramientas o mecanismos adecuados para observar, supervisar y reportar diferentes tipos de degradación del modelo que sobrepasen los límites

33. Las empresas usuarias del sistema de inteligencia artificial tienen la responsabilidad de conocer el nivel de precisión y contar con personal capacitado en la organización.

documentados como razonables para cada métrica de solidez, con el fin de hacerlos reproducibles para su corrección. Además, si los requisitos de solidez garantizados cambian tendrá que intervenir para corregirlos en aras de garantizar las métricas recogidas en la documentación.

En la construcción de sistemas robustos de IA resulta fundamental el análisis y la comprensión de los problemas que surgen en relación con los datos de producción que provienen del mundo real y con los datos de salida que son resultados ofrecidos por el modelo dada su repercusión directa en el nivel de precisión del sistema y, por consiguiente, en su nivel de confiabilidad.

Lo anterior presenta especial problemática en relación a los sistemas de IA de alto riesgo que siguen aprendiendo tras su introducción en el mercado o puesta en servicio, pues con el tiempo cambian o modifican su comportamiento creando escenarios donde introducen errores, fallos o resultados sesgados, no previstos, que influirán en los datos de entrada de futuras operaciones (bucles de retroalimentación) y que suponen una alteración de su solidez y de su precisión. En este sentido, estos sistemas deben desarrollarse de modo que se elimine o reduzca lo máximo posible el deterioro de las métricas de precisión, solidez y rendimiento con la adopción de las medidas de reducción de riesgos oportunas (artículo 15.4º).

La retroalimentación del sistema es un proceso iterativo en el que las decisiones y los resultados de un modelo se recopilan y utilizan continuamente, con una actualización constante de los datos de entrenamiento, los parámetros del modelo y los algoritmos del sistema a efectos de mejorar su rendimiento. El riesgo de deterioro a causa de errores, sesgos y fallos es elevado, más si el modelo se entrena no solo con datos generados por los humanos sino también con datos generados por IA.

Por tanto, los sistemas que se reentrenan siguiendo un proceso de actualización obligan a atender a los bucles o ciclos de retroalimentación que de IA cabe diferenciar entre bucles positivos y bucles negativos correspondiéndose con el primer tipo aquellos que generan resultados precisos que se alinean con las expectativas y preferencias de los usuarios a través de los comentarios positivos que dejan las personas, lo que a su vez refuerza la precisión de los resultados futuros. En otro sentido, los bucles de retroalimentación negativa de IA se presentan cuando los modelos de IA generan resultados inexactos y los usuarios reportan estos fallos a través de un ciclo de retroalimentación que, a cambio, intenta mejorar la estabilidad del sistema solucionando los errores.

Asimismo, en directa relación con la solidez del sistema, se presenta la conveniencia de establecer estrategias para predecir fallos del modelo que afecten negativamente a los derechos fundamentales o a la seguridad de las personas en la utilización de los sistemas de IA. En definitiva, utilizando palabras del RIA se trata de «preservar la solidez como resistencia a fallos, errores o incoherencias técnicas». En otras palabras, garantizar la calidad del modelo desde el principio hasta el final de su ciclo de vida.

La posible degradación puede presentarse mientras el modelo se entrena o cuando se usa para inferencia. De este modo, pueden aparecer problemas como la desviación del modelo (model deviation) que cabe atribuir a las diferencias surgidas entre lo que predice el modelo y la verdad. Para evitar este tipo de degradación se utilizan métricas de medición como la varianza, exactitud, precisión, recall o bias.

En definitiva, la herramienta para mitigar y controlar este tipo de diferencias o variaciones se dirigirá a controlar ese sobre aprendizaje del modelo.

Por otra parte, otro posible escenario se produce con la desviación o deriva del modelo en el tiempo, model (concept) drift, que se da cuando las predicciones del modelo aprendido se degradan debido a alteraciones en el entorno. Por tanto, las capacidades predictivas y la eficiencia disminuyen con el tiempo dado que el entorno es cambiante y sufre variaciones o alteraciones.

Un tercer escenario se produce con la desviación de los datos en el tiempo, data drift o covariate shift, que sucede cuando los datos de entrada de un modelo cambian. Se trata de la principal razón por la que se degrada la exactitud de un modelo en el tiempo (accuracy). El reentrenamiento del modelo puede ser una buena solución para conseguir que se readapte a los cambios producidos y pueda reajustarse en aras de garantizar su solidez.

Como se ha dicho en aquellos casos en que los datos de entrenamiento del modelo sean generados por IA estos problemas que ponen en riesgo la solidez del sistema irán en aumento dada la merma de su calidad y su impacto en los resultados de salida. Tanto es así que este rápido desarrollo de la IA generativa ha supuesto el estudio del fenómeno conocido como colapso del modelo que es un proceso degenerativo que afecta negativamente a modelos aprendidos porque los datos generados contaminan el conjunto de datos de entrenamiento de la próxima generación de modelo. En definitiva, el colapso se produce porque los modelos se entrenan con contenido generado por IA en vez de utilizar contenido generado por humanos derivando hacia una degradación de la calidad del modelo. Sería un bucle de retroalimentación porque los modelos entrenados con datos sintéticos multiplicarán sin cesar los errores, malinterpretarán los datos y las salidas serán incorrectas sin haber tenido en cuenta los hechos menos probables porque se salen de los patrones.

La consecuencia sería la contaminación de datos a gran escala.

El colapso puede producirse en diferentes momentos. En el modelo inicial cuando este empieza a perder información sobre las colas de distribución de los datos de entrenamiento o, por el contrario, en el último modelo cuando entrelaza diferentes modos de las distribuciones originales y converge en una distribución que en poco o nada se parece a la original.

En cuanto a las razones del colapso del modelo podemos establecer dos categorías principales. Por una parte, el error de aproximación estadística que es el error principal y viene causado por el número finito de muestras que, por el contrario, va desapareciendo a medida que el recuento de muestras se acerca al infinito. Por otra, el error de aproximación funcional que se produce cuando determinados modelos, como las redes neuronales, no logran capturar la verdadera función subyacente que debe aprenderse de los datos.

En definitiva, se requiere que los circuitos de retroalimentación de los modelos de IA sean sólidos y, por tanto, de calidad.

Como a nadie se le oculta la precisión resulta afectada por la presencia de sesgos siendo imprescindible que cuando se eligen las métricas se realice un análisis de sesgos que garantice su fiabilidad a partir de su imparcialidad. En este sentido, la presencia de datos sesgados, incompletos o ruidosos, la simplicidad o complejidad

excesivas del algoritmo o los sesgos implícitos o explícitos que portan las personas han de entenderse en relación con el rendimiento del modelo en cuanto errores sistémicos o desviaciones de resultado.

El sesgo admite diferentes tipos porque puede ser el resultado de las suposiciones, preferencias o limitaciones de los datos, el algoritmo o la persona involucrada en el proceso de modelado. El análisis y abordaje de los sesgos no debe hacerse exclusivamente desde una perspectiva técnica siendo necesario intervenir en el elemento humano, social e institucional donde los sesgos están insertados. De modo que cabe establecer tres categorías principales de sesgos de IA que deben ser gestionados:

— Sesgos sistémicos presentes en los conjuntos de datos de IA, normas, prácticas, procesos organizacionales a lo largo del ciclo vital de la IA y, evidentemente, en la sociedad en general que es la que usa estos sistemas.

— Sesgos computacionales y estadísticos presentes en los conjuntos de datos y en los procesos algorítmicos y que derivan de los errores sistémicos debido a que las muestras utilizadas no son representativas.

— Sesgos cognitivos humanos relacionados con la forma en que una persona o grupo percibe la información del sistema de IA que será utilizada para tomar una decisión o completar información que se está buscando; así como con la manera que tenemos de entender los propósitos y funciones de un sistema de IA.

Las métricas de sesgo del modelo pueden aplicarse en la fase de recopilación de datos y en una fase posterior para evaluar los resultados después de entrenar al modelo de modo que permiten detectar si las predicciones incluyen sesgos.

2. EVALUACIÓN DE LA PRECISIÓN Y LA SOLIDEZ PARA GARANTIZAR LA CALIDAD DEL SISTEMA

La precisión y solidez obtenida ha de ser evaluadas a través de su verificación y posterior validación. En otras palabras, tras la confirmación del cumplimiento de los objetivos previstos se ha de proceder a su validación con evidencias objetivas.

El objetivo determinar cómo se comporta el sistema y por qué lo hace de esa manera y poder aplicar esa información para mejorar su rendimiento. Las herramientas para evaluar su comportamiento son variadas destinadas a medir resultados a fin de determinar si se alcanza o no el umbral de confianza deseado. A la hora de medir los resultados los métodos estadísticos permiten establecer si se alcanza el nivel o umbral de confianza deseado.

La validación supone probar el modelo utilizando datos reales para comprobar la verificación de la estabilidad y eficacia del sistema. El modelo se prueba en un conjunto de datos separado que no haya sido usado durante el proceso de entrenamiento para permitir su generalización a datos nuevos e invisibles. La técnica de validación cruzada goza de gran popularidad y consiste en dividir el conjunto de datos en múltiples subconjuntos que son utilizados para entrenar y probar el en las diferentes combinaciones de estos subconjuntos.

Ahora bien, la calidad del modelo requiere algo más que la evaluación de su funcionamiento a través de métricas y técnicas de validación, pues hay que evaluar su interpretabilidad y equidad. En este sentido, nos referimos, por una parte, a la

capacidad para comprender y explicar las predicciones del modelo, esencial para generar confianza en los sistemas de IA y, por otra, a la equidad del modelo que supone que no discrimina a personas o grupos. En consecuencia, tanto la evaluación de la interpretabilidad como la evaluación de la equidad resultan decisivas para la validación del modelo.

La evaluación del modelo en términos de equidad permite determinar cómo afectan los resultados del modelo a determinados grupos sociales definidos en atención a atributos como el sexo, la raza o la edad entre otros, como se distribuyen los valores de predicción y como son los valores de las métricas de rendimiento entre estos grupos. En base a los resultados obtenidos se evalúa como se pueden mitigar o corregir las diferencias, definidos en atención a determinados atributos. Las métricas de equidad permiten evaluar los niveles de rendimiento para determinar la fiabilidad del sistema en tanto en cuanto no ofrezca riesgo de convertirse en fuente de discriminación a través de resultados injustos por la presencia de sesgos sistémicos que perjudicarán a los grupos tradicionalmente subrepresentados en la realidad social. Este resultado será prueba del buen o mal rendimiento de determinados conjuntos de datos que afecta a la robustez y confiabilidad del sistema³⁴.

Ya para cerrar afirmar que la validación de sistemas de IA supone realizar un proceso complejo porque integra la evaluación del funcionamiento del sistema en términos de precisión y solidez, la identificación y mitigación de sesgos en los datos y que se respeta la privacidad y la seguridad de los datos personales. Además, supone también que se han implementado medidas técnicas y organizativas validadas y recogidas documentalmente. Ahora bien, en ningún caso es un proceso que se cierra, sino que debe tener un carácter continuo, dada la obligación de garantizar la precisión, solidez y ciberseguridad del sistema durante todo el ciclo de vida del sistema de IA.

IV. CONCLUSIONES

Los sistemas de IA de alto riesgo nos sitúan ante un escenario en el que la adopción de decisiones puede tener un alto impacto en la salud, la seguridad y los derechos fundamentales, de modo que los resultados que ofrece el sistema deben ser precisos o lo más exactos posible, es decir, carentes de errores, imprecisiones y, por supuesto de sesgos.

No obstante, este objetivo prioritario no es nada fácil en los modelos de inteligencia artificial porque los sistemas pueden ser inexactos o imprecisos por diferentes causas como su continua evolución con el paso del tiempo y los procesos de retroalimentación, lo que afecta directamente y de forma negativa generando alto grado de desconfianza o de falta de fiabilidad.

La fiabilidad del sistema está directamente relacionada con el principio o exigencia de explicabilidad que, si bien es una responsabilidad primordial en la IA y que debe cumplirse por todos los implicados en el ciclo vital del sistema, no es fácil encaje

34. En este punto, nos remitimos al análisis de los datos de calidad (artículo 10) y a la gobernanza de datos en cuanto prácticas adecuadas de gestión y gobernanza de datos para lograr que los conjuntos de datos de entrenamiento, validación y prueba sean de buena calidad.

con el requisito de precisión técnica. El sistema es explicable en atención al nivel de explicación que se proporcione respecto del funcionamiento del sistema y de cómo se realizan los procesos de toma de decisiones, pero también depende de lo comprensibles que resulten dichas explicaciones. En consecuencia, cuanto mayor sea la precisión técnica más complejo será el sistema inteligente, es decir, menos explicable.

Por otra parte, los sistemas fiables son sistemas sólidos, que funcionan de forma precisa y consistente a lo largo del tiempo, con bajos niveles de incertidumbre y, por tanto, con mayor seguridad y confiabilidad. Para garantizar esta solvencia técnica del sistema los responsables deben adoptar cuantas medidas técnicas y organizativas resulten adecuadas para mitigar cualquier riesgo derivado de la quiebra de la solidez y resistencia. La experiencia en inteligencia artificial debe ser aprovechada para diseñar herramientas que se adapten a los diversos escenarios, es decir, medidas específicas que comprendan la complejidad de cada sistema y valore aspectos críticos como el rendimiento, la precisión y la equidad algorítmica.

Las pruebas de robustez y seguridad deberán ser exhaustivas para garantizar que los sistemas sean confiables y seguros en entornos reales y ante peligros reales. Ello sin olvidar que debe realizarse una evaluación para detectar sesgos y así asegurar que los sistemas de IA no estén sesgados y cumplan los requisitos éticos.

Por otra parte, consideramos que sería deseable que los requisitos sean más detallados y específicos para la evaluación de impacto y las pruebas técnicas que son de carácter obligatorio para los sistemas de IA de alto riesgo, a efectos de garantizar la seguridad y la calidad de estos sistemas.

En relación con la obligación de los proveedores de demostrar el cumplimiento de los requisitos establecidos parece conveniente desarrollar herramientas de certificación.

Y, por último, la colaboración internacional debe ir a más para poder prevenir y responder adecuadamente a los retos tecnológicos derivados del constante desarrollo de la IA. Tanto en lo que se refiere a una regulación armonizada global como en relación a compartir buenas prácticas y lecciones aprendidas.

No cabe cerrar este análisis sin una referencia a la garantía de los derechos fundamentales en especial a la privacidad, la no discriminación y la igualdad de oportunidades, lo que nos lleva a afirmar que es posible que en un futuro no muy lejano haya que regular requisitos más específicos y detallados para garantizar la protección de estos derechos.

Ciberseguridad en sistemas de inteligencia artificial de alto riesgo en el artículo 15 del Reglamento

MARCO EMILIO SÁNCHEZ ACEVEDO

Abogado. Profesor Doctor e investigador Universidad Católica de Colombia¹

En este capítulo el lector encontrará un análisis de la obligación de «ciberseguridad» derivada del artículo 15 de la nueva normativa en lo que concierne a los sistemas de Inteligencia Artificial (en adelante IA) catalogados como de alto riesgo. La metodología utilizada corresponde a un análisis e interpretación de documentos a partir de la exploración, comprobación e interpretación. El propósito de este capítulo es determinar cuáles son las obligaciones de ciberseguridad que deben cumplir los sistemas de inteligencia artificial catalogados como de riesgo alto. Para ello, se analizarán tres cuestiones; la primera tiene que ver con resolver la pregunta de si la ciberseguridad es para todos los sistemas de inteligencia artificial o solamente para los de alto riesgo, ello a partir de dos cuestiones de trascendental importancia, por un lado, la evolución tramitación y contenido de la obligación de ciberseguridad, y por otro sobre qué sistemas catalogados como de alto riesgo se exige dicha obligación. A continuación, se aborda el marco europeo de certificación de

-
1. Marco Emilio Sánchez Acevedo es abogado en Colombia y España, docente e investigador de la Universidad Católica de Colombia y colaborar en proyectos de investigación de la Universidad de Valencia — España, es doctorado en Tecnologías y Servicios de la Sociedad de la Información, línea de investigación Derecho Público y Tecnologías; magíster en Ciberseguridad y Ciberdefensa Nacional, y, Especialista en Derecho Administrativo, Constitucional y Gobierno electrónico. El presente capítulo de libro de investigación es resultado del trabajo adelantado por el autor dentro del proyecto de investigación «Análisis de datos para la toma de decisiones espaciales e inteligencia artificial para la administración», vinculado al grupo de investigación Derecho Público y Tecnologías, categoría Colciencias A1, Universidad Católica de Colombia. De igual forma, resultado de la colaboración con el grupo de investigación Régimen jurídico constitucional de las libertades, el gobierno abierto y el uso de las nuevas tecnologías (UVEG-GIUV2016-270), y del Proyecto «Derechos y garantías públicas frente a las decisiones automatizadas y el sesgo y discriminación algorítmicas» 2023-2025 (PID2022-136439OB-I00) financiado por MCIN/AEI/10.13039/501100011033/FEDER, UE, así como del proyecto nacional MICINN, «Derechos y garantías frente a las decisiones automatizadas en entornos de inteligencia artificial, IoT, Big Data y robótica» (PID2019-108710RB-I00, 2020-2022) de la Universidad de Valencia — España.

la ciberseguridad como instrumento para garantizar el cumplimiento de la obligación señalada en el artículo 15 para los sistemas de inteligencia artificial de alto riesgo, se determinan las obligaciones que se deben cumplir de ciberseguridad incorporando las implementadas por administraciones públicas y aquellas que tienen que ver con infraestructuras críticas. Por último, se aborda las evaluaciones de conformidad como instrumento para el cumplimiento de las garantías de ciberseguridad. Para finalizar se dan unas conclusiones.

I. ¿LA CIBERSEGURIDAD ES PARA TODOS LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL? LA OBLIGACIÓN DE CIBERSEGURIDAD DESDE SU PROPUESTA HASTA LA APROBACIÓN FINAL

La ciberseguridad desde hace ya varios años ha sido uno de los elementos centrales en el desarrollo de proyectos de tecnologías de la información y las comunicaciones en términos generales. En 2013, con la Estrategia de Ciberseguridad de la Unión Europea (JOIN/2013) se dio una respuesta política de la Unión a los retos relacionados con la ciberseguridad. El primer acto jurídico en el ámbito de la ciberseguridad de la Unión fue adoptado en 2016, y corresponde a la Directiva (UE) 2016/1148 del Parlamento Europeo y del Consejo. La Directiva (UE) 2016/1148 instauró un marco jurídico mínimo que permitiera mitigar las amenazas que se presentan a las redes y a los sistemas de información sobre todo en la prestación de servicios esenciales, de igual forma en buscar instrumentos que permitieran la continuidad de los servicios cuando se presenten incidentes de seguridad.

Los sistemas de información, la redes y en general las tecnologías hacen parte de la vida central y cotidiana de los ciudadanos, de las empresas, del gobierno². Esto genera una exposición mayor al conjunto de amenazas que se dan en el marco de las nuevas relaciones del ciberespacio, frente a esto se amplían los desafíos, los retos se deben entender a partir de la búsqueda de respuestas ante los riesgos, unos de mayor magnitud, otros de menor magnitud, pero que en cualquiera de los casos plantea la necesidad de enfrentarlos. El uso de sistemas de inteligencia artificial, como se ha venido viendo a lo largo del desarrollo de la nueva normativa analizada a través de este documento, presenta riesgos de distinta naturaleza. Es por ello, que el punto de partida sobre el que se debe analizar las obligaciones de ciberseguridad que se impongan a los sistemas de inteligencia artificial debe ser la Directiva (UE) 2022/2555 del Parlamento Europeo y del Consejo, de 14 de diciembre de 2022, sobre medidas para un alto nivel común de ciberseguridad en toda la Unión, por la que se modifica el Reglamento (UE) – 910/2014 y la Directiva (UE) 2018/1972 y por la que se deroga la Directiva (UE) 2016/1148 (Directiva NIS 2).

La Directiva (UE) 2022/2555 del Parlamento Europeo y del Consejo, de 14 de diciembre de 2022, sobre medidas para un alto nivel común de ciberseguridad en toda la Unión, establece un conjunto de medidas que buscan un alto nivel común de ciberseguridad en toda la unión y por ello incorpora un conjunto de obligaciones en cuanto a la necesidad de adoptar estrategias de ciberseguridad, designar autoridades

2. Sobre este tema consultar, Sánchez Acevedo, Marco Emilio y otros, *El derecho y las tecnologías de la información y la comunicación (TIC)*, Universidad Católica de Colombia, Bogotá, 2015.

competentes, designar autoridades de gestión de crisis, designar puntos de contacto únicos y equipos de respuestas a incidentes de seguridad informática; de igual forma, establece las medidas para gestionar los riesgos de ciberseguridad y las obligaciones de información, obligaciones de intercambio de información sobre ciberseguridad y obligaciones de supervisión y ejecución. La norma plantea unos sectores de alta crítica dentro de esos sectores está la energía que incorporan los subsectores de electricidad, calefacción refrigeración, aceite, gas, hidrógeno; El sector del transporte dentro del que se integran el transporte aéreo por carril por agua; también el sector bancario el de las infraestructuras del mercado financiero, salud, agua potable, infraestructura digital, agua residuales, infraestructura digital gestión de servicios tic empresa y administración pública y espacio.

Mitigar eficazmente los riesgos para la salud, la seguridad y los derechos fundamentales, requiere la imposición de un conjunto de requisitos se deben aplicar a los sistemas de inteligencia artificial y que se encuentran vinculados a la calidad del conjunto de los datos utilizados, a la gestión de documentación técnica, al mantenimiento de los registros, la entrega de información y transparencia, la solidez, precisión, supervisión humana y por supuesto a la ciberseguridad. La norma aprobada, si bien, tiene como ámbito de aplicación los sistemas de IA de alto riesgo, también es verdad, que las obligaciones de ciberseguridad no son exclusivas para estos sistemas, sino para todas las tecnologías de la información y las comunicaciones de manera proporcional a los fines e intereses perseguidos. El Reglamento sigue un enfoque basado en los riesgos que generan el uso de sistemas de IA i) un riesgo inaceptable, ii) un riesgo alto, y iii) un riesgo bajo o mínimo, para el caso que nos ocupa en la presente investigación, como ya se ha dicho, el objeto es solo los sistemas de IA de riesgo alto. Ello no implica que a pesar de que un sistema de información de IA no se encuentre en dicha categoría, está obligado al cumplimiento de normas de ciberseguridad.

La ciberseguridad es el instrumento que garantizará que los sistemas de IA hagan frente a los distintos tipos de ataques que se pueden presentar y que buscarán la explotación de vulnerabilidades. Para garantizar un nivel de ciberseguridad adecuado a los riesgos, se prevé que los proveedores de sistemas de IA de alto riesgo deben adoptar medidas adecuadas, teniendo en cuenta también, según proceda, la infraestructura de TIC subyacente. A los efectos del contenido del este capítulo se tendrá por Ciberseguridad «todas las actividades necesarias para la protección de las redes y sistemas de información, de los usuarios de tales sistemas y de otras personas afectadas por las ciberamenazas» siguiendo lo señalado en el artículo 2, punto 1, del Reglamento (UE) 2019/881³.

3. De acuerdo con los trabajos de la OCDE (véase por ejemplo, *Recommendation of the Council on Digital Security Risk Management for Economic and Social Prosperity in Digital Security Risk Management for Economic and Social Prosperity, OCDE Recommendation and Companion Document*, OCDE Publishing, Paris, 2015), la «ciberseguridad» puede plantearse a través de dimensiones 1) la tecnología, cuando se centra en el funcionamiento del entorno digital (a menudo llamado «seguridad de la información», «seguridad informática» o «seguridad de la red» por los expertos); 2) la aplicación de la ley o aspectos legales (por ejemplo, cibercrimen); 3) la seguridad nacional, la estabilidad internacional, incluidos aspectos como el papel de las TIC respecto de la inteligencia, la prevención de conflictos, la guerra, la ciberdefensa,

1. EVOLUCIÓN, TRAMITACIÓN Y CONTENIDO FINAL DE LA CIBERSEGURIDAD COMO REQUISITO EN LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL EN EL MARCO DE LA PROPUESTA ADOPTADA

La ciberseguridad ha estado presente en todo el proceso de tramitación del proyecto normativo⁴, tanto la propuesta inicial, que hace referencia a la necesidad de que los sistemas de IA garanticen la ciberseguridad, como en el trámite posterior se ha evidenciado el acuerdo en incorporar obligaciones de ciberseguridad.

La propuesta inicialmente presentada, en el considerando 43 inicia justificando que deben aplicarse a los sistemas de IA de alto riesgo requisitos, entre otros, a la ciberseguridad. En línea con ello, el considerando 49 incorporaba que los sistemas de IA de alto riesgo deben, entre otros, presentar un nivel adecuado de ciberseguridad con arreglo al estado de la técnica generalmente reconocido. Por otra parte, en el considerando 51 justificaba desde la afirmación que *«la ciberseguridad es fundamental para garantizar que los sistemas de IA resistan a las actuaciones de terceros maliciosos que, aprovechando las vulnerabilidades del sistema, traten de alterar su uso, conducta o funcionamiento o de poner en peligro sus propiedades de seguridad»*.

Ya en la propuesta normativa, en particular, el artículo 13 se presentaba una obligación de ciberseguridad vinculada a la transparencia y comunicación de información a los usuarios, en la medida que los sistemas de alto riesgo irán acompañados de instrucciones en el que se especificará el nivel de ciberseguridad. En línea con lo anterior, el artículo 15 de la propuesta presentaba que los sistemas de IA de alto riesgo se diseñarían y desarrollarían de modo que, en vista de su finalidad prevista, alcancen un nivel adecuado de, entre otros, ciberseguridad. Con ello vinculaba la ciberseguridad a la finalidad prevista y en su virtud se plantea un nivel adecuado a esa finalidad. En la misma línea, planteaba que las soluciones técnicas encaminadas a garantizar la ciberseguridad de los sistemas de IA de alto riesgo deben ser adecuadas a las circunstancias y los riesgos pertinentes.

Una de las cuestiones mas relevantes de la propuesta inicial es la presunción de conformidad con determinanos requisitos vinculados a la ciberseguridad en los sistemas de IA de alto riesgo que hayan sido certificados o para los que se haya expedido una delaración de conformidad con arreglo al esquema de de ciberseguridad en virtud del Reglamento (UE) 2019/881 del Parlamento

etc., y 4) la dimensión económica y social, que abarca la creación de riqueza, la innovación, el crecimiento, la competitividad y el empleo en todos los sectores económicos, las libertades individuales, la salud, la educación, la cultura, la participación democrática, la ciencia, el ocio y otras dimensiones del bienestar en que el entorno digital impulsa el progreso.

4. Véase, la Ley de Inteligencia Artificial (P9_TA(2023)0236), Enmiendas aprobadas por el Parlamento Europeo el 14 de junio de 2023 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican los actos legislativos de la Unión COM/2021/206, C9-0146/2021 y 2021/0106(COD) Recuperado de https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_ES.html

Europeo y del Consejo⁵, en la medida en que el certificado de ciberseguridad o la declaración de conformidad, o partes de estos, prevean estos requisitos.

En las enmiendas aprobadas por el Parlamento Europeo el 14 de junio de 2023 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo, en las que se modifican un par de actos legislativos de la Unión (COM/2021/0206 — C9-0146/2021 — 2021/0106(COD) y se incorporan algunos elementos puntuales, vinculados a las obligaciones de ciberseguridad y que fortalecen la propuesta inicial, no generando cambios sustanciales, sin perjuicio que se deben resaltar las siguientes:

En la enmienda 17 se incorpora un considerando 5 bis (nuevo), en el que se reconoce, entre otros, «(...) las preocupaciones en materia de ciberseguridad (...)».

En la enmienda 63, considerando (33 bis) se incorpora, entre otros, una motivación para justificar que «(...) No debe considerarse que los sistemas biométricos y basados en la biometría previstos en el Derecho de la Unión para posibilitar la ciberseguridad y las medidas de protección de los datos personales supongan un riesgo significativo de perjuicio para la salud, la seguridad y los derechos fundamentales».

En la enmienda 64 se le adiciona a la propuesta inicial (34) «(...) En el caso de la gestión y el funcionamiento de infraestructuras críticas, conviene considerar de alto riesgo a los sistemas de IA destinados a ser componentes de seguridad en la gestión y el funcionamiento del suministro de agua, gas, calefacción, electricidad e infraestructuras digitales críticas, pues su fallo o defecto de funcionamiento puede vulnerar la seguridad y la integridad de dichas infraestructuras críticas o poner en peligro la vida y la salud de las personas a gran escala y alterar de manera apreciable el desarrollo habitual de las actividades sociales y económicas. Los componentes de seguridad de las infraestructuras críticas, también de las infraestructuras digitales críticas, son sistemas utilizados para proteger directamente la integridad física de la infraestructura crítica o la salud y la seguridad de las personas y los bienes. Un fallo o un defecto de funcionamiento de esos componentes podría dar lugar directamente a riesgos para la integridad física de las infraestructuras críticas y, por tanto, a riesgos para la salud y la seguridad de las personas y los bienes. Los componentes destinados a ser utilizados exclusivamente con fines de ciberseguridad no deben considerarse componentes de seguridad. Entre dichos componentes de seguridad cabe señalar, por ejemplo, los sistemas de control de la presión del agua o los sistemas de control de las alarmas contraincendios en los centros de computación en nube».

En la enmienda 77 se le adiciona (43) que «Deben aplicarse a los sistemas de IA de alto riesgo requisitos referentes a la calidad y la pertinencia de los conjuntos de datos utilizados, la documentación técnica y el registro, la transparencia y la comunicación de información a los implementadores, la vigilancia humana, la solidez, la precisión y la ciberseguridad. Dichos requisitos son necesarios para

5. Reglamento (UE) 2019/881 del Parlamento Europeo y del Consejo, de 17 de abril de 2019, relativo a ENISA (Agencia de la Unión Europea para la Ciberseguridad) y a la certificación de la ciberseguridad de las tecnologías de la información y la comunicación y por el que se deroga el Reglamento (UE) N.º 526/2013 (Reglamento sobre la Ciberseguridad) (DO L 151 de 7.6.2019, p. 1).

mitigar de forma efectiva los riesgos para la salud, la seguridad y los derechos fundamentales, así como para el medio ambiente, la democracia y el Estado de Derecho, según corresponda en función de la finalidad prevista o el uso indebido razonablemente previsible del sistema, y no se dispone razonablemente de otras medidas menos restrictivas del comercio, con lo que se evitan restricciones injustificadas de este».

En la enmienda 85 se le adiciona (49) «Los sistemas de IA de alto riesgo deben funcionar de manera consistente durante todo su ciclo de vida y presentar un nivel adecuado de precisión, solidez y ciberseguridad con arreglo al estado de la técnica generalmente reconocido. (...)».

En la enmienda 87 se le adiciona (51) «La ciberseguridad es fundamental para garantizar que los sistemas de IA resistan a las actuaciones de terceros maliciosos que, aprovechando las vulnerabilidades del sistema, traten de alterar su uso, conducta o funcionamiento o de poner en peligro sus propiedades de seguridad. Los ciberataques contra sistemas de IA pueden dirigirse contra elementos específicos de la IA, como los conjuntos de datos de entrenamiento (p. ej., contaminación de datos) o los modelos entrenados (p. ej., ataques adversarios o contra la confidencialidad), o aprovechar las vulnerabilidades de los elementos digitales del sistema de IA o la infraestructura de TIC subyacente. Por lo tanto, para asegurar un nivel de ciberseguridad adecuado a los riesgos, los proveedores de sistemas de IA de alto riesgo (...)».

Enmienda 101 (60 octies) se le adiciona «A la vista de la naturaleza y la complejidad de la cadena de valor para los sistemas de IA, resulta esencial aclarar el papel de los agentes que contribuyen al desarrollo de dichos sistemas. (...) concretamente, los modelos fundacionales deben evaluar y mitigar los posibles riesgos y perjuicios mediante un diseño, unas pruebas y un análisis adecuados, aplicar medidas de gobernanza de datos —en particular, una evaluación de los sesgos— y cumplir requisitos de diseño técnico que garanticen niveles adecuados de rendimiento, previsibilidad, interpretabilidad, corregibilidad, seguridad y ciberseguridad, así como cumplir las normas medioambientales. (...)».

En la enmienda 110 se le adiciona (65) «En virtud del presente Reglamento, las autoridades nacionales competentes deben designar a los organismos notificados que realizarán las evaluaciones externas de la conformidad (...) en lo que respecta a su independencia, sus competencias y la ausencia de conflictos de intereses, así como a requisitos mínimos de ciberseguridad. (...)».

En la enmienda 115 se le adiciona (69) «(...) la Comisión debe tener en cuenta los riesgos de ciberseguridad y los riesgos vinculados a peligros. Con el fin de maximizar la disponibilidad y el uso de la base de datos por parte del público, la base de datos y la información que se pone a disposición a través de ella deben cumplir los requisitos establecidos en la Directiva 2019/882».

En la enmienda 306 se le adiciona (...) «ii) El nivel de precisión, solidez y ciberseguridad mencionado en el artículo 15 con respecto al cual se haya probado y validado el sistema de IA de alto riesgo y que puede esperarse de este, así como las circunstancias claramente conocidas o previsibles que podrían afectar al nivel de precisión, solidez y ciberseguridad esperado (...)».

En la enmienda 321 se le adiciona «(...) 1. Los sistemas de IA de alto riesgo se diseñarán y desarrollarán siguiendo el principio de seguridad desde el diseño y por defecto. En vista de su finalidad prevista, deben alcanzar un nivel adecuado de precisión, solidez, seguridad y ciberseguridad y funcionar de manera consistente en esos sentidos durante todo su ciclo de vida (...)».

En la enmienda 323 se adiciona al Artículo 15 — apartado 1 ter (nuevo) 1 ter. «A fin de abordar cualquier problema emergente en el mercado interior en relación con la ciberseguridad, la Agencia de la Unión Europea para la Ciberseguridad (ENISA) colaborará con el Comité Europeo de Inteligencia Artificial tal como se establece en el artículo 56, apartado 2, letra b)».

En la enmienda 399 se propone adicionar al artículo 28 ter (nuevo) en cuanto a las Obligaciones del proveedor de un modelo fundacional «(...) 1. Antes de comercializarlo o ponerlo en servicio (...) 2. A efectos de lo dispuesto en el apartado 1, el proveedor de un modelo fundacional: a) demostrará, mediante un diseño, ensayo y análisis adecuados, la detección (...) c) diseñará y desarrollará el modelo fundacional con el fin de alcanzar a lo largo de su ciclo de vida niveles adecuados de rendimiento, previsibilidad, interpretabilidad, corrección, seguridad y ciberseguridad evaluados mediante métodos adecuados, como la evaluación del modelo con la participación de expertos independientes, análisis documentados y pruebas exhaustivas durante la conceptualización, el diseño y el desarrollo; (...)».

En la enmienda 401 se pretende adicionar al artículo 29 — apartado 1 bis (nuevo) 1 bis. «En la medida en que los implementadores ejerzan control sobre el sistema de IA de alto riesgo, ellos: I(...); iii) garantizarán que las medidas de solidez y ciberseguridad pertinentes y adecuadas sean objeto de un seguimiento periódico de la eficacia y se ajusten o actualicen periódicamente».

En la enmienda 423 se pretende adicionar al artículo 33 — apartado 2 «Los organismos notificados satisfarán los requisitos organizativos, así como los de gestión de la calidad, recursos y procesos, necesarios para el desempeño de sus funciones, así como los requisitos mínimos de ciberseguridad establecidos para las entidades de la administración pública identificadas como operadores de servicios esenciales» de conformidad con la Directiva (UE) 2022/2555.

En la enmienda 505 se propone un artículo 53 bis (nuevo) «Modalidades y funcionamiento de los espacios controlados de pruebas para la IA 1(...) 2. La Comisión está facultada para adoptar actos delegados de conformidad con el procedimiento a que se refiere el artículo 73, a más tardar doce meses después de la entrada en vigor del presente Reglamento, y garantizará que: (...) h) los espacios controlados de pruebas facilitarán el desarrollo de herramientas e infraestructuras para la prueba, la evaluación comparativa, la evaluación y la explicación de las dimensiones de los sistemas de IA pertinentes para los espacios controlados de pruebas, como la precisión, la solidez y la ciberseguridad, así como la reducción al mínimo de los riesgos para los derechos fundamentales, el medio ambiente y la sociedad en su conjunto. 3. (...)».

En la enmienda 532 se propone un nuevo artículo 57 bis «Composición del consejo de administración 1. El consejo de administración estará compuesto por los siguientes miembros: (...) d) un representante de la Agencia de la Unión

Europea para la Ciberseguridad (ENISA); (...) Cada representante de una autoridad nacional de supervisión dispondrá de un voto. Los representantes de la Comisión, el SEPD, la ENISA y la FRA no tendrán derecho de voto. Cada miembro tendrá un suplente. El nombramiento de los miembros y suplentes del consejo de administración tendrá en cuenta la necesidad de equilibrio entre mujeres y hombres. Los miembros del consejo de administración y sus suplentes se harán públicos. (...)».

En la enmienda 557 se le adiciona a la propuesta inicial «(...) 4. Los Estados miembros garantizarán que la autoridad de supervisión disponga de recursos técnicos, financieros y humanos adecuados, así como de las infraestructuras para el desempeño eficaz de sus funciones con arreglo al presente Reglamento. En concreto, la autoridad nacional de supervisión dispondrá permanentemente de suficiente personal cuyas competencias y conocimientos técnicos incluirán un conocimiento profundo de las tecnologías de inteligencia artificial, datos y computación de datos, la protección de datos personales, la ciberseguridad, el Derecho en materia de competencia, los riesgos para los derechos fundamentales, la salud y la seguridad, y conocimientos acerca de las normas y requisitos legales vigentes. (...)».

En la enmienda 559 corresponde al artículo 59 — apartado 4 ter (nuevo) 4 ter. «Las autoridades nacionales de supervisión satisfarán los requisitos mínimos de ciberseguridad establecidos para las entidades de administración pública identificadas como operadores de servicios esenciales con arreglo a la Directiva (UE) 2022/2555».

En la enmienda 640 se presenta una propuesta de adicionar el artículo 70 — apartado 1 bis (nuevo) 1 bis. «Las autoridades involucradas en la aplicación del presente Reglamento de conformidad con el apartado 1 minimizarán la cantidad de datos solicitados para su divulgación a los datos estrictamente necesarios para la percepción del riesgo y la evaluación de dicho riesgo. Suprimirán los datos tan pronto como dejen de ser necesarios para el fin para el que se solicitaron. Establecerán medidas adecuadas y efectivas en materia de ciberseguridad, técnica y organización a fin de proteger la seguridad y la confidencialidad de la información y los datos obtenidos en el desempeño de sus funciones y actividades».

En la enmienda 755 que corresponde al Anexo IV — párrafo 1 — punto 2 — letra g se adiciona «los procedimientos de validación y prueba utilizados, incluida la información acerca de los datos de validación y prueba empleados y sus características principales; los parámetros utilizados para medir la precisión, la solidez y el cumplimiento de otros requisitos pertinentes dispuestos en el título III, capítulo 2, así como los efectos potencialmente discriminatorios; los archivos de registro de las pruebas y todos los informes de las pruebas fechados y firmados por las personas responsables, en particular en lo que respecta a los cambios predeterminados a que se refiere la letra f)».

Así las cosas, se deja en claro que la ciberseguridad, en esta fase de la tramitación, ocupó varios de los debates y se convierte en uno de los elementos esenciales a regular en la propuesta normativa.

Finalmente, la norma ya aprobada establece un conjunto de obligaciones de manera integral en materia de ciberseguridad, entre otras cuestiones, se destaca; i) no considerar sistemas de alto riesgo los sistemas biométricos destinados a ser utilizados únicamente con el fin de habilitar medidas de ciberseguridad y protección de datos personales (Considerando 54); ii) clasificar como de alto riesgo los sistemas de IA destinados a ser utilizados como componentes de seguridad en la gestión y el funcionamiento de las infraestructuras digitales críticas enumeradas en el anexo I, punto 8, de la Directiva (UE) 2022/2557, sin embargo, los componentes destinados a ser utilizados únicamente con fines de ciberseguridad no deben considerarse componentes de seguridad (Considerando 55); iii) deben aplicarse requisitos a los sistemas de IA de alto riesgo en lo que respecta, entre otros, a la gestión de riesgos y ciberseguridad (Considerando 66); los sistemas de IA de alto riesgo deben funcionar de manera uniforme a lo largo de su ciclo de vida y alcanzar un nivel adecuado de, entre otros, ciberseguridad, a la luz de su finalidad prevista y de conformidad con el estado de la técnica generalmente reconocido (74); La ciberseguridad es fundamental para garantizar que los sistemas de IA resistan a las actuaciones de terceros maliciosos que, aprovechando las vulnerabilidades del sistema, traten de alterar su uso, comportamiento o funcionamiento o de poner en peligro sus propiedades de seguridad, y para garantizar un nivel de ciberseguridad adecuado a los riesgos, los proveedores de sistemas de IA de alto riesgo deben adoptar medidas adecuadas, como los controles de seguridad, teniendo también en cuenta, cuando proceda, la infraestructura de TIC subyacente (considerando 76); los sistemas de IA de alto riesgo que se encuentren ámbito de aplicación del Reglamento 2022/0272, de conformidad con el artículo 8 del Reglamento 2022/0272⁶, podrán demostrar el cumplimiento del requisito de ciberseguridad mediante el cumplimiento de los requisitos esenciales de ciberseguridad establecidos en el artículo 10 y en el anexo I del Reglamento 2022/0272 y deben considerarse conformes con los requisitos de ciberseguridad establecidos si se demuestra la declaración UE de conformidad o en partes de la misma expedida con arreglo al Reglamento 2022/0272 (considerando 77); los proveedores de modelos de IA de uso general con riesgos sistémicos deben evaluar y mitigar los posibles riesgos sistémicos (114 y 115); los sistemas de IA de alto riesgo que hayan sido certificados o para los que se haya expedido una declaración de conformidad en el marco de un régimen de ciberseguridad contenido en el Reglamento (UE) 2019/881 del Parlamento Europeo y del Consejo, cumplen el requisito de ciberseguridad del presente Reglamento en la medida en que el certificado de ciberseguridad o la declaración de conformidad o partes de los mismos cubran el requisito de ciberseguridad.

En materia de ciberseguridad, la norma aprobada incorpora un conjunto de obligaciones directas en el articulado, todas estas vinculadas a los sistemas

6. Comité Económico y Social Europeo, Dictamen del Comité Económico y Social Europeo sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo relativo a los requisitos horizontales de la ciberseguridad para los productos con elementos digitales y por el que se modifica el Reglamento (UE) 2019/1020. (Ponente, Mensi, Maurizio), Recuperado en <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52022AE4103>

de alto riesgo, que deben atenderse de forma paralela a las obligaciones de transparencia a los responsables del despliegue (artículo 13) y precisión, solidez y ciberseguridad (artículo 15).

La primera obligación está vinculada a que «(...) 2. los sistemas de IA de alto riesgo irán acompañados de instrucciones de uso en un formato digital adecuado o de otro tipo que incluya información concisa, completa, correcta y clara que sea pertinente, accesible y comprensible para que los responsables del despliegue (...) las instrucciones de uso contendrán como mínimo (...) (ii) el nivel de precisión, incluidas sus métricas, robustez y ciberseguridad a que se refiere el artículo 15, con respecto al cual se ha probado y validado el sistema de IA de alto riesgo y que cabe esperar, así como cualquier circunstancia conocida y previsible que pueda repercutir en ese nivel previsto de precisión, robustez y ciberseguridad». De igual forma, «(...) Los sistemas de IA de alto riesgo se diseñarán y desarrollarán de forma que alcancen un nivel adecuado de precisión, solidez y ciberseguridad y funcionen de manera uniforme en esos sentidos durante todo su ciclo de vida. (...) Los sistemas de IA de alto riesgo serán resistentes a los intentos de terceros no autorizados de alterar su uso, su información de salida o su funcionamiento aprovechando las vulnerabilidades del sistema. Las soluciones técnicas encaminadas a garantizar la ciberseguridad de los sistemas de IA de alto riesgo serán adecuadas a las circunstancias y los riesgos pertinentes. Entre las soluciones técnicas destinadas a subsanar vulnerabilidades específicas de la IA figurarán, según corresponda, medidas para prevenir, detectar, combatir, resolver y controlar los ataques que traten de manipular el conjunto de datos de entrenamiento (“envenenamiento de datos”), o los componentes preentrenados utilizados en el entrenamiento (“envenenamiento de modelos”), la información de entrada diseñada para hacer que el modelo cometa un error (“ejemplos adversarios” o “evasión de modelos”), los ataques a la confidencialidad o los defectos en el modelo (...)».

En cuanto a los requisitos relativos a los organismos notificados, se establece el deber de estos organismos de cumplir requisitos de ciberseguridad adecuados (artículo 31).

De igual manera, el artículo 42 «Presunción de conformidad con determinados requisitos» en el numeral 2 (...) «Se presumirá que los sistemas de IA de alto riesgo que cuenten con un certificado o una declaración de conformidad en virtud de un esquema de ciberseguridad con arreglo al Reglamento (UE) 2019/88 cuyas referencias estén publicadas en el Diario Oficial de la Unión Europea cumplen los requisitos de ciberseguridad establecidos en el artículo 15 del presente Reglamento en la medida en que el certificado de ciberseguridad o la declaración de conformidad, o partes de estos, abarquen dichos requisitos.».

En cuanto a las obligaciones de los proveedores de modelos de IA de uso general con riesgo sistémico los proveedores de estos modelos velarán por que se establezca un nivel adecuado de protección de la ciberseguridad para el modelo de IA de uso general con riesgo sistémico y la infraestructura física del modelo.

En lo que respecta a los actos de ejecución para evitar que se produzca una fragmentación en la Unión, se establece la obligación para que la Comisión adopte actos de ejecución que especifiquen las disposiciones detalladas para el establecimiento, el desarrollo, la puesta en práctica, el funcionamiento y la supervisión de los espacios controlados de pruebas para la IA. En línea

con ello, los actos de ejecución mencionados en el apartado 1 del artículo 58 garantizarán los espacios controlados de pruebas para la IA faciliten el desarrollo de herramientas e infraestructuras para la prueba, la evaluación comparativa, la evaluación y la explicación de las dimensiones de los sistemas de IA pertinentes para el aprendizaje regulatorio, como la precisión, la solidez y la ciberseguridad, así como de medidas para reducir los riesgos para los derechos fundamentales y la sociedad en su conjunto.

Con la creación del Consejo Europeo de Inteligencia Artificial (regulado en el artículo 65), se le señalan obligaciones de cooperación, entre otros, en el ámbito de la ciberseguridad (artículo 66 literal h). En la misma línea se crea un foro consultivo (artículo 67) para proporcionar conocimientos técnicos y asesorar al Comité y a la Comisión. La Agencia de los Derechos Fundamentales de la Unión Europea, la Agencia de la Unión Europea para la Ciberseguridad, el Comité Europeo de Normalización (CEN), el Comité Europeo de Normalización Electrotécnica (Cenelec) y el Instituto Europeo de Normas de Telecomunicaciones (ETSI) serán miembros permanentes del foro consultivo (artículo 67, 5).

Por último, en lo que respecta a la documentación técnica incorporada en el anexo IV, señala que «La documentación técnica a que se refiere el artículo 11, apartado 1, incluirá como mínimo la siguiente información, aplicable al sistema de IA pertinente: (...) (h) h) las medidas de ciberseguridad adoptadas».

Así las cosas, la norma contiene un conjunto de obligaciones directas y otras que nos remite a la aplicación de obligaciones contenidas en otras normas, eso sí, siempre vinculadas a los sistemas de IA de alto riesgo.

2. LA CIBERSEGURIDAD EN SISTEMAS DE INTELIGENCIA ARTIFICIAL CATALOGADOS ALTO RIESGO

Las obligaciones de ciberseguridad están vinculadas a los sistemas de IA de alto riesgo contenidos en el anexo II y III del nuevo reglamento⁷. En idéntico sentido se considerará de alto riesgo que el producto del que el sistema de IA sea componente de seguridad con arreglo a la letra a), o el propio sistema de IA como producto, deba someterse a una evaluación de la conformidad realizada por un organismo independiente para su introducción en el mercado o puesta en servicio con arreglo a los actos legislativos de armonización de la Unión enumerados en el anexo I. y por ende con obligaciones de ciberseguridad de las contenidas en el artículo 15: la seguridad en juguetes⁸, embarcaciones de recreo y a las motos acuáticas⁹, ascensores y componentes de seguridad para ascensores¹⁰, aparatos y sistemas de protección para uso en atmósferas

7. Sobre este particular véase el apartado de sistemas de alto riesgo ya desarrollado.

8. Directiva 2009/48/CE del Parlamento Europeo y del Consejo, sobre la seguridad de los juguetes, 18 de junio de 2009, p. 1 (DO L 170 de 30.6.2009).

9. Directiva 2013/53/UE del Parlamento Europeo y del Consejo, relativa a las embarcaciones de recreo y a las motos acuáticas y por la que se deroga la Directiva 94/25/CE, 20 de noviembre de 2013, p. 90 (DO L 354 de 28.12.2013).

10. Directiva 2014/33/UE del Parlamento Europeo y del Consejo, sobre la armonización de las legislaciones de los Estados miembros relativas a los ascensores y componentes de seguridad para ascensores, 26 de febrero de 2014, p. 251 (DO L 96 de 29.3.2014).

potencialmente explosivas¹¹, comercialización de equipos radioeléctricos¹², comercialización de equipos a presión¹³, instalaciones de transporte por cable¹⁴, equipos de protección individual¹⁵, aparatos de gas¹⁶, productos sanitarios¹⁷, productos sanitarios para diagnóstico in vitro¹⁸, seguridad de la aviación civil¹⁹, homologación y la vigilancia del mercado de los vehículos de dos o tres ruedas y los cuatriciclos²⁰, la homologación y la vigilancia del mercado de los vehículos agrícolas y forestales²¹, equipos marinos²², interoperabilidad del sistema ferroviario en la Unión Europea²³, homologación y la vigilancia del mercado de los vehículos de motor y sus remolques, y de los sistemas, componentes y

11. Directiva 2014/34/UE del Parlamento Europeo y del Consejo, sobre la armonización de las legislaciones de los Estados miembros relativas a los aparatos y sistemas de protección para uso en atmósferas potencialmente explosivas, 26 de febrero de 2014, p. 309 (DO L 96 de 29.3.2014).
12. Directiva 2014/53/UE del Parlamento Europeo y del Consejo, sobre la armonización de las legislaciones de los Estados miembros relativas a la comercialización de equipos radioeléctricos y por la que se deroga la Directiva 1999/5/CE, 16 de abril de 2014, p. 62 (DO L 153 de 22.5.2014).
13. Directiva 2014/68/UE del Parlamento Europeo y del Consejo, relativa a la armonización de las legislaciones de los Estados miembros sobre la comercialización de equipos a presión, 15 de mayo de 2014, p. 164 (DO L 189 de 27.6.2014).
14. Reglamento (UE) 2016/424 del Parlamento Europeo y del Consejo, relativo a las instalaciones de transporte por cable y por el que se deroga la Directiva 2000/9/CE, 9 de marzo de 2016, p. 1 (DO L 81 de 31.3.2016).
15. Reglamento (UE) 2016/425 del Parlamento Europeo y del Consejo, sobre equipos de protección individual y por el que se deroga la Directiva 89/686/CEE del Consejo, 9 de marzo de 2016, p. 51 (DO L 81 de 31.3.2016).
16. Reglamento (UE) 2016/426 del Parlamento Europeo y del Consejo, de 9 de marzo de 2016, sobre los aparatos de gas y por el que se deroga la Directiva 2009/142/CE (DO L 81 de 31.3.2016, p. 99).
17. Reglamento (UE) 2017/745 del Parlamento Europeo y del Consejo, de 5 de abril de 2017, sobre productos sanitarios, por el que se modifican la Directiva 2001/83/CE, el Reglamento (CE) N.º 178/2002 y el Reglamento (CE) N.º 1223/2009 y se derogan las Directivas 90/385/CEE y 93/42/CEE del Consejo (DO L 117 de 5.5.2017, p. 1).
18. Reglamento (UE) 2017/746 del Parlamento Europeo y del Consejo, de 5 de abril de 2017, sobre productos sanitarios para diagnóstico in vitro y por el que se derogan la Directiva 98/79/CE y la Decisión 2010/227/UE de la Comisión (DO L 117 de 5.5.2017, p. 176).
19. Reglamento (CE) N.º 300/2008 del Parlamento Europeo y del Consejo, de 11 de marzo de 2008, sobre normas comunes para la seguridad de la aviación civil y por el que se deroga el Reglamento (CE) N.º 2320/2002 (DO L 97 de 9.4.2008, p. 72).
20. Reglamento (UE) N.º 168/2013 del Parlamento Europeo y del Consejo, de 15 de enero de 2013, sobre la homologación y la vigilancia del mercado de los vehículos de dos o tres ruedas y los cuatriciclos (DO L 60 de 2.3.2013, p. 52).
21. Reglamento (UE) N.º 167/2013 del Parlamento Europeo y del Consejo, de 5 de febrero de 2013, sobre la homologación y la vigilancia del mercado de los vehículos agrícolas y forestales (DO L 60 de 2.3.2013, p. 1).
22. Directiva 2014/90/UE del Parlamento Europeo y del Consejo, de 23 de julio de 2014, sobre equipos marinos y por la que se deroga la Directiva 96/98/CE del Consejo (DO L 257 de 28.8.2014, p. 146).
23. Directiva (UE) 2016/797 del Parlamento Europeo y del Consejo, de 11 de mayo de 2016, sobre la interoperabilidad del sistema ferroviario en la Unión Europea (DO L 138 de 26.5.2016, p. 44).

unidades técnicas independientes destinados a dichos vehículos²⁴, homologación de tipo de los vehículos de motor y sus remolques²⁵, normas comunes en el ámbito de la aviación civil²⁶.

Si bien se incorporan obligaciones en materia de ciberseguridad, también es cierto que la norma direcciona la ciberseguridad a la aplicación de las normas existentes y al nivel adecuado a la finalidad prevista y a las circunstancias. En consecuencia, se deben aplicar, entre otras, las normas de ciberseguridad dispuestas por la Unión Europea y en particular la Directiva (UE) 2022/2555 del Parlamento Europeo y del Consejo, de 14 de diciembre de 2022, sobre medidas para un alto nivel común de ciberseguridad en toda la Unión²⁷, por la que se modifica el Reglamento (UE) 910/2014 y la Directiva (UE) 2018/1972, y por la que se deroga la Directiva (UE) 2016/1148 (Directiva NIS 2), y el Reglamento (UE) 2019/881 del Parlamento Europeo y del Consejo, de 17 de abril de 2019, relativo a ENISA (Agencia de la Unión Europea para la Ciberseguridad) y a la certificación de la ciberseguridad de las tecnologías de la información y la comunicación y por el que se deroga el Reglamento (UE) 526/2013 («Reglamento sobre la Ciberseguridad»), Reglamento 2022/0272²⁸, entre otros.

24. Reglamento (UE) 2018/858 del Parlamento Europeo y del Consejo, de 30 de mayo de 2018, sobre la homologación y la vigilancia del mercado de los vehículos de motor y sus remolques, y de los sistemas, componentes y unidades técnicas independientes destinados a dichos vehículos, por el que se modifican los Reglamentos (CE) N.º 715/2007 y (CE) N.º 595/2009 y se deroga la Directiva 2007/46/CE (DO L 151 de 14.6.2018, p. 1).
25. Reglamento (UE) 2019/2144 del Parlamento Europeo y del Consejo, de 27 de noviembre de 2019, relativo a los requisitos de homologación de tipo de los vehículos de motor y sus remolques, así como de los sistemas, componentes y unidades técnicas independientes destinados a dichos vehículos, en lo que respecta a su seguridad general y a la protección de sus ocupantes y de los usuarios vulnerables de la vía pública, por el que se modifica el Reglamento (UE) 2018/858 del Parlamento Europeo y del Consejo y se derogan los Reglamentos (CE) N.º 78/2009, (CE) N.º 79/2009 y (CE) N.º 661/2009 del Parlamento Europeo y del Consejo y los Reglamentos (CE) N.º 631/2009, (UE) N.º 406/2010, (UE) N.º 672/2010, (UE) N.º 1003/2010, (UE) N.º 1005/2010 de la Comisión, (UE) N.º 1008/2010, (UE) N.º 1009/2010, (UE) N.º 19/2011, (UE) N.º 109/2011, (UE) N.º 458/2011, (UE) N.º 65/2012, (UE) N.º 130/2012, (UE) N.º 347/2012, (UE) N.º 351/2012, (UE) N.º 1230/2012 y (UE) 2015/166 (DO L 325 de 16.12.2019, p. 1).
26. Reglamento (UE) 2018/1139 del Parlamento Europeo y del Consejo, de 4 de julio de 2018, sobre normas comunes en el ámbito de la aviación civil y por el que se crea una Agencia de Seguridad Aérea de la Unión Europea, y se modifican los Reglamentos (CE) N.º 2111/2005, (CE) N.º 1008/2008, (UE) N.º 996/2010, (UE) N.º 376/2014 y las Directivas 2014/30/UE y 2014/53/UE del Parlamento Europeo y del Consejo, y se derogan los Reglamentos (CE) N.º 552/2004 y (CE) N.º 216/2008 del Parlamento Europeo y del Consejo y el Reglamento (CEE) N.º 3922/91 del Consejo (DO L 212 de 22.8.2018, p. 1), en lo que respecta al diseño, la producción y la comercialización de las aeronaves mencionadas en su artículo 2, apartado 1, letras a) y b), cuando se trate de aeronaves no tripuladas y de sus motores, hélices, piezas y equipos para controlarlas a distancia.
27. Norma que se aplica de forma articulada con el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo) y en la Directiva 2002/58/CE del Parlamento Europeo y del Consejo.
28. Dictamen del Comité Económico y Social Europeo sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo relativo a los requisitos horizontales de la

Esta última contiene los requisitos del esquema europeo de certificación de la ciberseguridad, entendido como el «conjunto completo, de disposiciones, requisitos técnicos, normas y procedimientos establecidos a escala de la Unión y que se aplican a la certificación o a la evaluación de la conformidad de los productos, servicios y procesos de TIC específicos», dicho esquema europeo es articulado a escala nacional a través del «esquema nacional de certificación de la ciberseguridad» entendido como el «conjunto completo de disposiciones, requisitos técnicos, normas y procedimientos desarrollados y adoptados por una autoridad pública nacional, y que se aplican a la certificación o a la evaluación de la conformidad de los productos, servicios y procesos de TIC incluidos en el ámbito de aplicación de dicho esquema específico», y, se materializa en términos prácticos como el «certificado europeo de ciberseguridad», que corresponde al «documento expedido por el organismo pertinente que certifica que determinado, producto, servicio o proceso de TIC ha sido evaluado para verificar que cumple los requisitos específicos de seguridad establecidos en un esquema europeo de certificación de la ciberseguridad».

II. EL MARCO EUROPEO DE CERTIFICACIÓN DE LA CIBERSEGURIDAD COMO INSTRUMENTO DE GARANTÍA DE LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL DE ALTO RIESGO

El marco europeo para la certificación de la ciberseguridad instaaura un esquema de certificación y confirmación de que los productos, procesos, los servicios asociados a las tecnologías de la información y las comunicaciones se han evaluado y cumplen con unos requisitos para proteger la autenticidad, integridad, disponibilidad y confidencialidad de los datos tanto que han sido almacenados o transmitidos o procesados, o cualquier servicio o función y que se puede acceder durante la vida el ciclo de vida tanto en los productos, servicios y procesos.

Con fundamento en el artículo 47 del Reglamento (UE) 2019/881 del Parlamento Europeo y del Consejo, de 17 de abril de 2019, se podría plantear que en los programas de trabajo evolutivos de la unión si incluyen productos, servicios o procesos, de tecnologías de la información y las comunicaciones y en particular sistemas de IA de los catalogados como de alto riesgo, y, con ello que los sistemas de inteligencia artificial tengan una certificación de seguridad independiente. Lo anterior justificado en el derecho o las políticas aplicables de la unión europea y particularmente las nuevas en torno a los sistemas de inteligencia artificial la propia demanda del mercado y la evolución de la ciber amenazas en entornos de inteligencia artificial.

El diseño de los esquemas europeos de certificación de ciberseguridad aborda varios objetivos clave para garantizar la seguridad en el ciclo de vida de productos, servicios o procesos de Tecnologías de la Información y Comunicación (TIC). Estos objetivos incluyen la protección contra accesos no autorizados, la preservación de la integridad y disponibilidad de los datos, la gestión adecuada de accesos autorizados, la detección y documentación de vulnerabilidades conocidas, el registro y verificación de actividades de acceso y uso, la eliminación de vulnerabilidades en productos

ciberseguridad para los productos con elementos digitales y por el que se modifica el Reglamento (UE) 2019/1020. Recuperado de <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52022AE4103>

y la rápida restauración de servicios en caso de incidentes. Además, se destaca la importancia de la seguridad por defecto y desde el diseño, así como la entrega de productos y servicios con programas y equipos actualizados y seguros, con mecanismos para actualizaciones de seguridad.

Los certificados europeos de ciberseguridad pueden especificar uno o más de los niveles de garantía siguientes para productos, servicios y procesos de TIC: «básico», «sustancial» o «elevado». El nivel de garantía asignado debe reflejar el riesgo asociado al uso previsto de un producto, servicio o proceso de TIC, considerando tanto la probabilidad como las posibles repercusiones de un incidente de ciberseguridad. Un certificado europeo de ciberseguridad o una declaración de conformidad de la UE, designada como nivel «básico», asegura que los productos, servicios y procesos de TIC cumplen con los requisitos de seguridad, minimizando los riesgos conocidos de ciberincidentes y ciberataques. La evaluación incluye, al menos, una revisión de la documentación técnica o actividades de evaluación equivalentes. En el caso de un certificado de nivel «sustancial», se garantiza que los productos, servicios y procesos de TIC cumplen con los requisitos de seguridad, minimizando riesgos de ciberseguridad conocidos y ataques de agentes con recursos limitados. La evaluación implica la revisión para demostrar la ausencia de vulnerabilidades conocidas y la verificación de la correcta aplicación de las funcionalidades de seguridad. Por último, un certificado de nivel «elevado» ofrece garantías de que los productos, servicios y procesos de TIC cumplen con los requisitos de seguridad, minimizando el riesgo de ciberataques sofisticados realizados por agentes con capacidades y recursos considerables.

Ante lo anterior, surge la pregunta de ¿cuál es el nivel de garantía que se exigirá a los sistemas de IA? La respuesta podría estar en la finalidad del sistema de IA, la norma prevé que «Los sistemas de IA de alto riesgo se diseñarán y desarrollarán de manera que alcancen, a la luz de su finalidad prevista, un nivel adecuado de (...) ciberseguridad». En la práctica, todo sistema de IA de alto riesgo, dada los niveles de riesgo que representan para los derechos se deberían integrar en el modelo de certificados de ciberseguridad «elevados». Disminuir la categoría a niveles inferiores implicaría una contradicción sustancial con la propia norma pues sería la minimización de riesgos de ciberataques, ciberincidentes y una revisión menor en cuanto a la técnica o actividades de evaluación.

De igual forma, en los sistemas de IA de alto riesgo se deberá dar cumplimiento a las obligaciones derivadas de la Directiva (UE) 2022/2555 del Parlamento Europeo y del Consejo, de 14 de diciembre de 2022, en lo que corresponde a la adopción de las estrategias nacionales de ciberseguridad, en las que se incorpore el apartado específico de los sistemas de IA de alto riesgo y designen o establezcan autoridades competentes, autoridades de gestión de crisis de ciberseguridad, puntos de contacto únicos sobre ciberseguridad y equipos de respuesta a incidentes de seguridad informática. De igual forma, la integración de estos en los modelos diseñados para la gestión de riesgos de ciberseguridad y obligaciones de notificación, y la identificación de las entidades cuyo tipo se enmarca en los anexos I o II; así como para las entidades identificadas como críticas con arreglo a la Directiva (UE) 2022/2557. Por otra parte, las obligaciones en cuanto al intercambio de información sobre ciberseguridad y las derivadas de las actividades de vigilancia, supervisión y control.

1. CERTIFICACIONES DE CIBERSEGURIDAD EN SISTEMAS DE INTELIGENCIA ARTIFICIAL DE ALTO RIESGO ¿OBLIGATORIAS O VOLUNTARIAS?

De la obligación de ciberseguridad impuesta en el artículo 15 de la nueva normativa de IA, se desprende la aplicación de las normas de ciberseguridad adoptadas anteriormente en el derecho de la Unión, en particular, el artículo 56 del Reglamento (UE) 2019/881²⁹ del Parlamento Europeo y del Consejo, de 17 de abril de 2019, en cuanto a las certificaciones de ciberseguridad y particularmente al numeral segundo de dicha disposición en la que se señala que la certificación de ciberseguridad será voluntaria, salvo que se disponga otra cosa en el derecho de la unión o de los Estados miembros.

Para los sistemas de IA de alto riesgo en los que se incorporó la obligación de ciberseguridad señalada en el artículo 15 debería entonces, por un lado, contar con el esquema europeo de certificación de ciberseguridad específico para sistemas de IA que cumpla los elementos señalados en el artículo 54 y, por otro lado, convertirse en obligatorio en desarrollo del derecho de la unión. Si fuere así deberán tener dicha certificación evaluaciones cada 2 años y al mismo tiempo la comisión debe determinar con base en los resultados de esa evaluación los productos servicios y procesos cubiertos con el esquema de certificación obligatoria de ciberseguridad de IA. De la misma manera, los fabricantes o proveedores de productos, servicios y procesos de TIC certificados o autoevaluados deben proporcionar la información sobre ciberseguridad complementaria señalada en el artículo 55, entre otras, orientaciones para el mantenimiento seguro de los productos, período de soporte, actualizaciones, información de vulnerabilidades.

En el desarrollo de certificados de ciberseguridad en los sistemas de IA también se deberá contemplar la revisión de pares contenida en el artículo 59, y que busca «alcanzar normas equivalentes en toda la Unión en lo que respecta a los certificados europeos de ciberseguridad expedidos y a las declaraciones de conformidad de la UE».

El esquema de certificación en ciberseguridad a los sistemas de IA de alto riesgo debe llevarse a las normas internas de cada uno de los Estados, seguramente, en unos más que otros, las adaptaciones serán considerables. Como ejemplo, y solo a título ilustrativo, véase la necesidad de adaptación de las normas españolas contenidas en la Orden PRE/2740/2007, de 19 de septiembre, por la que se aprueba el Reglamento de Evaluación y Certificación de la Seguridad de las Tecnologías de la Información, que tiene por objeto «la articulación del Organismo de Certificación (OC) del Esquema Nacional de Evaluación y Certificación de la Seguridad de las Tecnologías de la Información (ENECSTI) en el ámbito de actuación del Centro Criptológico Nacional, según lo dispuesto en la Ley 11/2002, de 6 de mayo, reguladora del Centro Nacional de Inteligencia, y en el Real Decreto 421/2004, de 12 de marzo, por el que se regula el Centro Criptológico Nacional, respectivamente». Así como, el reciente Real Decreto

29. Reglamento (UE) 2019/881 del Parlamento Europeo y del Consejo, de 17 de abril de 2019, relativo a ENISA (Agencia de la Unión Europea para la Ciberseguridad) y a la certificación de la ciberseguridad de las tecnologías de la información y la comunicación y por el que se deroga el Reglamento (UE) N.º 526/2013 («Reglamento sobre la Ciberseguridad»).

311/2022, de 3 de mayo, por el que se regula el Esquema Nacional de Seguridad y cuyo objeto regular el Esquema Nacional de Seguridad, establecido en el artículo 156.2 de la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público.

La seguridad de la información busca garantizar que una organización pueda alcanzar sus objetivos y realizar sus funciones mediante el uso de sistemas de información. Para lograr esto, se deben seguir principios fundamentales, que incluyen la consideración de la seguridad como un proceso integral, la gestión basada en riesgos, abordar la prevención, detección, respuesta y conservación, establecer líneas de defensa, mantener una vigilancia continua, realizar reevaluaciones periódicas y diferenciar responsabilidades. Estos principios son clave para establecer un enfoque efectivo y robusto en materia de seguridad de la información.

2. OBLIGACIONES DE CIBERSEGURIDAD EN LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL DE ALTO RIESGO

La ciberseguridad implica el desarrollo de capacidades en torno al proceso de identificación, protección, detección, respuesta y recuperación. Minimizar las vulnerabilidades y lograr que los riesgos no se materialicen requiere de un proceso de planeación y entendimiento de la ciberseguridad desde el inicio y durante todo el ciclo del proceso de desarrollo de los sistemas de inteligencia artificial de alto riesgo. La ciberseguridad se debe concebir a partir de la realización de un conjunto de acciones integrales en cada uno de los momentos de preparación, desarrollo, puesta en marcha y control del sistema de IA. Tanto el marco europeo de ciberseguridad como el marco nacional de ciberseguridad de cada uno de los estados de la Unión, atendiendo a los estándares técnicos, incluyen, integran y desarrollan, los modelos de ciberseguridad más conocidos. Andrew S. Tanenbaum señalaba «lo bueno de los estándares es que hay muchos donde elegir»; el modelo de madurez de los sistemas de gestión de seguridad de la información del ISM 3, el sistema de gestión de seguridad de la información de la ISO 27001, el modelo de madurez de la ciberseguridad comunitaria de White (CCSMM), el marco de ciberseguridad de la NIST, entre otros, que desarrollan de una u otra forma, las distintas etapas de para identificar, proteger, detectar, responder y recuperarse frente a ataques informáticos.

La obligación de ciberseguridad contenida en el artículo 15 implica el fortalecimiento de la función de ciberseguridad, desde la aplicación de estándares técnicos internacionales, para quienes se encuentren incluidos como sujetos obligados a cumplir las obligaciones derivadas del RIA y especialmente para los sistemas de inteligencia artificial de alto riesgo. La seguridad debe ser un principio inicial de dichos proyectos y en consecuencia deben establecerse las actividades que permitan gestionar los activos, identificar el entorno de los servicios, tener los niveles de gobernanza de la ciberseguridad identificados, contar con programas para la evaluación de los riesgos, identificar la estrategia para gestionar los riesgos, identificar el modelo de gestión de riesgos en la cadena de suministros; la realización de actividades de protección como por ejemplo la gestión de identidades, controles de acceso, procedimientos para la protección de la información, actividades de mantenimiento, incorporación de tecnologías para la protección y herramientas para la protección; sistemas que permitan detectar anomalías y eventos a partir de la incorporación de procesos de detección de dichas anomalías; desarrollo

de capacidades en torno a la respuesta frente a ataques informáticos, procesos y procedimientos de comunicaciones, procesos para análisis mitigación y mejora; por último, actividades de recuperación. En los sistemas de información de inteligencia artificial de alto riesgo se debe contar con planes de recuperación que permitan la continuidad de la actividad y la protección frente a hipotéticas lesiones de los derechos de los distintos actores del ecosistema en el que se materializa el sistema de IA.

Dentro de las obligaciones de ciberseguridad incorporadas en el anexo y que se refiere a la documentación técnica mencionada en el apartado 1 del artículo 11, se señala que contendrá como mínimo la siguiente información: «(...) La documentación técnica a que se refiere el apartado 1 del artículo 11 contendrá como mínimo la siguiente información, según proceda para el sistema de IA pertinente: (...) (ga) medidas de ciberseguridad implantadas».

Es de resaltar la vinculación e importancia que se da a la Agencia de la Unión Europea para la Ciberseguridad (ENISA) a fin de abordar cualquier problema emergente en el mercado interior en relación con la ciberseguridad, de tal forma que para ello colaborará con el Comité Europeo de Inteligencia Artificial.

También se establece la obligación para que «Los sistemas de IA de alto riesgo que continúan aprendiendo tras su introducción en el mercado o puesta en servicio se desarrollarán de tal modo que los posibles sesgos en la información de salida que influyan en los datos de entrada en futuras operaciones (“bucle de retroalimentación”) y la manipulación maliciosa de los datos de entrada utilizados para el aprendizaje durante el funcionamiento se subsanen debidamente con las medidas de mitigación oportunas», esta obligación implica que las medidas técnicas y administrativas implementadas deben prever la incorporación de estos elementos.

2.1. La ciberseguridad en sistemas de inteligencia artificial de alto riesgo implementados por autoridades

Cuando los sistemas de IA de riesgo alto sean utilizados por autoridades, estarán obligados al cumplimiento de las normas de ciberseguridad establecidas para estas autoridades. En el caso de España, por ejemplo, todo el sector público, en los términos en que este se define por el artículo 2 de la Ley 40/2015 de 1 de octubre, y de acuerdo con lo previsto en el artículo 156.2 de la misma. De igual forma, sin perjuicio de la aplicación de la Ley 9/1968 de 5 de abril, de Secretos Oficiales y otra normativa especial, los sistemas que tratan información clasificada y los sistemas de información de las entidades del sector privado, incluida la obligación de contar con la política de seguridad a que se refiere el artículo 12, cuando, de acuerdo con la normativa aplicable y en virtud de una relación contractual, presten servicios o provean soluciones a las entidades del sector público para el ejercicio por estas de sus competencias y potestades administrativas, están obligadas a cumplir las obligaciones derivadas del Real Decreto 311/2022, de 3 de mayo, por el que se regula el Esquema Nacional de Seguridad.

Dentro de las obligaciones de ciberseguridad que deberán cumplir las entidades que utilicen sistemas de inteligencia artificial de alto riesgo, está la organización e implantación del proceso de seguridad; gestión de riesgos, consistente en un proceso de identificación, análisis, evaluación y tratamiento de los mismos; gestión

de personal; profesionalidad; autorización y control de los accesos; protección de las instalaciones; adquisición de productos de seguridad y contratación de servicios de seguridad; mínimo privilegio; integridad y actualización del sistema; protección de la información almacenada y en tránsito; prevención ante otros sistemas de información interconectados; registro de la actividad y detección de código dañino; incidentes de seguridad; continuidad de la actividad; y mejora continua del proceso de seguridad. De igual manera se deberá integrar la utilización de infraestructuras y servicios comunes de las administraciones públicas en aras de lograr una mayor eficiencia y retroalimentación de las sinergias de cada colectivo. Se debe resaltar que el artículo 30 da la posibilidad de implementar perfiles de cumplimiento específicos, así como esquemas de acreditación de entidades de implementación de configuraciones seguras y el desarrollo de capacidades que permitan la auditoría de la seguridad, el informe del estado de la seguridad y la respuesta a incidentes de seguridad.

En lo que respecta a las actividades específicas de prevención, detección y respuesta a incidentes de seguridad, se deben cumplir con los estándares técnicos, así como con las normas de conformidad, que se concretan en cuatro: Administración Digital, ciclo de vida de servicios y sistemas, mecanismos de control y procedimientos de determinación de la conformidad con el ENS³⁰.

2.2. La ciberseguridad en los sistemas de inteligencia artificial de alto riesgo que hagan parte de actividades críticas o servicios esenciales

Dentro de los sistemas de IA de alto riesgo también se encuentran sistemas de IA destinados a ser utilizados como componentes de seguridad en la gestión y explotación de infraestructuras digitales críticas, tráfico rodado o del suministro de agua, gas, calefacción y electricidad. De igual forma, en la definición de incidente grave se incorporan los que recaen sobre infraestructuras críticas, al señalar «cualquier incidente o mal funcionamiento de un sistema de IA que, directa o indirectamente, conduzca, pueda haber conducido o pueda conducir a cualquiera de las siguientes situaciones» (...) (b) una interrupción grave e irreversible de la gestión y el funcionamiento de las infraestructuras críticas. En tal sentido, y atendiendo a las remisiones normativas y a la integración del conjunto de normas se deberán cumplir las obligaciones referidas a infraestructuras críticas.

Así las cosas, la Directiva 2008/114 del Consejo, de 8 de diciembre, sobre la identificación y designación de Infraestructuras Críticas Europeas y la evaluación de la necesidad de mejorar su protección y que en España es desarrollada a través de la Ley 8/2011, de 28 de abril, por la que se establecen medidas para la protección de las infraestructuras críticas y el Real Decreto 704/2011, de 20 de mayo, por el que se aprueba el Reglamento de protección de las infraestructuras tiene por objeto establecer medidas para la protección de las infraestructuras críticas, a fin de concretar las actuaciones de los distintos órganos integrantes del Sistema de Protección de Infraestructuras Críticas, incorpora un conjunto de obligaciones a las que estarán sometidos los distintos sujetos cuando incorporan sistemas de inteligencia artificial de alto riesgo.

30. Real Decreto 311/2022, de 3 de mayo, por el que se regula el Esquema Nacional de Seguridad.

Las obligaciones de ciberseguridad son aquellas que están contenidas en las referidas disposiciones resaltando a los efectos de este documento, los Planes de Seguridad del Operador que corresponde a los documentos estratégicos que definen las políticas generales de los operadores críticos para asegurar la seguridad de sus instalaciones o sistemas, que estos son evaluados y aprobados por el secretario de Estado de Seguridad. Estos planes deben incluir una metodología de análisis de riesgos que garantice la continuidad de los servicios, abordando amenazas físicas y lógicas, con criterios para la implementación de medidas de seguridad, también en los sistemas de IA que se encuentran definidos como de alto riesgo, así como los mecanismos para implantación, control y seguimiento.

2.3. Evaluaciones de conformidad como instrumento para la ciberseguridad en los sistemas de inteligencia artificial catalogados como de alto riesgo

Las certificaciones de conformidad son documentos emitidos por organismos de certificación o entidades autorizadas que atestatan que un producto, servicio, sistema o proceso cumple con ciertos estándares, normas o especificaciones previamente establecidos. Estas certificaciones son una forma de garantizar que un producto o servicio cumple con los requisitos y estándares establecidos por las autoridades competentes o por organizaciones especializadas. Al obtener una certificación de conformidad, una entidad demuestra que ha sido evaluada y ha demostrado cumplir con los criterios y requisitos específicos establecidos para su industria o sector. Esto puede abarcar aspectos como calidad, seguridad, eficiencia energética, sostenibilidad ambiental, seguridad de la información, entre otros. Las certificaciones de conformidad pueden ser obligatorias para ciertos productos o servicios, especialmente en áreas reguladas por normativas gubernamentales. También pueden ser voluntarias y buscadas por las empresas como una manera de destacar la calidad y el cumplimiento de estándares reconocidos, lo que puede generar confianza en los consumidores y en el mercado en general. Ejemplos comunes de certificaciones incluyen la certificación ISO 9001 para sistemas de gestión de calidad, la certificación CE en la Unión Europea, y diversas certificaciones de seguridad y estándares industriales en diferentes sectores.

La normalización europea tiene como antecedente legislativo específico, entre otros, cuatro actos diferentes que se requiere relacionar pues de ellos derivan las obligaciones de ciberseguridad a los que remite la nueva norma en sistemas de inteligencia artificial de alto riesgo: la Directiva 98/34/CE del Parlamento Europeo y del Consejo, de 22 de junio de 1998, por la que se establece un procedimiento de información en materia de las normas y reglamentaciones técnicas y de las reglas relativas a los servicios de la sociedad de la información³¹, la Decisión no 1673/2006/CE del Parlamento Europeo y del Consejo, de 24 de octubre de 2006, relativa a la financiación de la normalización europea³², y la Decisión 87/95/CEE,

31. Directiva 98/34/CE del Parlamento Europeo y del Consejo, de 22 de junio de 1998, por la que se establece un procedimiento de información en materia de las normas y reglamentaciones técnicas y de las reglas relativas a los servicios de la sociedad de la información (DO L 204 de 21 de julio de 1998, p. 37).

32. Decisión N.º 1673/2006/CE del Parlamento Europeo y del Consejo, de 24 de octubre de 2006, relativa a la financiación de la normalización europea (DO L 315 de 15 de

de 22 de diciembre de 1986, relativa a la normalización en el campo de la tecnología de la información y de las telecomunicaciones³³, y Reglamento n.º 1025/2012 del Parlamento Europeo y del Consejo de 25 de octubre de 2012 sobre la normalización europea, por el que se modifican las Directivas 89/686/CEE y 93/15/CEE del Consejo y las Directivas 94/9/CE, 94/25/CE, 95/16/CE, 97/23/CE, 98/34/CE, 2004/22/CE, 2007/23/CE, 2009/23/CE y 2009/105/CE del Parlamento Europeo y del Consejo y por el que se deroga la Decisión 87/95/CEE del Consejo y la Decisión n.º 1673/2006/CE del Parlamento Europeo y del Consejo.

La regulación nueva de sistemas de IA de alto riesgo señala al respecto que «Un sistema de IA que sea a su vez un producto cubierto por la legislación de armonización de la Unión enumerada en el anexo II se considerará de alto riesgo si debe someterse a una evaluación de la conformidad por parte de terceros con vistas a la comercialización o puesta en servicio de dicho producto con arreglo a la legislación mencionada». (...) y que «Un sistema de IA destinado a ser utilizado como componente de seguridad de un producto cubierto por la legislación mencionada en el apartado 1 se considerará de alto riesgo si se le exige que se someta a una evaluación de la conformidad por parte de terceros con vistas a la comercialización o puesta en servicio de dicho producto con arreglo a la legislación mencionada. Esta disposición se aplicará con independencia de que el sistema de IA se comercialice o se ponga en servicio independientemente del producto».

En este sentido, si los sistemas de IA de alto riesgo o los sistemas de IA de uso general que hayan sido certificados o para los que se haya emitido una declaración de conformidad en el marco de un régimen de ciberseguridad con arreglo al Reglamento (UE) 2019/881 del Parlamento Europeo y del Consejo y cuyas referencias se hayan publicado en el Diario Oficial de la Unión Europea «se presumirán conformes con los requisitos de ciberseguridad establecidos en el artículo 15 (...) en la medida en que el certificado de ciberseguridad o la declaración de conformidad o sus partes cubran dichos requisitos».

III. CONCLUSIONES

La ciberseguridad implica el desarrollo de capacidades en torno al proceso de identificación, protección, detección, respuesta y recuperación. Minimizar las vulnerabilidades y lograr que los riesgos no se materialicen requiere de un proceso de planeación y entendimiento de la ciberseguridad desde el inicio y durante todo el ciclo del proceso de desarrollo de los sistemas de inteligencia artificial de alto riesgo.

Mitigar eficazmente los riesgos en el contexto de la inteligencia artificial, implica imponer requisitos que aborden varios aspectos, incluyendo la calidad de los datos, la gestión de documentación técnica, el mantenimiento de registros, la transparencia, la solidez, la precisión, la supervisión humana y, por supuesto, la ciberseguridad. Aunque la normativa se centra en los sistemas de IA de alto riesgo, se reconoce que las obligaciones de ciberseguridad no son exclusivas para estos sistemas, sino que

noviembre de 2006, p. 9).

33. Decisión 87/95/CEE, de 22 de diciembre de 1986, relativa a la normalización en el campo de la tecnología de la información y de las telecomunicaciones (DO L 36 de 7 de febrero de 1987, p. 31).

se aplican a todas las tecnologías de la información y las comunicaciones de manera proporcional a los fines e intereses perseguidos.

Las obligaciones de ciberseguridad están vinculadas a los sistemas de IA de alto riesgo contenidos en el anexo II y III del nuevo reglamento y aquellos sometidos a una evaluación de la conformidad por parte de terceros con vistas a la comercialización o puesta en servicio de dicho producto.

Si bien se incorporan obligaciones en materia de ciberseguridad, también es cierto que la norma direcciona la ciberseguridad a la aplicación de las normas existentes y al nivel adecuado a la finalidad prevista. En consecuencia, se deben aplicar, entre otras, las normas de ciberseguridad dispuestas por la Unión Europea y en particular la Directiva (UE) 2022/2555 del Parlamento Europeo y del Consejo, de 14 de diciembre de 2022, sobre medidas para un alto nivel común de ciberseguridad en toda la Unión, por la que se modifica el Reglamento (UE) – 910/2014 y la Directiva (UE) 2018/1972, y por la que se deroga la Directiva (UE) 2016/1148 (Directiva NIS 2), y el Reglamento (UE) 2019/881 del Parlamento Europeo y del Consejo, de 17 de abril de 2019, relativo a ENISA (Agencia de la Unión Europea para la Ciberseguridad) y a la certificación de la ciberseguridad de las tecnologías de la información y la comunicación y por el que se deroga el Reglamento (UE) – 526/2013 («Reglamento sobre la Ciberseguridad»), Reglamento 2022/0272, entre otros.

En los sistemas de inteligencia artificial de alto riesgo se direcciona a la aplicación del marco europeo para la certificación de la ciberseguridad que instaura un esquema de certificación y confirmación de que los productos, procesos, los servicios asociados a las tecnologías de la información y las comunicaciones se han evaluado y cumplen con unos requisitos para proteger la autenticidad, integridad, disponibilidad y confidencialidad de los datos tanto que han sido almacenados o transmitidos o procesados, o cualquier servicio o función y que se puede acceder durante la vida el ciclo de vida tanto en los productos, servicios y procesos.

Los sistemas de IA de alto riesgo se deberá dar cumplimiento a las obligaciones derivadas de la Directiva (UE) 2022/2555 del Parlamento Europeo y del Consejo, de 14 de diciembre de 2022, en lo que corresponde a la adopción de las estrategias nacionales de ciberseguridad, en las que se incorpore el apartado específico de los sistemas de IA de alto riesgo y designen o establezcan autoridades competentes, autoridades de gestión de crisis de ciberseguridad, puntos de contacto únicos sobre ciberseguridad y equipos de respuesta a incidentes de seguridad informática.

Por último, se debe resaltar la vinculación e importancia que se da a la Agencia de la Unión Europea para la Ciberseguridad (ENISA) a fin de abordar cualquier problema emergente en el mercado interior en relación con la ciberseguridad, de tal forma que para ello colaborará con el Comité Europeo de Inteligencia Artificial.

Vigilancia poscomercialización en los sistemas de inteligencia artificial de alto riesgo en el Reglamento. Descripción, medidas y casos de USO

IDOIA SALAZAR

Doctora. Profesora de la Universidad CEU San Pablo. Presidenta de Odiseia

MIGUEL ÁNGEL LIÉBANAS

Criminólogo experto en Sistemas Inteligentes. Odiseia.

CEO de Human Trends

I. LA VIGILANCIA POSCOMERCIALIZACIÓN EN EL REGLAMENTO

1. INTRODUCCIÓN

La rápida evolución y adopción de los sistemas de inteligencia artificial (IA) en diversas áreas, desde la medicina hasta la seguridad nacional, han traído consigo una serie de beneficios significativos. Sin embargo, la complejidad inherente y las capacidades en constante evolución de estos sistemas plantean desafíos únicos en términos de seguridad, privacidad, ética y gobernanza. Específicamente, los sistemas de IA de alto riesgo, aquellos cuyo mal funcionamiento o uso indebido podría tener graves consecuencias para los individuos o la sociedad, requieren una consideración especial. En este contexto, la vigilancia poscomercialización emerge como un componente crítico para asegurar que estos sistemas operen de manera segura, efectiva y ética a lo largo de su ciclo de vida.

Este artículo 72 RIA es el relativo a la «Vigilancia poscomercialización por parte de los proveedores y plan de vigilancia poscomercialización para sistemas de IA de alto riesgo». Dicho precepto destaca la importancia de implementar un plan de vigilancia postcomercialización robusto y sistemático para estos sistemas de IA de alto riesgo. Así, se muestra que estos planes son fundamentales para identificar y mitigar los riesgos emergentes asociados con el uso a largo plazo de la IA. Además, se incide en su papel crucial en la construcción de la confianza del público y en el fomento de la responsabilidad y la transparencia por parte de los desarrolladores y usuarios de la IA. En este sentido, se abordan temas clave como la identificación de indicadores de riesgo, la monitorización continua del rendimiento, la gestión de

la retroalimentación de los usuarios y la adaptación a las dinámicas tecnológicas y sociales cambiantes. Por otro lado, también se recalcan las implicaciones de una vigilancia postcomercialización insuficiente, incluidos los riesgos de perjuicios inconscientes, la pérdida de confianza pública y las posibles barreras regulatorias que podrían inhibir la innovación responsable.

En cualquier caso, se pretende subrayar que la vigilancia postcomercialización no es simplemente una obligación regulatoria, sino una oportunidad estratégica para los desarrolladores y usuarios de IA. Es una manera de garantizar que los sistemas de IA —sean o no de alto riesgo— no solo cumplan con sus objetivos iniciales, sino que también se deben adaptar y mejorar de manera responsable en respuesta a los desafíos emergentes y a las expectativas de la sociedad.

2. ¿QUÉ ES UN PLAN DE VIGILANCIA POSTCOMERCIALIZACIÓN Y QUÉ INCLUYE?

El artículo 3 en su apartado 25 define el «sistema de vigilancia poscomercialización» como **«todas las actividades realizadas por los proveedores de sistemas de IA destinadas a recoger y examinar la experiencia obtenida con el uso de sistemas de IA que introducen en el mercado o ponen en servicio, con objeto de detectar la posible necesidad de aplicar inmediatamente cualquier tipo de medida correctora o preventiva que resulte necesaria»**. Así pues, un Plan de Vigilancia postcomercialización supone un conjunto de procesos y herramientas orientados a recabar datos de un sistema para transformarlos en una serie de indicadores sobre su actividad con el objetivo de supervisar los sistemas de inteligencia artificial (IA) después de su lanzamiento al mercado. En este sentido, dicho plan incluiría una serie de tareas, siempre teniendo en cuenta la finalidad prevista del sistema. Serían las siguientes:

- Recolección de datos sobre el rendimiento y la seguridad del sistema.
- Evaluación de las posibles causas de problemas detectados.
- Implementación de soluciones para corregir problemas.
- Comunicación de los resultados y recomendaciones a las partes interesadas.

A continuación, se describen con un mayor detalle cada una de ellas:

A) *Recolección de datos sobre el rendimiento y la seguridad del sistema*

Esto implica la monitorización proactiva del sistema de IA para recoger información sobre su operatividad y cualquier evento adverso o desviaciones del comportamiento esperado. Los datos pueden provenir de una variedad de fuentes, incluidos registros de errores, *feedback* de los usuarios y otros sistemas de detección de anomalías. En este sentido es importante tener en cuenta varios componentes:

— *Definición de métricas relevantes:* Antes de recolectar datos, es fundamental definir qué métricas reflejan de manera precisa el rendimiento y la seguridad del sistema. Estas pueden incluir la precisión de las predicciones, la velocidad de procesamiento, la tasa de fallos, la frecuencia de falsos positivos o negativos y otros indicadores de estabilidad y fiabilidad. Lo abordaremos en mayor profundidad en la sección de Indicadores.

— *Sistemas de seguimiento en tiempo real*: Implementar sistemas que monitoricen continuamente el sistema de IA en busca de anomalías. Estos sistemas deben ser capaces de registrar eventos en tiempo real y proporcionar alertas tempranas de posibles problemas de seguridad o rendimiento.

— *Recopilación de datos de retroalimentación*: Los usuarios del sistema suelen ser una fuente rica en retroalimentación. Pueden reportar problemas que no son evidentes para los sistemas automáticos de monitorización, proporcionando así una visión más holística del rendimiento y la seguridad del sistema de IA.

— *Análisis de impacto*: **Más allá de la recolección de datos técnicos, es importante entender el impacto de la IA en el entorno en el que se despliega. Esto puede incluir el análisis de cómo las decisiones de la IA afectan a las personas, los procesos de negocio y otros sistemas tecnológicos.**

— *Intercambio seguro de datos*: Teniendo en cuenta que estos datos pueden incluir información sensible, es crucial asegurarse de que la recolección y el almacenamiento de los mismos se realicen de manera segura y en conformidad con las regulaciones de privacidad de datos aplicables.

— *Benchmarks y pruebas de estrés*: Regularmente, el sistema de IA debe someterse a pruebas de rendimiento contra benchmarks o estándares de la industria para evaluar su funcionamiento bajo diferentes condiciones y cargas de trabajo.

— *Registro y documentación*: Mantener un registro completo y documentación detallada de todos los datos recogidos es importante para el seguimiento a largo plazo y la auditoría del sistema. Esto también ayuda a establecer una línea base para entender su evolución a lo largo del tiempo.

— *Incorporación de nuevos datos*: Los sistemas de IA pueden desviarse de su rendimiento esperado a medida que se exponen a situaciones no previstas en su fase de entrenamiento. Incorporar nuevos datos recolectados durante la operación en el entrenamiento de la IA puede ayudar a que el sistema se adapte y mejore con el tiempo.

— *Evaluación continua de la adecuación de los datos*: A medida que el entorno y los contextos cambian se debe evaluar si los datos en los que se basa la IA siguen siendo representativos y adecuados para la tarea que se supone que debe realizar.

B) Evaluación de las posibles causas de problemas detectados

Al identificar problemas dentro de un sistema, se procede a una fase de análisis profundo para descifrar las causas fundamentales detrás de estos contratiempos. Este análisis puede requerir el empleo de metodologías avanzadas en el manejo de datos, como el aprendizaje automático y la minería de datos, para revelar patrones y correlaciones que no sean inmediatamente obvios.

En este sentido, el primer paso involucra un escrutinio meticuloso del incidente, que puede incluir la revisión de registros, la identificación de las condiciones específicas bajo las cuales el problema se manifestó y la interacción con los usuarios impactados para comprender su experiencia. Después se utilizan herramientas de diagnóstico para examinar el estado interno del sistema de IA, incluyendo la revisión de registros de depuración y el uso de monitores de rendimiento, con el fin de entender cómo opera el sistema.

Por otro lado, la aplicación de un análisis causal es fundamental para determinar las conexiones entre diversos factores y el problema identificado, analizando la secuencia de eventos que condujo al incidente y examinando tanto los datos de entrada como las decisiones tomadas por el sistema de IA. Para verificar las hipótesis sobre las causas potenciales, se llevan a cabo experimentos controlados o simulaciones para evaluar si el problema puede reproducirse bajo las mismas condiciones.

Debemos tener en cuenta que es muy importante revisar el código fuente y los algoritmos utilizados por el sistema de IA para detectar posibles fallos o errores en la lógica que podrían estar generando los problemas. Además, se emplean técnicas de minería de datos y aprendizaje automático para descubrir patrones ocultos que puedan estar contribuyendo a la situación problemática.

Como hemos visto anteriormente, la revisión de los datos utilizados para entrenar el sistema de IA es otro aspecto crítico. De esta manera se asegura que estos estén completos, precisos y no estén sesgados, ya que deficiencias en los datos de entrenamiento pueden traducirse en un rendimiento inadecuado del sistema. La retroalimentación de los usuarios finales ofrece perspectivas valiosas sobre el comportamiento del sistema en entornos reales y cómo estos comportamientos están relacionados con los problemas detectados.

En este ámbito es igualmente importante la evaluación del entorno operativo en el que se despliega la IA, ya que factores externos, como cambios en el hardware, el software complementario o las condiciones ambientales, pueden influir en el rendimiento del sistema.

Finalmente, la consulta con expertos en el área, como ingenieros de software, científicos de datos o especialistas en el dominio de aplicación de la IA, también puede proporcionar perspectivas adicionales sobre las causas de los problemas.

C) *Implementación de soluciones para corregir problemas*

Poniendo como base la evaluación de las causas, se diseñan e implementan soluciones. Esto puede incluir la actualización de algoritmos, la modificación de conjuntos de datos, la revisión de procesos de decisión automática o la mejora de los protocolos de seguridad.

La agilidad y la eficiencia en este proceso es importante mitigar los riesgos y prevenir la escalada de problemas. En este sentido, es necesario tener en cuenta la priorización de problemas, el desarrollo de las correcciones, hacer pruebas rigurosas, y revisar las implicaciones éticas y regulatorias. Antes de entrar brevemente en cada punto, es importante recalcar que es preferible implementar la solución de manera gradual, primero en un entorno de prueba, luego a un grupo pequeño de usuarios reales, y finalmente a toda la base de usuarios, para minimizar el riesgo.

— Respecto a la Priorización de Problemas: Basándonos en la gravedad y el impacto de los problemas identificados, se establece una prioridad para abordarlos. Esto implica considerar el riesgo para los usuarios y la organización, así como la frecuencia y las consecuencias del problema. Dado que el análisis de riesgos del sistema inteligente se habrá desarrollado previamente, podremos emplearlo como modelo para determinar la priorización.

— Respecto al desarrollo de correcciones: Trabajo para desarrollar correcciones específicas, ya sea en el código, en los algoritmos o en los datos utilizados por la IA.

En algunos casos, esto puede significar reentrenar modelos con nuevos datos o ajustar los parámetros del modelo.

— Respecto a la rigurosidad de las pruebas: Antes de implementar cualquier solución, se realizan pruebas exhaustivas para asegurarse de que no solo se resuelve el problema sino que también no introduce nuevos problemas. Esto puede incluir pruebas unitarias, pruebas de integración, pruebas de sistema y pruebas de aceptación del usuario.

— Respecto a la revisión de las implicaciones Éticas y Regulatorias: Cada solución propuesta debe revisarse para asegurarse de que cumple con las normativas aplicables (RIA u otras aplicables en función del país donde esté implementada la solución) y se adhiere a los estándares éticos, especialmente en términos de privacidad y equidad.

D) Comunicación de los resultados y recomendaciones a las partes interesada

Es vital mantener un diálogo abierto y transparente con todas las partes involucradas, incluyendo reguladores, usuarios finales y el público en general. La comunicación efectiva sobre cómo se están manejando los problemas y las mejoras implementadas es esencial para mantener la confianza en el sistema de IA.

Estas tareas se integran en un marco de gobernanza y gestión de riesgos que también debe incluir la adaptabilidad y la mejora continua. A medida que el sistema de IA aprende y evoluciona, también lo hace el entendimiento de sus riesgos potenciales, requiriendo un enfoque dinámico para la gestión de la vigilancia postcomercialización.

3. ¿POR QUÉ ES NECESARIO UN PLAN DE VIGILANCIA POST-COMERCIALIZACIÓN?

La creación de un plan de vigilancia post-comercialización para los sistemas de IA de alto riesgo no solo es una necesidad contemplada en el RIA, sino que es muy recomendable para mantener la eficiencia y la seguridad del sistema a lo largo del tiempo. Además, este tipo de vigilancia es clave para mantener la confianza del público y para que los sistemas se adapten a las cambiantes necesidades y datos.

Este elemento «la continuidad» en la seguridad y eficiencia de los sistemas de IA es crucial; sin un seguimiento adecuado, estos pueden enfrentar problemas inesperados o una disminución en el rendimiento debido a cambios en los patrones de datos o en su entorno operativo. La vigilancia post-comercialización permite la detección temprana y la corrección de estos problemas, previniendo daños significativos y preservando la confianza en la tecnología.

Además, la adaptación a nuevos datos y contextos es importante en el dinámico entorno tecnológico actual. Los sistemas de IA, especialmente aquellos basados en aprendizaje automático, requieren actualizaciones regulares para mantener su relevancia y eficacia. Un sistema de vigilancia efectivo garantiza que estas actualizaciones se realicen de manera oportuna, permitiendo que la IA responda adecuadamente a situaciones nuevas y no previstas. Por otro lado, la innovación responsable es un objetivo clave en este plan. Al monitorizar el desempeño y los impactos de los sistemas de IA después de su lanzamiento, los desarrolladores pueden identificar áreas para mejoras y avances tecnológicos. Esto no solo previene riesgos, sino que también fomenta una innovación ética y sostenible.

El cumplimiento regulatorio, específicamente el RIA, también es una consideración importante, dado que las regulaciones en torno a la IA están en constante evolución. Un plan de vigilancia post-comercialización asegura que los sistemas de IA permanezcan en conformidad con las normativas vigentes, evitando sanciones legales y protegiendo a los usuarios.

Finalizamos este apartado resaltando nuevamente la importancia de la confianza del usuario y la transparencia. Un sistema de vigilancia que promueva la rendición de cuentas puede fortalecer la confianza del público en la IA, demostrando un compromiso continuo con la seguridad y la responsabilidad.

4. EL PLAN DE VIGILANCIA POSTCOMERCIALIZACIÓN EN EL REGLAMENTO

El Título VIII del RIA concreta la normativa relativa al seguimiento postcomercialización, el intercambio de información y la vigilancia del mercado. Concretamente, el capítulo I hace referencia exclusiva al seguimiento posterior a la comercialización.

El capítulo cuenta únicamente con el artículo 61 enfocado en el seguimiento postcomercialización por parte de los proveedores y plan de seguimiento postcomercialización para los sistemas de IA de alto riesgo.

Primero, el apartado 61.1 indica la obligatoriedad de definir, implementar y documentar el sistema de vigilancia de forma proporcionada a los sistemas inteligentes desplegados.

«Los proveedores establecerán y documentarán un sistema de vigilancia postcomercialización de manera proporcionada a la naturaleza de las tecnologías de inteligencia artificial y a los riesgos del sistema de IA de alto riesgo.» (art. 61.1, título VIII, RIA).

En el siguiente apartado se indica un resumen de las funciones de dicho sistema de vigilancia y cuál es el objetivo del mismo: evaluar que los requisitos del título III, capítulo 2 se mantienen durante todo el ciclo de vida del sistema inteligente.

«El sistema de seguimiento poscomercialización recopilará, documentará y analizará de forma activa y sistemática los datos pertinentes que puedan facilitar los implantadores o que puedan recopilarse a través de otras fuentes sobre el funcionamiento de los sistemas de IA de alto riesgo a lo largo de su vida útil, y permitirá al proveedor evaluar la conformidad continua de los sistemas de IA con los requisitos establecidos en el título III, capítulo 2. Cuando proceda, el seguimiento posterior a la comercialización incluirá un análisis de la interacción con otros sistemas de IA. Esta obligación no cubrirá los datos operativos sensibles de los implantadores que sean autoridades policiales» (art 61.2, título VIII, RIA).

El tercer apartado del artículo 61 se centra en cómo desarrollar el sistema de vigilancia en base al diseño de un plan de vigilancia poscomercialización. En el apartado 2 **«Cómo abordar el requisito de la Vigilancia Poscomercialización» del presente análisis se abordará en profundidad dicho apartado.**

Por último, el apartado 4 establece una serie de excepciones en las cuáles no será necesario desarrollar un sistema de vigilancia poscomercialización:

Sistemas de IA de alto riesgo cubiertos por los actos jurídicos mencionados en la sección A del anexo II siempre que cuenten con un sistema y plan de vigilancia que tenga un nivel de protección equivalente. Por ejemplo, los sistemas de alto riesgo relacionados con la seguridad de los juguetes.

También se aplica a sistemas de IA de alto riesgo del punto 5, anexo III comercializados o puestos en servicio por entidades financieras.

De esta forma, el RIA establece la necesidad del sistema de vigilancia postcomercialización, indica que su objetivo es mantener los criterios de control de los sistemas inteligentes de título III, capítulo 2 durante todo el ciclo de vida del sistema y determina que el plan de vigilancia será la base sobre la que diseñar e implementar el sistema.

5. QUIÉN DEBE REALIZAR EL SISTEMA DE VIGILANCIA POSTCOMERCIALIZACIÓN

La implementación y gestión de un sistema de vigilancia postcomercialización, según lo estipulado en el marco regulatorio del RIA, recae primordialmente en los proveedores de los sistemas de IA. Este enfoque garantiza que los sistemas de IA, una vez implementados y en funcionamiento, continúen cumpliendo con los estándares y regulaciones establecidos a lo largo de su ciclo de vida. Es imperativo que los proveedores asuman la responsabilidad de evaluar y asegurar la conformidad continua de sus sistemas con los requisitos legales y éticos pertinentes, incluyendo aspectos de seguridad, privacidad y transparencia.

En este sentido, los proveedores deben establecer procedimientos robustos para el monitoreo constante de sus sistemas de IA. Esto implica no solo la revisión técnica del funcionamiento del sistema, sino también la consideración de cómo los cambios en el entorno operativo o en los datos pueden afectar el desempeño de la IA.

Además, los proveedores tienen la responsabilidad de diseñar mecanismos para la recopilación y análisis de *feedback*, incluyendo la notificación de incidentes y comportamientos anómalos por parte de los usuarios. Esto significa que los sistemas de IA deben ser diseñados con capacidades para registrar y reportar cualquier fallo o desviación en su comportamiento, facilitando así un proceso eficiente de retroalimentación entre los usuarios y los proveedores.

La responsabilidad de los usuarios.

Por otro lado, los usuarios (no finales) de sistemas de IA también juegan un papel importante en este ecosistema al actuar como observadores activos de la tecnología en uso. Se espera que los usuarios notifiquen cualquier incidente, fallo o comportamiento inusual del sistema al proveedor. Esta colaboración entre usuarios y proveedores es esencial para la detección temprana de problemas y para garantizar que se tomen medidas correctivas de manera oportuna.

La sinergia entre proveedores y usuarios, respaldada por un marco regulatorio claro como el RIA, facilita un entorno donde los sistemas de IA no solo son vigilados y evaluados constantemente, sino que también se promueve una mejora continua. Esto asegura que los sistemas de IA mantengan altos niveles de confiabilidad, seguridad

y conformidad con las normativas vigentes, al mismo tiempo que se adapta a las necesidades cambiantes de la sociedad y a los avances tecnológicos.

II. CÓMO ABORDAR EL REQUISITO DE LA VIGILANCIA POSCOMERCIALIZACIÓN

Tras presentar el concepto de la vigilancia poscomercialización vamos a comprender cómo abordar este requisito desde una perspectiva procedimental y técnica. Durante los siguientes apartados se presentarán algunos conceptos técnicos para su implementación pero siempre acompañados de una breve explicación que facilitará su comprensión sin requerir de ningún conocimiento previo.

1. PLAN DE VIGILANCIA Y SISTEMA DE VIGILANCIA. ELEMENTOS CLAVE DE LA VIGILANCIA

El principal objetivo de la Vigilancia Poscomercialización es comprobar que los requisitos establecidos en el capítulo III del RIA se cumplen durante todo el ciclo de vida del sistema inteligente. Para lograr dicho objetivo, necesitamos de dos componentes interrelacionados.

Primero, es necesario contar con un conjunto de procesos y protocolos con los que definir la actividad de vigilancia *per se*. Por otra parte, se requerirá de un sistema de monitorización bajo la perspectiva técnica para obtener todas las métricas necesarias del sistema inteligente, procesarlas en su debido tiempo para analizarlas y, si se da el caso, obtener las alertas correspondientes ante incidentes. Estos dos sistemas podemos definirlos de la siguiente forma:

— *Plan de Vigilancia Poscomercialización*. Aunque el RIA no ofrece una definición del Plan de Vigilancia, se puede inferir a través del articulado que se trata de un conjunto de protocolos y diseños que dan estructura y operatividad al Sistema de Vigilancia Poscomercialización. En otras palabras, se trata del plan que se debe de seguir para lograr una monitorización satisfactoria del sistema. Dentro de dicho plan se deberá de establecer qué tareas desarrollar, las responsabilidades del personal asociado al sistema, el propio diseño técnico del Sistema de Vigilancia y documentar todo su contenido en la documentación técnica del sistema inteligente. Por lo tanto, el diseño del Sistema de Vigilancia formará parte del contenido del Plan de Vigilancia:

«El sistema de vigilancia poscomercialización se basará en un plan de vigilancia poscomercialización. El plan de vigilancia poscomercialización formará parte de la documentación técnica a que se refiere el anexo IV» (art 61.3, título VIII, RIA).

— Respecto al contenido del Plan el propio RIA indica que:

«La Comisión adoptará un acto de ejecución en el que se establecerán disposiciones detalladas que constituyan un modelo para el plan de vigilancia poscomercialización y la lista de elementos que deberán incluirse en él a más tardar seis meses antes de la fecha de aplicación del presente Reglamento» (art. 61.3, título VIII, RIA).

Por lo tanto, aún no contamos con el modelo ni los requisitos del plan de vigilancia poscomercialización. Sin embargo, en el presente análisis se indicará el contenido mínimo de todo plan de vigilancia que con alta probabilidad formará parte de los requisitos establecidos por la comisión.

— *Sistema de Vigilancia Poscomercialización*. El RIA indica que el Sistema de Vigilancia son:

«todas las actividades realizadas por los proveedores de sistemas de IA destinadas a recoger y examinar la experiencia obtenida con el uso de sistemas de IA que introducen en el mercado o ponen en servicio, con objeto de detectar la posible necesidad de aplicar inmediatamente cualquier tipo de medida correctora o preventiva que resulte necesaria;» (art. 3.25, capítulo I, RIA).

Aunque en el texto se haga referencia a actividades resulta más intuitivo pensar en el sistema como un conjunto de procesos automatizados (y excepcionalmente manuales) que tienen como objetivo obtener los datos necesarios para evaluar que se siguen cumpliendo los requisitos del capítulo III aplicados al sistema inteligente. También, este sistema deberá de poder detectar cualquier falla en el sistema con la mayor brevedad posible, siempre actuando de forma preventiva si es posible.

De esta forma, el Plan de Vigilancia abordará los procedimientos necesarios para desarrollar la monitorización del sistema inteligente y el Sistema de Vigilancia será la herramienta que permitirá llevar a cabo dicha labor. El Plan definirá el qué, cuándo, cómo y quién realizará la vigilancia y el Sistema será la herramienta para apoyar toda esa labor.

2. Diseño del Sistema de Vigilancia

El objetivo de esta sección no es adentrarnos en los requisitos técnicos y las herramientas útiles para desarrollar el Sistema de Vigilancia sino obtener una intuición de qué componentes forman dicho sistema y cuáles son los requisitos que debemos de esperar del sistema.

Para obtener una visión global del sistema conviene comenzar por el final, es decir, por los outputs o resultados del sistema. En este caso son el panel de vigilancia del sistema y el sistema de alertas preconfigurado.

Primero, el panel de vigilancia permitirá a los responsables de vigilancia del sistema monitorizar que todos los indicadores (concepto que abordaremos en la siguiente sección) se mantienen en su rango de normalidad y el sistema funciona conforme a lo esperado.

En segunda instancia, para mantener una vigilancia continuada y poder responder de forma rápida a los incidentes, incluso llegando a prevenirlos, es necesario un sistema de alertas con los protocolos de comunicación necesarios para mantener al tanto a los responsables ante cualquier cambio de los indicadores del sistema.

Ahora se definirá de dónde obtendrán los datos el panel de vigilancia y el sistema de alertas para evaluar el estado actual del sistema. Para ello se empleará un sistema de base de datos en el que se almacenarán todos los registros de los indicadores seleccionados. El tipo de base de datos dependerá de la arquitectura del propio sistema inteligente y el caso de uso desarrollado. En definitiva, este sistema de almacenamiento funcionará como el hub de información proveniente de los registros de actividad de los sistemas inteligentes desplegados y ofrecerá dicha información al panel de vigilancia y el sistema de alertas.

Por último, se implementará un sistema de envío de registros desde el punto de despliegue del sistema inteligente al sistema de almacenamiento de registros. A modo de ejemplo, si el sistema inteligente realiza su procesamiento en el propio

dispositivo, como una cámara de seguridad con detección inteligente de anomalías, se deberá de implementar un sistema de envío de registros desde los dispositivos al sistema de almacenamiento con una periodicidad determinada en el diseño del sistema de vigilancia.

De esta forma, podemos dividir la arquitectura del Sistema de Vigilancia en tres bloques:

— *Sistema de envío de registros*: realizará el envío de los indicadores del sistema desde el dispositivo en el que se lleve a cabo el procesamiento. Este dispositivo puede ser un servidor, un sistema IoT o cualquier sistema en el que se lleve a cabo el procesamiento del sistema inteligente.

— *Sistema de almacenamiento de registros*: almacenará todos los indicadores generados por parte de los sistemas inteligentes.

— *Panel de Vigilancia y Sistema de Alertas*: procesará la información recibida ofreciendo accionabilidad y accesibilidad a los responsables de la vigilancia.

3. EL CONCEPTO DE INDICADOR

Previamente se ha hecho alusión al término indicador pero no se ha profundizado en su significado ni en su aplicación práctica en el sistema de vigilancia. ¿Cuál es la importancia de los indicadores en esta labor? Vamos a verlo con un ejemplo práctico.

Si nos dicen que vigilemos la temperatura de un coche porque durante los últimos días ha funcionado mal, ¿qué es lo que preguntaremos? Probablemente, pediríamos que nos dijeran cuánto es la temperatura normal y a partir de cuánta temperatura se considera un problema, tanto por temperatura alta como por temperatura baja. Con esos datos ya tendríamos una escala y podríamos vigilar que la temperatura del coche se mantuviera estable.

El concepto de indicador es exactamente el anterior. Podemos definirlo como un dato enmarcado dentro de una escala que nos permite inferir que nos acercamos a un escenario concreto o estamos saliendo del escenario que tratamos de medir. En nuestro ejemplo, el dato es la temperatura, la escala son los límites que nos ha dicho el mecánico y los escenarios son el funcionamiento normal del vehículo o el estado averiado del mismo.

En el contexto de los sistemas inteligentes, los indicadores se basarán en datos sobre el funcionamiento del sistema, por ejemplo, en el caso de la cámara de vigilancia, número de predicciones o de imágenes procesadas por minuto. Para cada dato deberemos de establecer su escala de normalidad y definir qué escenarios tratamos de controlar. Por ejemplo, una reducción drástica en el número de imágenes procesadas podría indicar una sobrecarga del dispositivo y conllevar un mal funcionamiento.

No existe una lista concreta de indicadores para nuestros sistemas inteligentes pero deberemos de crearla junto con el equipo técnico para tratar de monitorizar cómo de cerca nos encontramos de los riesgos detectados en el análisis de riesgos desarrollado y de un mal funcionamiento general del sistema.

Por supuesto, no todos los registros se podrán considerar indicadores. Existen registros logs que indicarán errores del sistema previamente desconocidos en formato

de texto que deberán de ser evaluados por el equipo técnico. No obstante, es vital contar con un esquema de indicadores lo más completo posible desde la perspectiva del sistema inteligente, el dispositivo que lo soporta, el uso por parte de los usuarios finales y los riesgos de ciberseguridad destacados.

Finalmente, el de vigilancia y el sistema de alertas emplearán dichos indicadores como base para hacer accionable todo el sistema de vigilancia en conjunto.

4. MEDIDAS A DESARROLLAR EN EL PLAN DE VIGILANCIA

Aún no contamos con el modelo ni los requisitos del plan de vigilancia poscomercialización. Dicho lo anterior, sí que existen ciertos pilares básicos que se deben de introducir en todo Plan de Vigilancia y que con alta probabilidad serán requisitos en el modelo ofrecido por la Comisión. En concreto:

— *Vigilancia Continua*: se trata de una monitorización con una periodicidad muy reducida en el tiempo (minutos u horas en función del sistema) que permitirá a los responsables detectar cambios abruptos en el funcionamiento del sistema inteligente. El sistema de alertas permitirá que esta labor se lleve a cabo de forma reactiva en el menor tiempo posible de respuesta si acontece un escenario anómalo. Sin embargo, en muchos casos, este sistema de alerta temprana no debe de ser la única labor de vigilancia y se debe de desarrollar una verificación continua de que todos los indicadores se mantienen estables.

— *Vigilancia Periódica*: en este caso, en lugar de una monitorización en tiempo real, esta vigilancia se realizará con una periodicidad más dilatada en el tiempo (días, semanas o meses) con el objetivo de evaluar cambios paliativos y dilatados en el tiempo que pasen desapercibidos en una monitorización en tiempo real. Por ejemplo, si evaluamos la precisión de un sistema inteligente hora tras hora puede que no se observe ningún cambio pero si desarrollamos una evaluación semanal podremos observar si se ha producido algún cambio significativo.

— *Asignación de responsables*: el Plan deberá asignar las personas responsables de llevar a cabo las actividades de vigilancia y el mantenimiento del sistema de vigilancia.

— *Formación*: Las personas seleccionadas como responsables deberán de conocer acerca del funcionamiento del panel de vigilancia, el sistema de alertas, los indicadores del sistema y los protocolos de notificación establecidos.

— *Protocolo de notificación de incidentes*: resulta de vital importancia desarrollar un protocolo de comunicación y registro de los incidentes detectados en el sistema. En el caso de ser incidentes graves, este protocolo deberá de complementarse con las medidas establecidas en el RIA para estos casos como la notificación a la Autoridad de Vigilancia.

Como ya se ha comentado, aún no se conoce el modelo y requisitos oficiales del Plan de Vigilancia pero con alta probabilidad los puntos explicados formarán parte de los seleccionados.

5. VALIDEZ DEL SISTEMA Y EL PLAN DE VIGILANCIA

Una vez desarrollado el Sistema y el Plan de Vigilancia cabe preguntarse hasta cuándo resultará válido dicho sistema para la monitorización correcta del sistema

inteligente. La respuesta es directa: siempre que el sistema inteligente no sufra ninguna modificación que altere el Plan o el Sistema de Vigilancia o se realice un nuevo análisis de riesgos que exponga nuevos escenarios de riesgo que deban de contar con nuevos indicadores.

III. CONCLUSIONES

La vigilancia postcomercialización es un componente indispensable en el desarrollo y despliegue de sistemas de IA de alto riesgo, ya que asegura que estos sistemas sean seguros, eficaces y éticamente responsables a lo largo de su ciclo de vida. Este proceso requiere una colaboración estrecha entre proveedores, usuarios y reguladores, y debe ser visto como una oportunidad para mejorar continuamente la tecnología de IA, promover su aceptación social y fomentar una innovación responsable en el campo. Así, a raíz del análisis realizado, detallamos las siguientes conclusiones:

La vigilancia postcomercialización es esencial para monitorizar y mantener la seguridad, la eficacia y el cumplimiento ético y regulatorio de los sistemas de IA de alto riesgo a lo largo de su ciclo de vida. Este proceso continuo ayuda a identificar y mitigar problemas emergentes que podrían no ser evidentes en las fases de diseño y prueba.

La principal responsabilidad de implementar sistemas de vigilancia robustos recae sobre los proveedores de sistemas de IA de alto riesgo. Esto incluye la monitorización constante del rendimiento del sistema, la adaptación a cambios en el entorno operativo y la respuesta a los desafíos éticos y legales que surgen durante el uso del sistema. Por otro lado, los usuarios no finales (o implementadores) de sistemas de IA de alto riesgo también desempeñan un papel importante en la vigilancia postcomercialización. Estos proporcionan retroalimentación sobre el rendimiento del sistema y reportar cualquier incidente o anomalía. Esta colaboración es fundamental para la detección temprana de problemas y la implementación oportuna de soluciones.

Uno de los mayores retos de la vigilancia postcomercialización es la capacidad de los sistemas de IA para adaptarse a entornos operativos en constante cambio y a conjuntos de datos que evolucionan. Esto requiere mecanismos de vigilancia flexibles y dinámicos que puedan ajustarse a nuevas condiciones y desafíos.

La vigilancia postcomercialización juega un papel crucial en garantizar que los sistemas de IA cumplan continuamente con las regulaciones y normas éticas en evolución. Esta cuestión no solo protege a los usuarios y a la sociedad, sino que también asegura la confianza y aceptación de la IA.

En conclusión, la vigilancia postcomercialización es un componente indispensable en el desarrollo y despliegue de sistemas de IA de alto riesgo, asegurando que estos sistemas sean seguros, eficaces y éticamente responsables a lo largo de su ciclo de vida. Este proceso requiere una colaboración estrecha entre proveedores, usuarios y reguladores, y debe ser visto como una oportunidad para mejorar continuamente la tecnología de IA, promover su aceptación social y fomentar una innovación responsable en el campo.

**INTELIGENCIA ARTIFICIAL DE USO GENERAL,
SISTEMAS QUE NO SON DE ALTO RIESGO Y LOS
SISTEMAS DEL ARTÍCULO 50**

Inteligencia artificial de uso general, modelos fundacionales (y «Chat GPT») en el Reglamento de inteligencia artificial

JOSÉ ANTONIO CASTILLO PARRILLA¹

Investigador Ramón y Cajal — Universidad de Granada

I. INTRODUCCIÓN

La IA se encuentra plenamente inserta en la vida cotidiana²: asistentes virtuales de agenda, traductores, generadores de subtítulos de video, herramientas en plataformas para sugerencias de contenido, y un larguísimo etcétera de ejemplos. La Comisión Europea afirma que la IA se refiere a sistemas que muestran un comportamiento inteligente al analizar su entorno y tomar acciones (con cierto grado de autonomía) para lograr objetivos específicos³. Los sistemas basados en IA pueden estarlo puramente en software (ej, sistemas de recomendación o motores de búsqueda), o integrarse en hardware (robots, drones, o aplicaciones IoT). En 2019 el Grupo de Expertos de Alto Nivel en IA de la Comisión Europea definió la IA como sistemas de software (o hardware) diseñados por humanos que, dado un objetivo complejo, actúan en la dimensión física o digital percibiendo su entorno mediante la adquisición de datos, interpretando los datos recogidos, estructurados o no, razonando sobre el conocimiento, o procesando la información, derivada de estos datos y decidiendo la(s) mejor(es) acción(es) a tomar para alcanzar el objetivo dado⁴.

1. Esta publicación es parte del contrato RYC2021-031430-I, financiado por MCIN/AEI/10.13039/501100011033 y por la Unión Europea, «NextGenerationEU». Se desarrolla en el marco del Proyecto GOIA, financiado por MCIN/AEI/TED2021-12902B-C22 y por la Unión Europea, «NextGenerationEU». Gran parte de la investigación se realizó durante la estancia desarrollada en la Universidade Nova de Lisboa durante los meses de marzo a mayo de 2024, financiada por el Plan Propio de la Universidad de Granada.
2. https://ec.europa.eu/commission/presscorner/detail/en/statement_23_6474
3. Comisión, Comunicación «Artificial Intelligence for Europe», Bruselas, 25 de abril de 2018, COM (2018) 237 final, <https://digital-strategy.ec.europa.eu/en/library/communication-artificial-intelligence-europe>, p. 1.
4. HLEG-AI — High Level Expert Group on Artificial Intelligence, (2019): A definition of AI: main capabilities and disciplines, Comisión Europea, Bruselas, abril, <https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines>

La relevancia de los modelos y sistemas IA de uso general ha ido creciendo considerablemente en los últimos años, especialmente desde la eclosión de Chat GPT a finales de 2022, habiendo provocado un auténtico terremoto⁵. Chat GPT es una herramienta de IA generativa de texto (*large language model* o LLM) que se lanzó el 30 de noviembre de 2022⁶, inicialmente de manera gratuita, y en apenas cinco días superó el millón de usuarios, y superando los 180 millones de usuarios activos en noviembre del año siguiente⁷. El 14 de marzo de 2024 OpenAI lanza GPT-4⁸, que en menos de seis meses superó los cien millones de usuarios activos semanales⁹. Chat GPT no es la única herramienta de IA generativa¹⁰, pero sí una de las que goza de mayor popularidad actualmente, habiendo suscitado un intenso debate acerca de los riesgos que este tipo de herramientas suponen; hasta el punto de que por ejemplo el Comité Europeo de Protección de Datos inició en abril de 2023 un grupo de trabajo sobre Chat GPT¹¹ tras la decisión de algunas autoridades nacionales de protección de datos como la AEPD¹² o el Garante Privacy (Italia)¹³ de iniciar de oficio actuaciones de investigación¹⁴ por posible incumplimiento de la normativa de protección de datos.

El impacto social y la popularidad de esta herramienta ha sido tal que si la Propuesta de RIA de 2021¹⁵ no mencionaba modelos fundacionales o IA de uso general, apenas tres meses después del lanzamiento de GPT-4 las Enmiendas al texto presentadas por el Parlamento Europeo¹⁶ dedicaron varios nuevos Considerandos y artículos a los modelos fundacionales; y el texto finalmente presentado y aprobado

5. Novelli, C. y otros, «Generative AI in EU Law: Liability, Privacy, Intellectual Property and Cybersecurity», Cornell University, <https://arxiv.org/abs/2401.07348>
6. <https://openai.com/blog/chatgpt>, p. 1.
7. <https://www.primeweb.com.mx/chatgpt-usuarios-estadisticas> De entre los datos destacados llama la atención que cerca del 80% de jóvenes de entre 18 y 29 años han utilizado o han visto utilizar a alguien la herramienta.
8. <https://openai.com/research/gpt-4>
9. <https://www.theverge.com/2023/11/6/23948386/chatgpt-active-user-count-openai-developer-conference>
10. Otras herramientas de IA generativa se han desarrollado en el ámbito del llamado «arte de inteligencia artificial», como Stable Difussion, Midjourney o DALL-E. En IA generativa de texto Chat GPT no es tampoco única: Microsoft ha lanzado Copilot, y Nvidia RTX.
11. https://www.edpb.europa.eu/news/news/2023/edpb-resolves-dispute-transfers-meta-and-creates-task-force-chat-gpt_en
12. <https://www.aepd.es/prensa-y-comunicacion/notas-de-prensa/aepd-inicia-de-oficio-actuaciones-de-investigacion-a-openai>
13. <https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9870847>
14. La Comisión Nacional de Protección de Datos (CNPD, Portugal) únicamente manifestó su interés por esta cuestión pero no inició investigación de oficio sobre Chat GPT: <https://observador.pt/2023/04/03/chatgpt-cnpd-leu-com-muito-interesse-decisao-de-bloqueio-em-italia-mas-nao-preve-para-ja-algo-semelhante-em-portugal/>
15. <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex%3A52021PC0206>
16. Parlamento Europeo (2023): Enmiendas aprobadas sobre la Propuesta de Reglamento por el que se establecen normas armonizadas en materia de inteligencia artificial (COM (2021) 0206), https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_ES.html

como RIA¹⁷ dedica un Capítulo completo a la regulación de los modelos IA de uso general, amén de diversas menciones y obligaciones específicas en otros apartados del RIA.

También da cuenta de la importancia del tema el hecho de que la inclusión (y el alcance) de la IA de uso general en el RIA fuese uno de los últimos aspectos de debate en la negociación final del texto. En noviembre de 2023 se hicieron públicas entre diversos Estados Miembros en relación con los entonces llamados modelos fundacionales durante la fase de diálogos tripartitos, a pesar de existir consenso acerca de la necesidad de incluir ciertas normas de transparencia: Alemania, Francia o Italia eran partidarias de que se favoreciese la elaboración de códigos de conducta, pero sin un régimen de sanciones ya fijado por el RIA¹⁸, mientras que España defendía la inclusión de obligaciones más allá de la transparencia, e incluso de abordar el reto de los derechos de autor¹⁹.

II. INTELIGENCIA ARTIFICIAL GENERAL, DE USO GENERAL, MODELOS FUNDACIONALES E INTELIGENCIA ARTIFICIAL GENERATIVA

La primera gran clasificación en que se dividen las herramientas de IA es la que distingue aquellas basadas en reglas lógicas, de las basadas en datos²⁰. Las primeras, también conocidas como sistemas expertos, son capaces de desarrollar muy bien tareas en campos delimitados y relativamente sencillos a partir de la incorporación de reglas lógicas y del conocimiento de expertos (que, nuevamente, diseñan reglas lógicas que la máquina incorpora). El ejemplo más popular a día de hoy quizás siga siendo Deep Blue²¹. Las herramientas IA basadas en datos son alimentadas con grandes cantidades de datos que les permiten, a través de diversas técnicas (*machine learning*, redes neuronales, *deep learning*, árboles de decisión...) resolver problemas inespecíficos o cuya solución no puede alcanzarse a través de razonamiento deductivo (análisis de textos o imágenes, predicción de comportamiento, o sistemas de recomendación²²). Si bien los sistemas IA basados en datos tienen siempre un cierto porcentaje de error, éste se irá reduciendo a medida que son capaces de obtener y procesar mayores cantidades de datos²³.

17. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_ES.pdf

18. <https://www.euractiv.com/section/artificial-intelligence/news/france-germany-italy-push-for-mandatory-self-regulation-for-foundation-models-in-eus-ai-law/>

19. <https://www.euractiv.com/section/artificial-intelligence/interview/eu-ai-act-cannot-turn-away-from-foundation-models-spains-state-secretary-says/>

20. Valls Prieto, J., *Inteligencia artificial, derechos humanos y bienes jurídicos*, Aranzadi, Navarra, 2021, p. 20.

21. La herramienta Deep Blue de IBM derrotó en 1997 al campeón mundial de ajedrez del momento Gary Kasparov en el segundo match (a 6 partidas), que se jugó en ese año, tras un primero que tuvo lugar en 1996 y del que Deep Blue «aprendió» (<https://www.ibm.com/history/deep-blue>).

22. Los sistemas de recomendación (por ejemplo, utilizados en plataformas o técnicas de marketing online en redes sociales y motores de búsqueda) se basan en una técnica llamada *reinforcement learning* (aprendizaje por refuerzo): se permite al sistema IA tomar decisiones libremente, y se le premia cuando acierta. El objetivo del sistema IA es maximizar las recompensas.

23. HLEG-AI, op. cit., pp. 3-4.

En el campo de la IA suele distinguirse entre IA estrecha o débil e IA general o fuerte. Se califica de IA general o fuerte un sistema IA capaz de desarrollar actividades típicamente humanas; mientras que se entiende como IA estrecha o débil aquella capaz de desarrollar una o algunas tareas específicas. Si bien la mayor parte de los sistemas IA desarrollados hasta 2019 podrían ser calificados como IA estrecha o débil²⁴, los últimos avances y su popularización parecen justificar que el RIA haya decidido regular la que llama IA de uso general. La expresión IA general, por tanto, no debe identificarse con los términos que emplea el RIA (modelo IA de uso general / sistema IA de uso general), sino meramente como «IA fuerte».

En los textos que se han sucedido entre junio de 2023 y marzo de 2024 han tenido lugar algunos matices que conviene aclarar en este punto. El cambio de terminología más relevante, sin embargo, es la sustitución de «modelos fundacionales» por «modelo IA de uso (o propósito) general». Estos cambios en la terminología deben tenerse presentes no sólo para analizar cuánto de las propuestas del Parlamento Europeo ha quedado integrado en el texto del RIA; sino, por lo que se refiere a España, porque el RD 817/2023, de 8 de noviembre, siguió la terminología del Parlamento Europeo y habla por tanto de modelos fundacionales y sistemas IA de propósito general²⁵.

Para evitar complicaciones innecesarias, en adelante hablaremos simplemente de modelos IA de uso general siguiendo la terminología del RIA. Los modelos IA de uso general son modelos IA (que pueden estar entrenados con una gran cantidad de datos utilizando autosupervisión a gran escala), que presentan un grado considerable de generalidad y son capaces de realizar de manera competente una gran variedad de tareas distintas, así como de integrarse en diversos sistemas o aplicaciones posteriores (art. 3.63 RIA).

La IA generativa, por último, es un tipo de IA basada en modelos de IA de uso general capaz de generar de manera flexible contenidos de texto, audio, imágenes o video. Se trata, por lo tanto, de una categoría concreta de modelos de IA de uso general²⁶. La IA generativa plantea relativos a la generación de nuevos contenidos y el respeto a los derechos de propiedad intelectual²⁷: los modelos de IA generativa siguen el mismo funcionamiento que los modelos de IA general, es decir, se nutren de una gran cantidad de información de entrada, que en este caso puede estar protegida en su mayoría por normas de propiedad intelectual, con la consiguiente necesidad de recabar el oportuno consentimiento en su caso (Cons. 105 RIA).

24. HLEG-AI, op. cit., p. 5.

25. Real Decreto 817/2023, de 8 de noviembre, que establece un entorno controlado de pruebas para el ensayo del cumplimiento de la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial, cfr. arts. 3.6 y 3.5. Disponible en: https://www.boe.es/diario_boe/txt.php?id=BOE-A-2023-22767

26. Cons. 99 RIA, y Cons. 111 y 51 RIA respecto de los modelos IA de uso general con riesgo sistémico.

27. Cfr. Enmienda 102 del Parlamento Europeo.

III. ¿QUÉ ES Y QUÉ NO ES INTELIGENCIA ARTIFICIAL DE USO GENERAL EN EL REGLAMENTO? MODELOS DE USO GENERAL CON RIESGO SISTÉMICO Y EXCLUSIÓN DE LOS MODELOS ESPECÍFICAMENTE DESTINADOS A INVESTIGACIÓN

Aunque los modelos IA son componentes esenciales de los sistemas IA, no constituyen sistemas por sí mismos, ya que necesitan de otros componentes (como por ejemplo, una interfaz de usuario) para convertirse en sistemas IA (Cons. 97). Por lo tanto, entre modelos y sistemas IA existe una relación de todo/parte, donde el sistema IA es el todo y los modelos IA pueden ser una parte. Dentro de los sistemas IA basados en datos (esto es, aquellos que para lograr sus objetivos utilizan inferencias basadas en información de entrada para producir diversos productos como información de salida *ex art.* 3.1 RIA), se consideran sistemas IA de uso general aquellos que están basados en un modelo IA de uso general (art. 3.66 RIA).

Dos características fundamentales definen un modelo IA como de uso general: (1) la generalidad, y (2) la capacidad de realizar de manera competente una amplia variedad de tareas diferenciadas. ¿Qué debe entenderse por generalidad? El RIA no la define en el art. 3, pero sí proporciona un criterio de referencia: serán generales aquellos modelos que: (1) tengan al menos mil millones de parámetros, y (2) se hayan entrenado con un gran volumen de datos utilizando autosupervisión a gran escala (Cons. 97).

Los modelos de IA generativa son una subcategoría dentro de los modelos de IA de uso general (Cons. 99). También pueden entenderse como subcategoría los modelos IA de uso general con riesgos sistémicos (Cons. 111 y art. 51 RIA).

Debe considerarse que un sistema IA de uso general presenta riesgos sistémicos cuando: (1) tenga capacidades de gran impacto, o (2) unas repercusiones considerables en el mercado interior debido a su alcance. Dentro de la segunda posibilidad no parece que se incluyan contenidos encaminados a desinformación, ya que esto se reconduce fundamentalmente a meras obligaciones de transparencia²⁸. Son capacidades de gran impacto aquellas que igualan o superan las mostradas por los modelos IA de uso general más avanzados, lo cual dependerá del estado de la técnica de cada momento. Un criterio de referencia importante para determinar si un modelo IA de uso general tiene capacidades de gran impacto es el umbral de FLOPS, esto es, la cantidad acumulada de cálculos utilizados para el entrenamiento del modelo IA de uso general, medida en operaciones de coma flotante (art. 3.67 RIA). Este umbral deberá ir ajustándose en función de la evolución tecnológica, de manera que refleje los cambios tecnológicos e industriales, las mejoras algorítmicas o el aumento de la eficiencia del hardware (Cons. 111).

El Anexo XIII del RIA establece criterios de referencia a tener en cuenta para valorar si un modelo IA de uso general tiene capacidades de gran impacto (art. 51.1.b RIA). Estos criterios son: (1) el número de parámetros del modelo; (2) la calidad y el tamaño del conjunto de datos; (3) la cantidad de cálculo utilizada para entrenar el modelo, medida en FLOP o en combinación con otras variables; (4) las modalidades de entrada y salida del modelo (texto a texto, texto a imagen...); (5) los parámetros de referencia y las evaluaciones de las capacidades del modelo;

28. Watcher et al., op. cit., p. 41.

(6) la importancia de sus repercusiones debido a su alcance; y (7) el número de usuarios finales registrados.

Se trata de criterios abiertos: se establece una lista ejemplificativa, y respecto de cada uno de los criterios no se introducen umbrales mínimos, salvo en dos: el umbral de FLOPS (en art. 51.2 RIA), y el alcance del modelo, que se dará por supuesto cuando se haya puesto al alcance de al menos 10000 usuarios profesionales registrados establecidos en la UE²⁹. Este sistema es coherente con el hecho de que será la Comisión la que designe un modelo IA de uso general como modelo con riesgo sistémico (art. 52.1 RIA). Por otra parte, la norma prevé su propia capacidad de adaptación de manera ágil en el art. 52.3 RIA, que faculta a la Comisión para adoptar actos delegados para modificar los criterios y los umbrales referidos en función de los avances tecnológicos (art. 97 RIA y 290 TFUE). Todo ello redundará en beneficio de la seguridad jurídica de los proveedores de sistemas de IA, y particularmente por lo que a este punto se refiere, de los proveedores de modelos IA de uso general³⁰.

La seguridad jurídica se ve reforzada también en el procedimiento para la consideración de un modelo de IA de uso general como modelo con riesgo sistémico (art. 52 RIA). Podríamos decir que se trata de un procedimiento «en dos capas»: (1) responsabilidad proactiva del proveedor; y (2) actuaciones de revisión y cierre por parte de la Comisión. El procedimiento es acertado en la medida en que no sólo favorece la seguridad jurídica del tráfico de estos productos, sino que deposita una parte importante de la misma en quienes se benefician económicamente de los modelos, dejando a las instituciones (la Comisión en este caso) un papel de control y cierre del sistema.

La primera fase (art. 52.1 RIA) descansa fundamentalmente en el proveedor del modelo IA de uso general, que debe notificar a la Comisión que ha superado o prevé superar los requisitos que lo califican como modelo con riesgo sistémico, aportando la documentación necesaria. Asimismo, puede aportar junto a esta documentación argumentos que avalarían la ausencia de riesgos sistémicos en el caso concreto debido a las características específicas del modelo IA (art. 52.2 RIA), que la Comisión valorará a efectos de, finalmente, su designación o no como modelo IA de uso general con riesgo sistémico (art. 52.3 RIA)³¹. El precepto debe interpretarse entendiendo que el

29. Anexo XIII, apdo. f. Con todo, si atendemos a la literalidad de este apartado, podemos ver que se trata igualmente de un criterio abierto: se entiende superado el alcance cuando se llega a los 10000 usuarios profesionales registrados establecidos en la UE; pero nada impide considerar que el modelo tiene un alcance significativo con una cifra inferior, más aun teniendo en cuenta que éste es sólo uno de los criterios que habrá de tener en cuenta la Comisión para designar un modelo IA de uso general como modelo con riesgo sistémico.

30. También redundará en beneficio de la seguridad jurídica, y de la transparencia, la previsión de publicación de una lista actualizada de modelos IA de uso general con riesgo sistémico por parte de la Comisión, siempre aportando información suficiente sobre los mismos pero sin poner en peligro derechos de propiedad intelectual, industrial y secretos comerciales de los modelos (art. 52.6 RIA).

31. Este proceso puede repetirse en el futuro (nunca antes de 6 meses desde su designación) si el proveedor del modelo ya designado como modelo IA de uso general con riesgo sistémico solicita reevaluación por la Comisión, siempre que aporte razones nuevas desde su designación que justificasen un cambio (art. 52.5 RIA).

proveedor está obligado a notificar a la Comisión en el plazo de dos semanas desde el día siguiente al que ha cumplido cierto requisito fácilmente contrastable por sí mismo: superar el umbral FLOP o el número de usuarios del Anexo XIII.f que en cada momento estén vigentes, u otros umbrales de similar claridad que se estableciesen en el futuro³². Transcurrido este plazo entraríamos en la segunda fase, en la que la Comisión puede revisar y en su caso completar la tarea de responsabilidad proactiva de los proveedores de modelos IA de uso general.

La que hemos llamado fase de revisión y cierre por parte de la Comisión se traduce en varias vías por las que la Comisión puede determinar que un modelo IA de uso general presenta riesgos sistémicos: (1) si entiende que proveedor no ha sido capaz de demostrar la ausencia de riesgos una vez cumplidos los parámetros objetivos que le obligan a notificar (art. 52.3 RIA); (2) si lo designa de oficio cuando tiene conocimiento de que presenta riesgos sistémicos y no lo ha notificado (art. 52.1 y Cons. 113 RIA); o (3) si lo designa de oficio a raíz de una alerta cualificada por parte de grupo de expertos científicos independientes, cuando se cumplan los requisitos del Anexo XIII (art. 52.4 y 90.1.b RIA).

La seguridad jurídica de este procedimiento de designación se cierra con una consideración, quizás no del todo explícita en el articulado: un modelo de IA de uso general debe considerarse como de riesgo sistémico a partir de (1) que la Comisión recibe la notificación por parte del proveedor, o (2) de que designa al modelo como modelo con riesgo sistémico de acuerdo con las tres vías que acabamos de ver. No aclara la norma el valor del silencio. Puede asumirse que el proveedor debe considerarse como proveedor de modelo IA de uso general con riesgo sistémico: (1) desde que notifica a la Comisión haber superado los umbrales objetivos sin aportar argumentos que cuestionen la calificación; (2) desde que notifica a la Comisión aportando argumentos que cuestionen la calificación y la Comisión no contesta, o contesta negativamente; y (3) desde que la Comisión le notifica su designación. La calificación de un modelo de IA de uso general como modelo con riesgo sistémico comporta una serie de obligaciones detalladas en el art. 55 RIA, por lo que resulta exigible un mínimo de seguridad jurídica para el proveedor, que se vería colmado en menor medida si la exigibilidad de estas obligaciones no estuviese asociada a

32. La redacción del art. 52.1 RIA es mejorable por varias razones. En primer lugar, el art. 51.1.a RIA no habla de «un requisito», sino de varios, la mayoría de los cuales no están asociados a umbrales objetivos salvo dos: FLOP y alcance del modelo medido en número de usuarios profesionales activos registrados en la UE. Hay que entender, por tanto, que cuando un modelo IA de uso general supere el umbral FLOP del art. 51.2 o el umbral del Anexo XIII.f, o los que en el futuro los sustituyesen, debe notificar a la Comisión esta circunstancia. El segundo aspecto de mejora en la claridad de la redacción del art. 52.1 es el referido al tiempo: tanto en la determinación del día a quo como en la determinación del plazo para notificar a la Comisión: el proveedor está obligado a notificar «sin demora y, en cualquier caso, antes de transcurridas dos semanas desde que se cumpla dicho requisito o se sepa que va a cumplirse». El día a quo, por tanto, puede ser el momento en que se cumple con cierto requisito, o el momento en que se sabe (¿cómo?) que se va a cumplir. El plazo admite tres alternativas: (1) sin demora (¿qué significa?), (2) dos semanas desde que se sabe que se va a cumplir cierto requisito, o (3) dos semanas desde que ya se ha cumplido cierto requisito. Semejante indeterminación no favorece la seguridad jurídica que el engranaje de los arts. 51 y 52, y el Anexo XIII propician.

algún tipo de comunicación con el proveedor como las que se han señalado. Entre estas obligaciones no figura, como tampoco ocurre con carácter general, la de que el modelo produzca resultados fiables³³.

Interesa destacar de la definición del RIA no sólo qué son los modelos IA de uso general, sino también qué no son: no se consideran modelos IA de uso general los que «se utilizan para actividades de investigación, desarrollo o creación de prototipos antes de su comercialización» (art. 3.63 RIA *in fine*). Se pretende de esta forma que el RIA no socave las actividades de investigación y desarrollo, en coherencia con su declaración de apoyo a la innovación y respeto a la libertad de ciencia (Cons. 25). Esto implica que quedan excluidos del ámbito de aplicación del RIA aquellos sistemas y modelos IA desarrollados *específicamente* y puestos en servicio únicamente con fines de investigación y desarrollo científicos (Cons. 25 RIA)³⁴. Debe distinguirse si los fines de investigación y desarrollo científico son el único uso posible del sistema o modelo IA o si éste puede utilizarse, entre otras cosas, para actividades de investigación y desarrollo científicos. En el segundo caso el RIA no se aplicará antes de su introducción en el mercado, pero una vez introducido en el mercado el contenido del RIA le resultará plenamente aplicable (art. 2.8 RIA).

Se entiende por introducción en el mercado «la primera comercialización en el mercado de la Unión de un sistema de IA o de un modelo de IA de uso general» (art. 3.9 RIA³⁵). Asimismo, cuando el proveedor de un modelo de IA de uso general integre un modelo propio en un sistema de IA propio que se comercialice o se ponga en servicio, debe entenderse que se ha introducido en el mercado (Cons. 97). No se considerarán introducción en el mercado las pruebas en condiciones reales que cumplan los requisitos establecidos en los artículos 57 o 60 (art. 3.57 RIA)³⁶. Aquellos modelos IA de uso general que ya operen en el mercado de la UE durante el primer año desde que entre en vigor el RIA tendrán 36 meses desde la entrada su entrada en vigor para acomodarse a sus requisitos (art. 111.3 RIA).

IV. ALGUNOS RETOS NORMATIVOS DE LA INTELIGENCIA ARTIFICIAL GENERATIVA

El uso de IA generativa plantea supone un avance indudable, pero también presenta riesgos en diversos ámbitos como la responsabilidad civil derivada del uso de la IA, el derecho a recibir una información veraz, la propiedad intelectual, la protección de datos.

La responsabilidad civil derivada del uso de la IA debe observarse a la luz de dos Directivas que aún se encuentran en fase de propuesta: la nueva Directiva de responsabilidad por productos defectuosos, y la Directiva sobre responsabilidad por

33. Watcher et al., op. cit., p. 40.

34. Énfasis añadido.

35. Cfr. también art. 3.2 Reglamento UE 2019/1020.

36. Los artículos 57 y 60 forman parte del Capítulo VI (arts. 57 a 63), destinado a medidas de apoyo a la innovación. El art. 57 establece requisitos para espacios controlados de pruebas para IA, y el art. 60 requisitos para pruebas de sistemas IA de alto riesgo en condiciones reales fuera de los espacios controlados.

IA. En la medida en que los aspectos de responsabilidad civil derivada del uso de IA quedan fuera del RIA no nos ocuparemos de este aspecto³⁷.

Por lo que se refiere al derecho a recibir una información veraz, el RIA destaca que los modelos de IA de uso general pueden plantear riesgos sistémicos, como la desinformación (Cons. 110 RIA), lo cual puede poner en riesgo procesos democráticos y electorales (Cons. 120 RIA), amén de manipulación a gran escala, fraude, suplantación de identidad y engaño a consumidores (Cons. 133 RIA). El riesgo de desinformación aumenta con las ultrafalsificaciones: imágenes, audio o vídeo generados o manipulados por IA que se asemejan a personas, objetos, lugares u otras entidades o sucesos reales y que pueden inducir a una persona a pensar erróneamente que son auténticos o verídicos (art. 3.60 RIA)³⁸. La UE lleva al menos desde 2017³⁹ desarrollando estrategias contra la desinformación que han culminado hasta la fecha en un Código de buenas prácticas reforzado sobre desinformación en 2022⁴⁰. El tratamiento de esta problemática va más allá del RIA y de los modelos IA de uso general por lo que, como en el caso anterior, simplemente lo apuntamos.

Los retos relacionados con la propiedad intelectual, podemos destacar como retos no sólo los eventuales impedimentos de reutilización de material por parte de modelos IA de uso general (a los que el RIA les presta alguna atención, como veremos seguidamente), sino también la necesaria distinción entre un uso del modelo IA como mera herramienta, y un uso creativo de éste⁴¹. En cuanto a la reutilización de material previo, al margen de las disposiciones al respecto en el RIA, no cabe esperar que estos casos se judicialicen al menos por lo que se refiere a las herramientas de IA generativa de grandes compañías, que ya se han adelantado anunciando compensaciones económicas por las posibles infracciones que se hubieran cometido⁴².

En cuanto a la primera cuestión, tanto el RIA como la Directiva 2019/790 ofrecen algunas respuestas: a los resultados producidos por modelos IA de uso general se les aplica una lógica normativa similar a la de las obras derivadas (Cons. 105 RIA), lo que lleva a establecer una serie de obligaciones para los proveedores de modelos IA de uso general en coordinación con el art. 4.3 de la Directiva UE 2019/790. El art. 4 de la Directiva UE 2019/790 mandata a los Estados

37. Lo han tratado en España con carácter general Muñoz García, C. (Regulación de la inteligencia artificial en Europa: incidencia en los regímenes jurídicos de protección de datos y de responsabilidad por productos, Tirant lo Blanch, Valencia, 2023), o Navas Navarro, S. (Daños ocasionados por sistemas de inteligencia artificial: especial atención a su futura regulación, Comares, Granada, 2022), y específicamente respecto de la IA Generativa Novelli et al., op. cit., pp. 2-7.

38. La definición fue introducida por la enmienda 203 del Parlamento Europeo a la Propuesta originaria de RIA, si bien el término ya se menciona desde 2021.

39. Comisión, Comunicación «Luchar contra la desinformación en línea: un enfoque europeo», Bruselas, 26 de abril de 2018, COM (2018) 236 final, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0236>

40. <https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation>

41. Novelli et al., op. cit., p. 14.

42. CMA — Competition & Markets Authority (2024): AI Foundation Models — Technical update report, https://assets.publishing.service.gov.uk/media/661e5a-4c7469198185bd3d62/AI_Foundation_Models_technical_update_report.pdf, p. 54.

miembros para establecer excepciones a ciertos derechos de autor⁴³ en relación con las reproducciones y extracciones de obras y otras prestaciones accesibles de forma legítima para fines de minería de textos y datos, siempre que el uso de las obras y otras prestaciones no esté reservado expresamente por los titulares de derechos de manera adecuada de acuerdo con el art. 4.1 y 3 de la Directiva UE 2019/790. En la Enmienda 399 del Parlamento Europeo el propuesto art. 28.ter contemplaba tres obligaciones para los proveedores de modelos fundacionales de IA generativa: (1) obligaciones de transparencia (informar a personas de que están interactuando con un sistema IA)⁴⁴; (2) diseñar y desarrollar el modelo de manera que se garanticen salvaguardas frente a la generación de contenidos que infrinjan normativa de propiedad intelectual; y (3) documentar y poner a disposición del público un resumen detallado de los datos de entrenamiento. Esta última obligación es aplicable con carácter general para modelos IA de uso general en virtud del art. 54 RIA.

La segunda cuestión (sobre cuándo un resultado producido con IA generativa debe considerarse susceptible de protección) no permite una solución *a priori* sino que necesitará de un examen caso por caso. Sin embargo, sí pueden ofrecerse criterios que por otra parte no son nuevos sino que conectan con la clásica distinción entre fotografía y mera fotografía del art. 128 TRLPI⁴⁵: hablaremos de obras de propiedad intelectual (o industrial) siempre que la herramienta IA sea utilizada como mero instrumento, de manera que sea posible reconocer una actividad humana genuina en las elecciones tanto de resultados como de instrucciones introducidas en la herramienta IA; cuestión que no se dará si el modelo IA de uso general trabaja de manera autónoma⁴⁶. Queda por determinar, en cada caso, qué constituye una actividad humana genuina y reconocible como tal.

Por lo que se refiere a los aspectos relativos al tratamiento de datos, debe recordarse en primer lugar que la IA generativa es una herramienta alimentada con grandes cantidades de datos, personales y no personales. Debe recordarse igualmente, la vis expansiva de la noción de dato personal y de tratamiento de datos personales⁴⁷. Esto comporta necesariamente que exista riesgo de infracción de la normativa de protección de datos a lo largo de toda la cadena de valor de los datos (desde la recogida y tratamiento de los datos hasta incluso los resultados obtenidos de su procesado) y da cuenta de la importancia de la protección de datos desde el diseño y por defecto que prevé el RGPD en su art. 25. Incluso puede inferirse información personal a partir de ingeniería inversa, lo cual lleva a la necesidad de plantear técnicas

43. Arts. 5.a y 7.1 de la Directiva 96/9/CE; 2 de la Directiva 2001/29/CE; 4.1.a y b de la Directiva 2009/24/CE; y 15 de la Directiva 790/2019/UE.

44. Watcher et al., op. cit., p. 40.

45. Sobre el particular, Bondía Román, F., «Los derechos sobre las fotografías y sus limitaciones», Anuario de Derecho civil, Tomo LIX, Fasc. III, julio-septiembre 2006, https://www.boe.es/biblioteca_juridica/anuarios_derecho/abrir_pdf.php?id=ANU-C-2006-30106501114, pp. 1065-1114.

46. Novelli et al., op. cit., pp. 18-19.

47. Romeo Casabona, C., «Datos personales (Comentario al artículo 4.1 RGPD)», en Troncoso Reigada, A. (dir.), Comentario al Reglamento General de Protección de Datos y a la Ley Orgánica de Protección de Datos Personales y Garantía de los Derechos Digitales, Aranzadi, Navarra, 2021, pp. 574.

de privacidad diferencial⁴⁸ que podrían incluso quedarse cortas en este contexto tecnológico.

Los problemas relativos al tratamiento de datos se refieren a (1) si y (2) cómo deben tratarse los datos personales en el entrenamiento de las herramientas de IA generativa. Pueden destacarse siete problemas relacionados con el tratamiento de datos⁴⁹: (1) la base de legitimación para el tratamiento de los datos en entrenamiento del modelo; (2) la base de legitimación para el tratamiento de datos en el caso de las instrucciones del modelo (prompts)⁵⁰; (3) requisitos de información; (4) problemas relativos a la inversión de modelos, fuga de datos y ejercicio del derecho de supresión; (5) decisiones automatizadas⁵¹; (6) protección de menores; y (7) respeto a los principios de limitación de finalidad y minimización de datos. Nos centraremos en el primero de ellos⁵².

En cuanto a la base de legitimación para el tratamiento de datos personales en el entrenamiento, debe recordarse que todo tratamiento de datos debe realizarse, como mínimo, conforme a una de las bases de legitimación recogidas en el art. 6 RGPD, incluso si se trata de empresas establecidas fuera de la UE pero que ofrecen servicios en la UE⁵³ y antes de la introducción en el mercado del modelo. Si bien el consentimiento resulta de partida la base de legitimación que mayor seguridad jurídica ofrece al responsable del tratamiento, no ocurre así en el caso de la IA generativa debido a los enormes costes que conllevaría para éste asegurarse de que todos los interesados hubieran emitido un consentimiento específico, libre, inequívoco e informado respecto de la enorme variedad de actividades de tratamiento de datos que van a tener lugar⁵⁴. Por ello parece más apropiado acudir al interés legítimo como base de legitimación, siempre que se supere el juicio de ponderación entre los intereses legítimos del responsable del tratamiento (el gestor de la herramienta IA)

48. La privacidad diferencial es un método matemático que permite una mejor protección de la privacidad gracias a la incorporación de ruido aleatorio suficiente en la información original. El resultado no pierde valor por aplicación de la ley de los grandes números, pero la introducción del ruido permite una negación plausible de que los datos de una persona concreta formen parte del conjunto de análisis (Dwork, C., «Differential privacy», en Bugliesi, M. y otros. (eds.), *Automata, Languages and Programming*, Springer, Berlín-Heidelberg, 2006, pp. 1-12).

49. Novelli, et al., op. cit., pp. 7-14.

50. En muchas ocasiones el uso de herramientas de IA generativa conlleva una suerte de diálogo con la herramienta, en el que se le puede «contar» información de otra persona sin su consentimiento; información que podrá utilizar para su propio entrenamiento.

51. A raíz de la amplia concepción de la noción de «decisión» defendida por el TJUE en el caso SCHUFA, cabría preguntarse si no estaremos en este caso también ante decisiones plenamente automatizadas del art. 22 RGPD (Novelli et al., op. cit., p. 12).

52. Para un examen detallado, Novelli, C. y otros, «Generative AI in EU Lawcit...» pp. 7-14. Por otra parte, en esta obra, Jiménez López, J., «Protección de datos y Reglamento de Inteligencia Artificial».

53. Art. 3.2.a RGPD.

54. Se han dado incluso situaciones en las que herramientas como Chat GPT 3.5 ha proporcionado listas de consentimiento explícito para el uso de datos de manera incorrecta (Watcher, S.; Mittlestadt, B.; Russell, C., «Do large language models have a legal duty to tell the truth?», *Royal Society Open Science*, mayo 2024, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4771884, p. 11, nota 88).

y los derechos y libertades fundamentales de los interesados; análisis que deberá tener lugar caso por caso, esto es, herramienta por herramienta⁵⁵. También se ha destacado la utilidad que puede suponer usar datos sintéticos, si bien por el momento las capacidades de uso de datos sintéticos a gran escala no lo permiten⁵⁶, amén de que si nos atenemos al art. 4.1 RGPD el origen sintético de los datos no evitaría su calificación como datos personal.

La cuestión se complica algo más si tenemos en cuenta que no sólo es expansivo el concepto de dato personal, sino también el concepto de categorías especiales de datos personales del art. 9.1 RGPD. Esto pudo observarse en la STJUE de 7 de julio de 2023, que entiende que incluso aquellos datos que permiten revelar información comprendida en alguna de las categorías especiales de datos del art. 9.1 RGPD son ya datos de categorías especiales independientemente de que la información revelada sea o no exacta⁵⁷. En el art. 9.2 RGPD, que recoge las excepciones a la prohibición general de tratamiento de categorías especiales de datos, no existe un equivalente al interés legítimo. Podrían explorarse excepciones como la relativa a actividades de investigación (art. 9.2.j RGPD) o la referida a datos que el interesado ha hecho manifiestamente públicos (art. 9.2.e RGPD)⁵⁸. Por lo que se refiere a la última, habrá de tenerse en cuenta si el interesado pretendió, de manera explícita y a través de una clara acción afirmativa, hacer públicos estos datos⁵⁹, y aún con todo ello las excepciones del art. 9.2 RGPD deben interpretarse de manera restrictiva. Por tanto: (1) los datos no deben referirse a personas distintas de aquella que los hizo públicos⁶⁰; (2) no podrá entenderse que la mera consulta de páginas web es un dato que el interesado hace manifiestamente público⁶¹; y (3) habrán de tenerse en cuenta las expectativas razonables del interesado (es decir, si podía esperar que sus datos servirían para entrenar modelos y herramientas de IA) en el momento de la recolección de sus datos⁶². En cuanto a las actividades de investigación, cabe destacar el apoyo declarado del RIA a la innovación, el respeto a la libertad de ciencia y su voluntad de no socavar la actividad de investigación y desarrollo (Cons. 25 RIA)⁶³. Sin embargo, sigue siendo el RGPD quien delimita qué es «investigación» respecto del tratamiento de datos. A pesar de que apuesta por un concepto amplio de investigación⁶⁴, excluye del mismo

55. Gil González, E.; De HerT, P., «Understanding the Legal Provisions That Allow Processing and Profiling of Personal Data-an Analysis of GDPR Provisions and Principles», ERA Forum 2019, vol. 4, 2019, <https://research.tilburguniversity.edu/en/publications/understanding-the-legal-provisions-that-allow-processing-and-prof>, pp. 618-619; Novelli et al., op. cit., p. 8.

56. CMA, op. cit., p. 41.

57. STJUE de 7 de julio de 2023 (C-251/22), cons. 68-73.

58. Vid. CEPD (2024), Report of the work undertaken by the ChatGPT Taskforce, https://www.edpb.europa.eu/our-work-tools/our-documents/other/report-work-undertaken-chatgpt-taskforce_en, p. 7, cons. 18.

59. STJUE de 7 de julio de 2023 (C-251/22), cons. 77.

60. STJUE de 7 de julio de 2023 (C-251/22), cons. 75.

61. STJUE de 7 de julio de 2023 (C-251/22), cons. 79.

62. STJUE de 7 de julio de 2023 (C-251/22), cons. 117.

63. El tenor literal de este Considerando es prácticamente idéntico al de la Enmienda 11 del Parlamento Europeo.

64. Martín Urganga, A., «Protección de datos y fomento de la investigación científica: la necesidad de un equilibrio adecuado», en TRONCOSO REIGADA, A. (dir.), Comentario al Reglamento General de Protección de Datos y a la Ley Orgánica de Protección

las actividades destinadas a explotación comercial (Cons. 159 y 162 RGPD). Todo ello permite afirmar la conveniencia de diseñar una nueva excepción en el marco del art. 9.2 RGPD referida al uso de datos personales para el entrenamiento de modelos y sistemas IA de uso general con las salvaguardas adecuadas para preservar el equilibrio entre el interés social en los beneficios derivados del entrenamiento de la IA y la protección de los derechos y libertades de los ciudadanos⁶⁵.

V. APLICABILIDAD DEL REGLAMENTO COMO NORMA GENERAL Y EVOLUCIÓN NORMATIVA DEL TRATAMIENTO DE LA INTELIGENCIA ARTIFICIAL DE USO GENERAL

Las normas armonizadas del RIA respecto de sistemas IA de alto riesgo son normas generales y, por lo tanto, deben entenderse sin perjuicio de aquellas otras relativas a protección de datos, protección de los consumidores, derechos fundamentales, empleo, protección de los trabajadores y seguridad de los productos (Cons. 9 RIA). Particularmente, por lo que se refiere a los modelos IA de uso general, el RIA no afecta a la normativa de la UE en materia de derechos de autor (Cons. 108).

El RIA es también una norma de introducción y seguimiento de productos en el mercado de la UE, tal como podemos ver en sus arts. 1 y 2 y en el Cons. 118. El art. 1.2.e RIA anuncia, precisamente, que se establecen normas armonizadas para la introducción de modelos IA de uso general en el mercado de la UE, reglas que resultan aplicables a los proveedores que pretendan introducirlos en el mercado de la UE independientemente de si tienen su sede en la UE o en un tercer Estado (art. 2.1.a RIA). Por lo tanto, debe interpretarse de conformidad con las normas que se refieren a este aspecto como los Reglamentos UE 765/2008 y 1020/2019 o la Decisión 768/2008/EC (Cons. 9 RIA), y sobre todo con el Reglamento UE 2019/1020, relativo a la vigilancia del mercado y la conformidad de los productos. Tanto es así que el art. 18 del Reglamento UE 2019/1020 actúa como norma general en materia de derechos procedimentales de los proveedores (Cons. 164 *in fine* RIA).

Es importante en este sentido tener en cuenta que muchos sistemas y modelos IA se introducen en el mercado de la UE directamente en el entorno digital, por ejemplo, en plataformas o motores de búsqueda de muy gran tamaño (VLOP/VLOSE). En estos casos el RIA complementa lo dispuesto en el Reglamento de Servicios Digitales en materia de gestión de riesgos. Particularmente, las plataformas y motores de búsqueda de muy gran tamaño deben realizar una evaluación de riesgos sistémicos derivados del diseño, funcionamiento y uso de sus servicios y, en su caso, adoptar las oportunas medidas paliativas. Los requisitos que establece el RIA aplicables a sistemas y modelos IA se presumen cumplidos si se ha cumplido ya con el RSD salvo si se detectan riesgos sistemáticos diferentes de los que cubre el RSD (Cons. 118). El art. 34.1 RSD detalla una lista de potenciales riesgos sistemáticos de VLOP/VLOSE, algunos de los cuales pueden encontrarse (parcialmente) reflejados en el Anexo III del RIA. Sin embargo, no parece que se trate de una lista cerrada si tenemos en cuenta ya no solo la forma abierta en que se describen los riesgos sistemáticos en el art.

de Datos Personales y Garantía de los Derechos Digitales, Aranzadi, Navarra, 2021, p. 1221.

65. Novelli et al., op. cit., p. 9.

34.1 RSD, sino la previsión publicación de informes anuales de la Junta de Servicios Digitales en cooperación con la Comisión que incluyan la detección y evaluación de los riesgos sistemáticos más destacados y recurrentes notificados por las VLOP/VLOSE (art. 35.2 RSD). Por ello, parece que el Cons. 118 RSD está pensando en riesgos sistemáticos que circunstancialmente se detecten después de remitida la información en cumplimiento del RSD y a los que específicamente se refiera el RIA.

Como ya se ha dicho, las normas en materia de modelos IA de uso general fueron introducidas una vez elaborado el grueso de la propuesta de RIA y tras la eclosión de herramientas como ChatGPT. Fruto de esta introducción de última hora de las normas relativas a IA de uso general es también la distribución de competencias en materia de supervisión (Cons. 161): si un sistema IA está basado en un modelo IA de uso general y tanto uno como otro son del mismo proveedor, la supervisión se llevará a cabo a nivel de la UE por parte de la Oficina IA⁶⁶, que tendrá poderes de autoridad de vigilancia del mercado de acuerdo con el Reglamento UE 2019/1020⁶⁷. En el resto de casos serán las autoridades nacionales de vigilancia del mercado las encargadas de supervisar, si bien cuando un sistema IA de uso general pueda ser utilizado directamente por los responsables de su despliegue para finalidades consideradas de alto riesgo las autoridades nacionales deberán cooperar con la Oficina de la IA (OIA)⁶⁸ siguiendo el procedimiento de asistencia mutua transfronteriza previsto en el capítulo VI del Reglamento UE 2019/1020 (arts. 22 a 24).

Antes de cerrar este apartado acerca de la evolución de la normativa sobre sistemas y modelos IA de uso general conviene una breve referencia a la enmienda 213 del Parlamento Europeo, que finalmente no ha pasado a la versión final del RIA. En esta enmienda se propone la introducción de un artículo relativo a principios generales aplicables a todos los sistemas IA, una suerte de homólogo (salvando las distancias) al artículo 5 RGPD. En tanto que principios generales, habrían resultado aplicables a todo sistema y modelo de IA, incluidos los de uso general. De acuerdo con esta propuesta, los operadores de modelos y sistemas IA deben cumplir, en aras de promover un enfoque europeo coherente, centrado en el ser humano y con una IA ética y fiable, los siguientes principios: (1) intervención y vigilancia humanas; (2) solidez y seguridad técnicas; (3) privacidad y gobernanza de datos; (4) transparencia (habría que añadir, explicable); (5) diversidad, no discriminación y equidad; y (6) bienestar social y medioambiental. No es éste el lugar para reflexionar acerca de si debían ser estos seis principios u otros⁶⁹, pero sí al menos para lamentar que finalmente no se haya introducido un artículo de este tipo habida cuenta del vasto desarrollo reflexivo acerca de principios éticos de la IA.

66. La Oficina IA se establece por Decisión (DOIA) de la Comisión de 24 de enero de 2024 (art. 3.47 y art. 64 RIA). Decisión disponible en: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32024D01459>

67. Vid. art. 3.4 y 10 Reglamento UE 2019/1020.

68. En la Propuesta RIA de 2021 se habla del Comité Europeo de la IA. Esta denominación cambia tras las Enmiendas 524 y siguientes del Parlamento Europeo.

69. Puede consultarse para ello a Cotino Hueso, L., «Ética en el diseño para el desarrollo de una inteligencia artificial, robotica y big data confiables y su utilidad desde el derecho», Revista catalana de dret públic, n.º 58, 2019, <https://revistes.eapc.gencat.cat/index.php/rcdp/article/view/10.2436-rcdp.i58.2019.3303/n58-cotino-es.pdf>, pp. 36-40.

VI. OBLIGACIONES DE LOS PROVEEDORES DE INTELIGENCIA ARTIFICIAL DE USO GENERAL EN EL REGLAMENTO

Podemos agrupar las obligaciones de proveedores de IA de uso general en tres: (1) obligaciones previas a la comercialización (art. 54 RIA); (2) generales (art. 53 RIA); y específicas de proveedores de modelos IA de uso general que comportan riesgo sistémico (art. 55 RIA). Junto con todo ello, debe tenerse presente que la Comisión podrá solicitar al proveedor del modelo IA de uso general no sólo la documentación a que se refieren los artículos 53 y 55 RIA, sino cualquier otra que considere necesaria para evaluar el cumplimiento del RIA por parte del proveedor (art. 91.1 RIA).

La propuesta original de obligaciones para los proveedores de modelos IA de uso general (llamados entonces «modelos fundacionales») se recoge en la Enmienda 399 del Parlamento Europeo, que proponía un nuevo artículo 28.ter.2, donde se recogían obligaciones previas a la comercialización del modelo, y posteriores a su comercialización. El sistema ha cambiado considerablemente en el texto definitivo: las obligaciones previas a la comercialización del modelo son únicamente aplicables a proveedores de terceros países (art. 54.1 RIA), se ha añadido la necesidad de nombrar un representante autorizado en la UE y de cooperar con la OIA. Respecto de las obligaciones posteriores a la comercialización, se mantiene la obligación de conservar la documentación actualizada durante diez años (en relación con el plazo, art. 18.1 RIA), y se añade la de elaborar y publicar un resumen detallado del contenido de entrenamiento de la herramienta.

En primer lugar, tendrán que nombrar a representantes autorizados que estén establecidos en la UE (art. 54 RIA), debiendo fijar un contenido mínimo del mandato: (1) comprobar que se ha elaborado la documentación técnica, de la que deberá conservar copia a disposición de la OIA y las autoridades nacionales competentes; (2) que el proveedor cumple con las obligaciones del art. 53 y, en su caso, del 55; (3) facilitar a la OIA, previa solicitud motivada, la información que demuestre el cumplimiento de dichas obligaciones; (4) y cooperar con la OIA y autoridades nacionales en acciones emprendidas con modelos IA de uso general con riesgo sistémico.

En el artículo 53 RIA se recoge un elenco de obligaciones de carácter general, que pueden agruparse en tres tipos: (1) documentales (2) relativas a propiedad intelectual, y (3) de cooperación. No se incluye la obligación de que vigilar que los modelos IA de uso general proporcionen resultados fiables⁷⁰.

Por lo que se refiere a las obligaciones documentales, los proveedores de modelos de IA de uso general deben: (1) elaborar y mantener actualizada la documentación técnica del modelo (incluyendo información sobre el proceso de entrenamiento, realización de pruebas y resultados de evaluación); (2) elaborar y mantener actualizada información y documentación que permita a los proveedores de sistemas IA que tengan intención de integrar el modelo IA de uso general entender las capacidades y limitaciones de dicho modelo en aras de cumplir con el RIA; y (3) elaborar y poner a disposición del público un resumen suficientemente detallado del contenido utilizado para el entrenamiento del modelo, de acuerdo con el modelo facilitado por la OIA (arts. 53.1. a, b y d, y 53.7 RIA).

70. Watcher et al., op. cit., p. 40.

Los modelos de IA de uso general de código abierto no estarán obligados a cumplir con las obligaciones documentales de los apartados a y b del art. 53.1 RIA, pero sí deberán poner a disposición del público un resumen detallado del contenido utilizado para su entrenamiento de acuerdo con el art. 53.1.d RIA. Las obligaciones documentales de los apartados a y b del art. 53.1 RIA deben cumplirse de acuerdo con lo especificado en los Anexos XI y XII RIA respectivamente. Estos Anexos, como ocurre en otros puntos del RIA, pueden actualizarse en función de los avances tecnológicos mediante actos delegados del art. 97 RIA (art. 53.5 y 6 RIA respectivamente). Finalmente, toda la información y documentación elaborada en el marco del art. 53 está sujeta a deber de confidencialidad en los términos del artículo 78 RIA (art. 53.7 RIA). En la medida en que dicha documentación incluya datos personales el deber de confidencialidad del art. 78 RIA deberá completarse con el art. 5.1.f RGPD.

Respecto de las obligaciones documentales, debe incluirse las relativas a la elaboración de documentación (que deberá mantenerse actualizada) sobre el modelo IA de uso general por parte de los proveedores posteriores, como parte de las responsabilidades de los proveedores de modelos IA de uso general a lo largo de la cadena de valor de La. Debe tenerse presente que los modelos de IA de uso general pueden constituir la base de sistemas de etapas posteriores suministrados por otros proveedores que, por lo tanto, necesitarán entender bien los modelos y sus capacidades, tanto por motivos técnicos como para poder cumplir con el RIA y demás normativa (Cons. 101 RIA).

Por lo que se refiere a las obligaciones relativas a derechos de autor, éstas vienen recogidas de manera no especialmente clara en el art. 53.1.c RIA, del que ya hemos hablado más arriba⁷¹.

En cuanto a los deberes de cooperación, los proveedores de modelos IA de uso general deben cooperar con la Comisión y las autoridades nacionales competentes en aras de facilitar el cumplimiento del RIA de acuerdo con su artículo 53.3. Este deber genérico de cooperación se concreta en diversos puntos del RIA, y no se circunscribe a la relación entre proveedores de modelos IA de uso general y autoridades, sino también entre autoridades entre sí, en el marco del principio de cooperación leal del art. 4.3 TUE. Así, deben tenerse en cuenta los deberes de cooperación con la OIA en el marco del art. 75.2 RIA, así como las facultades de revisión y supervisión tanto de la OIA (art. 75.1 y 3 RIA), como de la Comisión (art. 88.2 RIA). También, por lo que se refiere a los sistemas IA de uso general, en la medida en que pueden utilizarse como sistemas de IA de alto riesgo bien por sí solos o como partes de éstos, los proveedores de sistemas de IA de uso general deben cooperar estrechamente con los proveedores de sistemas de IA de alto riesgo correspondientes (Cons. 85 RIA).

Además de las obligaciones que con carácter general se establecen para los proveedores de modelos IA de uso general, cuando éstos son calificados como modelos IA de uso general con riesgo sistémico deben cumplir con las obligaciones específicas que establece el art. 55 RIA.

La primera de éstas es evaluar los modelos con vistas a detectar y reducir el riesgo sistémico (art. 55.1.a RIA). Esta evaluación debe realizarse de acuerdo con protocolos

71. Vid. supra, apdo. sobre retos normativos de la IA generativa.

y herramientas normalizados según el estado de la técnica e incluir pruebas de simulación de adversarios con el modelo. Las pruebas de simulación de adversarios o pruebas de robustez permiten identificar vulnerabilidades a través de la simulación sistemas atacantes dentro de una red y sugerir mejoras que permitan una mejor comprensión y mejora continua del modelo y refuercen su seguridad y fiabilidad⁷². También deberán detectar el origen, evaluar y reducir los riesgos sistémicos a escala de la UE que puedan derivarse del desarrollo, introducción en el mercado o uso del modelo IA de uso general con riesgo sistémico (art. 55.1.b RIA); vigilar, documentar y notificar sin demora indebida a la OIA información relativa a incidentes graves y posibles medidas correctoras; y también velar porque se establezca un nivel adecuado de protección de la ciberseguridad y la infraestructura física del modelo (art. 55.1.c y d RIA).

El deber de confidencialidad del art. 53.7 RIA se reitera en el art. 55.3 RIA, con el mismo tenor literal. Cabe preguntarse por la necesidad del art. 55.3 RIA si el art. 53 ya es aplicable con carácter general a todos los proveedores de modelos IA de uso general, sean estos modelos con riesgo sistémico o no lo sean. Finalmente, los códigos de buenas prácticas permiten a los proveedores de modelos IA de uso general con riesgo sistémico demostrar el cumplimiento de sus obligaciones (art. 55.2 RIA).

Finalmente, los códigos de buenas prácticas son una herramienta fundamental para el cumplimiento adecuado de las obligaciones derivadas del RIA por parte de los proveedores de modelos IA de uso general, al tiempo que les facilitan demostrar el cumplimiento de dichas obligaciones (Cons. 117 y arts. 53.4 y 55.2 RIA). Se trata de una introducción que no estaba prevista en la Propuesta de 2021 ni tampoco en las Enmiendas del Parlamento Europeo. Las normas armonizadas, por su parte, son a los efectos del RIA todas aquellas especificaciones técnicas adoptadas por un organismo de normalización reconocido cuya observancia no es obligatoria pero que son de aplicación repetida o continua, elaboradas a raíz de una petición de la Comisión (art. 3.27 RIA y 2.1.c Reglamento UE 1025/2012). Queda siempre en manos del proveedor utilizar métodos alternativos diferentes de los códigos de buenas prácticas y las normas armonizadas.

VII. SUPERVISIÓN Y SEGUIMIENTO, Y RÉGIMEN SANCIONADOR

La Comisión tiene atribuidas competencias de supervisión y control del cumplimiento de las obligaciones de los proveedores de modelos de IA de uso general, cuya ejecución delega en la OIA, que debe poder adoptar las medidas necesarias para supervisar la aplicación efectiva y el cumplimiento de las obligaciones de los proveedores de modelos IA de uso general establecidas en el RIA, pudiendo para ello requerir información, evaluar e imponer medidas a proveedores de IA de uso general. Puede contar, además, con asesoramiento por parte del grupo de expertos científicos previsto en el RIA, que deben seleccionarse sobre la base de conocimientos científicos o técnicos en el ámbito de la IA y desempeñar sus funciones con imparcialidad y objetividad (Cons. 162 y 164, y arts. 68 y 88 RIA).

72. Hannon, B.; Kumar, Y.; LI, J. J.; Morreale, P., «From Vulnerabilities to Improvements—A Deep Dive into Adversarial Testing of AI Models», *Congress in Computer Science, Computer Engineering & Applied Computing*, 2023, pp. 2645-2649.

La OIA tiene facultades de supervisión cuando un sistema IA se basa en un modelo IA de uso general y un mismo proveedor desarrolla modelo y sistema, teniendo a estos efectos la consideración de autoridad de mercado de acuerdo con el Reglamento UE 2019/1020 (art. 75.1 RIA). Asimismo, es competente para tomar medidas en relación con la aplicación y cumplimiento del RIA por parte de proveedores de modelos IA de uso general, así como para la observancia de los códigos de buenas prácticas que hayan sido aprobados (art. 89.1 RIA); y para realizar tareas de evaluación en el marco del art. 92 RIA.

Respecto del régimen sancionador en el RIA, puede dividirse en dos categorías: adopción de medidas y sanciones propiamente dichas, el primero en manos de la Comisión y el segundo de los Estados miembros y del SEPD (cuando se tratase de conductas contrarias al RGPD). La concreción de las sanciones queda en manos de los Estados miembros, salvo por lo que se refiere a las multas a proveedores de modelos IA de uso general que actúen de forma deliberada o negligente, que de acuerdo con el art. 101 RIA son impuestas por la Comisión (art. 101.1 RIA).

La Comisión es competente para: (1) solicitar la adopción de medidas encaminadas al oportuno cumplimiento de las obligaciones de los proveedores de modelos IA de uso general; (2) exigir a un proveedor que aplique medidas de reducción de riesgos cuando la evaluación realizada de acuerdo con el art. 92 RIA apunte a la existencia de un riesgo sistémico a escala de la UE; y (3) restringir la comercialización del modelo (art. 93.1.a, b y c RIA). Antes de solicitar la adopción de medidas, la OIA puede entablar un diálogo estructurado con el proveedor del modelo IA de uso general, encaminado a evitar la actuación unilateral de la Comisión, ya que si durante el diálogo estructurado el proveedor se compromete a adoptar medidas de reducción del riesgo sistémico la Comisión puede adoptar una decisión por la cual convierta en vinculantes los compromisos del proveedor y declare que no hay motivos para actuar (art. 93.2 y 3 RIA).

Respecto multas, deberán ser de una «cuantía apropiada» (Cons.169 RIA), y serán «efectivas, proporcionadas y disuasorias» (art. 101.3 RIA) teniendo en cuenta: (1) la naturaleza, gravedad y duración de la infracción (2) los principios de proporcionalidad y adecuación, y (3) los compromisos que hubieran contraído los proveedores en virtud del art. 93.3 RIA o de la adhesión a códigos de buenas prácticas. El RIA fija un tope máximo de cuantía de multa que podrá imponer la Comisión. La cuantía máxima de estas multas será del 3% del volumen de negocios mundial total correspondiente al ejercicio financiero anterior o de 15 millones de euros. Los supuestos en que la Comisión puede imponer multas son: (1) infracción de las normas del RIA; (2) no haber atendido a una solicitud de documentación o información en el marco del art. 91 RIA o haber facilitado información inexacta, incompleta o engañosa; (3) haber incumplido una medida solicitada en virtud del art. 93; o (4) no haber dado acceso a la Comisión al modelo IA de uso general para realizar una evaluación en el marco del art. 92 RIA.

La imposición de multas debe realizarse de acuerdo con ciertas reglas de procedimiento recogidas en el art. 101. Tanto la actuación de la Comisión como del TJUE, si nos atenemos al tenor literal del RIA, serán de propia iniciativa cuando se den las circunstancias previstas. No obstante, podríamos preguntarnos si cabe la posibilidad de que particulares denuncien infracciones del RIA o de la Comisión en

la imposición de multas a pesar de que no esté expresamente previsto en la norma. En otras palabras, ¿es suficiente para invocar una norma de la UE que ésta enuncie una obligación clara e incondicional o debe estar prevista también la posibilidad de denuncia de infracción por particulares? La respuesta a esta pregunta fue resuelta por la conocida Sentencia del TJUE de 17 de septiembre de 2002, que entendió que la garantía de operatividad de la normativa de la UE exigía también que su cumplimiento pudiera instarse también en el marco de procesos civiles iniciados por particulares⁷³. Ello es coherente con la asunción de que es el juez nacional el encargado del cumplimiento del Derecho de la UE en cada Estado Miembro. Este criterio ha sido propuesto respecto de normas UE recientes que tampoco contemplan de manera explícita la posibilidad de que ciudadanos particulares interpongan acciones para reclamar su cumplimiento, como ocurre con el Reglamento UE 2019/1150⁷⁴. Parece, a nuestro juicio, razonable entender que este criterio es igualmente aplicable al RIA, lo cual puede ser especialmente interesante durante sus primeros años de funcionamiento, cuando es previsible que las diferentes normas de desarrollo previstas estén aún en fase de elaboración.

VIII. CONCLUSIONES

La eclosión de herramientas como Chat GPT a finales de 2022 ha ocasionado un terremoto social y también normativo. Buena muestra del terremoto normativo es la introducción de disposiciones específicas relativas a modelos de IA de uso general en la versión definitiva del RIA, incorporando (y ampliando) las enmiendas presentadas por el Parlamento Europeo en junio de 2023 a la Propuesta de RIA de 2021, que por primera vez introducen disposiciones específicas relativas a los entonces llamados «modelos fundacionales». También es buena muestra del enorme impacto producido por estas herramientas el hecho de que la necesidad y modo de incluirlas en el texto del RIA fue uno de los últimos puntos críticos de debate en la negociación que tuvo lugar a finales de 2023.

Los modelos IA de uso general son modelos IA entrenados con gran cantidad de datos utilizando supervisión a gran escala que presentan un grado considerable de generalidad y son capaces de realizar de manera competente una gran variedad de tareas distintas así como de integrarse en diversos sistemas o aplicaciones posteriores (art. 3.63 RIA). Un sistema IA se considerará de uso general cuando esté basado en un modelo IA de uso general (art. 3.66 RIA). Dentro de los modelos IA de uso general deben destacarse dos subcategorías: los modelos de IA generativa, que son aquellos capaces de generar de manera flexible contenidos de texto, audio, imágenes o video; y los modelos de uso general que presentan riesgo sistémico.

Un modelo IA de uso general presenta riesgo sistémico si tiene capacidades de gran impacto (de acuerdo con el Anexo XIII RIA, que podrá actualizarse según la evolución de la técnica), o si presenta unas repercusiones potenciales considerables en el mercado interior debido a su alcance. Los proveedores de modelos IA que presenten riesgo sistémico deberán cumplir con una serie de obligaciones añadidas

73. STJUE de 17 de septiembre de 2002 (C-251/22), cons. 30 y 31.

74. Jens-Uwe, F., «Individual Private Rights of Action under the Platform-to-Business Regulation», *European Business Law Review*, vol. 34, tomo 4, 2023, pp. 559-560.

a las que debe cumplir todo proveedor de modelos IA de uso general (arts. 53 y 54 RIA, para obligaciones generales y específicas de modelos IA con riesgo sistémico). Esto justifica una cierta necesidad de seguridad jurídica para estos proveedores, que deberán considerarse dentro de esta categoría siempre que superen ciertos estándares y así lo comuniquen a la Comisión o bien cuando la Comisión les comunique que tienen dicha condición por haber superado ciertos estándares.

Los modelos de IA generativa, por su parte, presentan ciertos retos relacionados con la responsabilidad civil derivada del uso de la IA, la desinformación⁷⁵, la protección de datos personales y la propiedad intelectual⁷⁶. Por lo que se refiere al tratamiento de datos personales, los problemas se refieren a si y cómo deben tratarse los datos personales tanto en el entrenamiento de las herramientas de IA como posteriormente en su uso. Debe recordarse que los modelos IA de uso general se caracterizan por utilizar grandes cantidades de datos, lo que arroja dudas acerca de la base de legitimación (y, en su caso, excepciones a la prohibición general de tratamiento de datos de categorías especiales) para el entrenamiento del modelo, o relativos a los datos personales que los propios usuarios introducen en las directrices (prompts) a la herramienta.

En cuanto a los aspectos relativos a propiedad intelectual, por un lado los resultados producidos por modelos IA de uso general presentan cierta similitud con las obras derivadas, si bien en la medida en que se producen a través de minería de textos y datos debe tenerse en cuenta lo dispuesto en el art. 4.3 de la Directiva 2019/790, a la que el RIA se remite. También es interesante preguntarse cuándo un resultado producido por una herramienta de IA generativa puede ser protegible: el criterio a tener en cuenta, a nuestro juicio, debe seguir siendo si es posible reconocer una actividad humana genuina en el uso de la herramienta, lo que se traduce en preguntarse si las instrucciones introducidas revisten suficiente complejidad y creatividad (algo similar a la clásica distinción entre obra fotográfica y mera fotografía). Lo difícil será determinar en la práctica cuándo es reconocible esta actividad humana genuina. Una última cuestión, no jurídica pero importante, son las repercusiones sociales que puede tener la facilidad de acceso a herramientas que facilitan huir de elaborar pensamientos humanos complejos, o como mínimo huir de la mera disciplina de la actividad de escribir, por ejemplo.

El RIA es una norma de introducción y seguimiento de productos en el mercado de la UE (sistemas y modelos de IA). Esto se ve en las remisiones en diversos puntos al Reglamento UE 2019/1020. Ello explica también que las obligaciones que se fijan para modelos IA de uso general tengan que ver fundamentalmente con información, documentación y cooperación con autoridades (Comisión, OIA, y autoridades nacionales). Se distinguen tres tipos de obligaciones para proveedores de modelos IA de uso general: previas a la comercialización del modelo (aplicables sólo a proveedores que no estén establecidos en la UE), generales, y específicas respecto de proveedores de modelos IA de uso general que presenten riesgo sistémico. Estas obligaciones

75. Watcher et al., op. cit., p. 5.

76. BSI — Oficina Federal para la Seguridad de la Información del Gobierno alemán (2024): Generative AI Models — Opportunities and Risks for Industry and Authorities, https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/Generative_AI_Models.html, p. 9.

fueron introducidas en un primer momento por la Enmienda 399 del Parlamento Europeo; sin embargo, sufrieron modificaciones considerables en su estructura y contenido. En la Enmienda 399 las obligaciones previas a comercialización no quedaban restringidas a proveedores de fuera de la UE y eran algo más extensas. Las generales, en cambio, no quedaban tan detalladas como en el art. 53 RIA. No estaban previstas obligaciones específicas para modelos IA que presentasen riesgos sistémicos en la medida en que esta categoría tampoco figuraba.

El sistema normativo aplicable a los modelos IA de uso general se cierra en el RIA con previsiones relativas a supervisión y seguimiento, y a régimen sancionador. La Comisión tiene atribuidas competencias de supervisión y control del cumplimiento normativo, cuya ejecución delega en la OIA, que en su norma de funcionamiento (Decisión de 24 de enero de 2024) contempla varias previsiones de control respecto de modelos IA de uso general.

El régimen sancionador se divide en dos categorías: medidas y multas. Respecto de las medidas, la Comisión es competente para solicitar la adopción de medidas encaminadas al cumplimiento normativo, exigir a un proveedor que aplique medidas de reducción de riesgos cuando exista riesgo sistémico y, en última instancia, restringir la comercialización del modelo. La OIA juega un papel importante en este contexto en la medida en que puede entablar diálogos estructurados encaminados a evitar la actuación de la Comisión. En cuanto a las multas, éstas deben ser de cuantía apropiada, efectivas, proporcionadas y disuasorias con carácter general (art. 101.3 RIA). Con carácter general, son los Estados quienes deben fijar la cuantía de las multas; pero en el caso de los modelos IA de uso general será la Comisión la competente para imponerlas de acuerdo con los topes establecidos del 3% del volumen de negocio mundial del ejercicio financiero anterior o 15 millones de euros. A modo de cierre, debe entenderse posible que particulares acudan a la jurisdicción ordinaria nacional para denunciar el incumplimiento del RIA aunque no esté expresamente previsto en la norma.

Códigos de conducta, sellos o certificaciones para los sistemas de inteligencia artificial que no son de alto riesgo (artículo 95 del Reglamento)

LORENZO COTINO HUESO

Catedrático de Derecho Constitucional de la Universitat de València. Valgrai¹

I. LA REGULACIÓN EN EL ARTÍCULO 95 DE CÓDIGOS DE CONDUCTA PARA SISTEMAS QUE NO SON DE ALTO RIESGO

El RIA regula esencialmente obligaciones para los sistemas de alto riesgo. No obstante, también contiene algunas disposiciones sobre los modelos de IA de uso general (Capítulo V, arts. 51-56) e impone algunas obligaciones de transparencia a «determinados» sistemas de IA en el artículo 50, Capítulo IV.

La Comisión Europea estima que el 90% de los sistemas de IA² y dos tercios de los sistemas de IA públicos no serán clasificados como de alto riesgo.³ Bajo el modelo de riesgos del RIA, los sistemas que no son de alto riesgo no estarán sujetos a las obligaciones del reglamento. Sin embargo, estos sistemas seguirán sujetos a otras

1. cotino@uv.es. OdiseIA. El presente estudio es resultado de investigación de los siguientes proyectos: MICINN Proyecto «Derechos y garantías públicas frente a las decisiones automatizadas y el sesgo y discriminación algorítmicas» 2023-2025 (PID2022-136439OB-I00) financiado por MCIN/AEI/10.13039/501100011033/; «La regulación de la transformación digital ...» Generalitat Valenciana «Algorithmic law» (Prometeo/2021/009, 2021-24); «Algorithmic Decisions and the Law: Opening the Black Box» (TED2021-131472A-I00) y «Transición digital de las Administraciones públicas e inteligencia artificial» (TED2021-132191B-I00) del Plan de Recuperación, Transformación y Resiliencia. Estancia Generalitat Valenciana CIAEST/2022/1., Grupo de Investigación en Derecho Público y TIC Universidad Católica de Colombia; Estancia Generalitat Valenciana CIAEST/2022/1, Convenio de Derechos Digitales-SEDIA Ámbito 5 (2023/C046/00228673) y Ámbito 6. (2023/C046/00229475).
2. Comisión Europea, Renda. A. (project leader), *Study to Support an Impact Assessment of Regulatory Requirements for Artificial Intelligence in Europe. Final Report (D5)*, abril 2021. p. 143, <https://op.europa.eu/es/publication-detail/-/publication/55538b70-a638-11eb-9585-01aa75ed71a1>
3. JRC, Tangi, L. y otros: *AI Watch European landscape on the use of Artificial Intelligence by the Public Sector*, JRC Science For Policy Report, Unión Europea. 2022, p. 58.

normativas relevantes, como el RGPD en caso de tratamiento de datos personales. Asimismo, para garantizar la seguridad de los productos, se aplicará como red de seguridad el Reglamento (UE) 2023/988 del Parlamento Europeo y del Consejo, de 10 de mayo de 2023, relativo a la seguridad general de los productos (Considerando 166). Dicho reglamento dispone que «establece normas esenciales sobre la seguridad de los productos de consumo introducidos en el mercado» (art. 1.2º) y viene a ser el «coche escoba» regulatorio, pues se aplica a los productos «en la medida en que no existan disposiciones específicas con la misma finalidad en el Derecho de la Unión que regulen la seguridad de los productos de que se trate.» (art. 2.1º).

Para los sistemas que no son de alto riesgo, desde la versión inicial del RIA en 2021⁴ se incluyó un Título IX relativo a los «Códigos de conducta», con el objetivo de que estos cumplan voluntariamente los requisitos obligatorios para los sistemas de IA.⁵ Estos códigos podrían también incluir compromisos voluntarios sobre sostenibilidad medioambiental, accesibilidad para personas con discapacidad, participación de partes interesadas en el diseño y desarrollo de sistemas de IA y diversidad en los equipos de desarrollo. La intención del RIA es que los sistemas que no son de alto riesgo sean «seguros una vez introducidos en el mercado o puestos en servicio» (Considerando 166). El artículo 95 ha tenido pocas variaciones, siendo relevante la aparición final de la Oficina de IA y la atribución de responsabilidades a la misma, inicialmente asignadas a la Comisión y al Comité. En la versión final, se han añadido elementos adicionales que estos Códigos pueden incluir, vinculados a las Directrices éticas de la Unión Europea, repercusiones medioambientales, alfabetización, diseño inclusivo o perjuicios para personas vulnerables.

Dado que la aplicación de las obligaciones del RIA u otras sería voluntaria y no obligatoria, el RIA conecta esta cuestión con el ámbito de la ética. Respecto a los sistemas que no son de alto riesgo, se «puede dar lugar a la adopción más amplia de una IA ética y fiable en la Unión» (Considerando 165). Se trata de la marca IA de la UE, conocida también como IA ética en el diseño.⁶

4. La cuestión se atendía en el Considerando 81 y en el artículo 69, además de la explicación general del Reglamento.

5. Sobre el tema puede seguirse: Stuurman, K. and Lachaud, E., *Regulating AI. A Label To Complete the Proposed Act on Artificial Intelligence* enero 2022 <http://dx.doi.org/10.2139/ssrn.3963890>

Galán, C., «The Certification as a Mechanism for Control of Artificial Intelligence in Europe» septiembre 2019. <http://dx.doi.org/10.2139/ssrn.3451741> también en «La certificación como mecanismo de control de la inteligencia artificial en Europa» en *bie3: Boletín IEEE*, n.º 14, 2019, pp. 622-637.

Cihon, P. y otros «AI Certification: Advancing Ethical Practice by Reducing Information Asymmetries» en *IEEE Transactions on Technology and Society*, LawAI Working Paper No. 5-2021, <https://doi.org/10.1109/TTS.2021.3077595>

6. Un análisis exhaustivo lo realizo en «Ética en el diseño para el desarrollo de una inteligencia artificial, robótica y big data confiables y su utilidad desde el derecho» en *Revista Catalana de Derecho Público* n.º 58 (junio 2019). <http://revistes.eapc.gencat.cat/index.php/rcdp/issue/view/n58> <http://dx.doi.org/10.2436/rcdp.i58.2019.3303>

El artículo 95 del RIA, que inicialmente constituía un capítulo completo y ahora es el Capítulo X, «Códigos de conducta y directrices», regula con carácter normativo bastante abierto:

— El fomento y facilitación de códigos y mecanismos de gobernanza por parte de la Oficina IA y los Estados miembros.

— La aplicación voluntaria y variable de las obligaciones de los sistemas de alto riesgo a los que no son de alto riesgo, dirigida especialmente a proveedores, pero también a desplegados.

— Además del cumplimiento voluntario del RIA, se señala que estos Códigos voluntarios puedan introducir «requisitos adicionales» como los de las Directrices éticas de la Unión Europea, repercusiones medioambientales, alfabetización, diseño inclusivo o perjuicios para personas vulnerables.⁷

— En la elaboración de códigos se deben tener en cuenta las mejores prácticas y soluciones técnicas, y deben desarrollarse «sobre la base de objetivos claros e indicadores clave de resultados para medir la consecución de dichos objetivos» (art. 95.2°).

— Los códigos de conducta pueden ser elaborados por proveedores, desplegados, organizaciones que los representen, sociedad civil y el mundo académico,⁸ con mención a los intereses de pymes y empresas emergentes.

Además del artículo En otros apartados del RIA se hacen algunas menciones a los códigos y certificaciones. Así, se menciona que un objetivo de un sandbox puede ser también el aprendizaje de la aplicación no sólo del RIA, sino, también de los códigos de conducta (art. 58.2° e). El Comité tiene la función de «emitir recomendaciones y dictámenes por escrito» sobre la elaboración y aplicación de códigos de conducta y de buenas prácticas (art. 66.e)i). Cada tres años, la Comisión debe evaluar la repercusión y eficacia de los códigos de conducta voluntarios para fomentar la aplicación de los requisitos para los sistemas de IA de alto riesgo a los que no lo son, y posiblemente otros requisitos adicionales (Considerando 174). No se aborda ahora los importantes Códigos de buenas prácticas para el ámbito de la IA de uso general regulados en el artículo 56.

II. FINALMENTE EL REGLAMENTO NO HA INCLUIDO UNOS PRINCIPIOS OBLIGATORIOS PARA TODO TIPO DE SISTEMA INTELIGENCIA ARTIFICIAL

Hay que llamar la atención de que las enmiendas al RIA del Parlamento de la UE en 2023 incluyeron la regulación de unos principios «Principios generales aplicables a todos los sistemas de IA» (artículo 4 bis, nuevo). Se pretendía seguir el esquema del RGPD cuando reconoce sus esenciales principios del artículo 5. Como es sabido, en protección de datos durante más de treinta años los «principios» son los pilares fundamentales del marco jurídico de protección de datos, es más, constituyen reglas concretas aplicables a los tratamientos.

7. Así se señala en el Considerando 165.

8. «como las organizaciones empresariales y de la sociedad civil, el mundo académico, los organismos de investigación, los sindicatos y las organizaciones de protección de los consumidores», Considerando 165.

Tan es así que su mero incumplimiento directamente implica la comisión de infracciones.

Cabe destacar que las enmiendas al RIA del Parlamento de la UE en 2023 incluyeron los «Principios generales aplicables a todos los sistemas de IA» (artículo 4 bis, nuevo).⁹ Se pretendía seguir el esquema del RGPD, donde los principios del artículo 5¹⁰ son pilares fundamentales además de reglas concretas aplicables cuyo incumplimiento implica infracciones.

Se proponía incluir principios para todo sistema de IA, de alto riesgo o no, así como para los modelos fundacionales. Se proclamaban y con concreción los principios de «intervención y vigilancia humanas» (a),¹¹ «solidez y seguridad técnicas» (b),¹² «privacidad y gobernanza de datos» (c),¹³ «transparencia» (d),¹⁴ «diversidad, no discriminación y equidad» (e)¹⁵ y «bienestar social y medioambiental» (f)¹⁶. Se prescribía que «Todos los operadores [...] se esforzarán al máximo por desarrollar y utilizar los sistemas de IA o modelos fundacionales con arreglo a los siguientes principios generales que establecen un marco de alto nivel para promover un enfoque europeo coherente centrado en el ser humano con respecto a una inteligencia artificial ética y fiable».

También es cierto que en aquella propuesta del RIA por el Parlamento de la UE, los principios se proclamaban con cautela, si se me permite, con «el freno echado», modulando y restringiendo su alcance, afirmando «sin crear nuevas

9. Enmienda 213.

10. El artículo 5 RGPD los regula: licitud, lealtad y transparencia, limitación de la finalidad, adecuación, limitación, necesidad y proporcionalidad de los datos (minimización de datos), exactitud, limitación del plazo de conservación, integridad y confidencialidad y responsabilidad proactiva.

11. «a) “Intervención y vigilancia humanas”: los sistemas de IA se desarrollarán y utilizarán como una herramienta al servicio de las personas, que respete la dignidad humana y la autonomía personal, y que funcione de manera que pueda ser controlada y vigilada adecuadamente por seres humanos.»

12. «b) “Solidez y seguridad técnicas”: los sistemas de IA se desarrollarán y utilizarán de manera que se minimicen los daños imprevistos e inesperados, así como para que sean sólidos en caso de problemas imprevistos y resistentes a los intentos de modificar el uso o el rendimiento del sistema de IA para permitir una utilización ilícita por parte de terceros malintencionados.»

13. «c) “Privacidad y gobernanza de datos”: los sistemas de IA se desarrollarán y utilizarán de conformidad con las normas vigentes en materia de privacidad y protección de datos, y tratarán datos que cumplan normas estrictas en términos de calidad e integridad.»

14. «d) “Transparencia”: los sistemas de IA se desarrollarán y utilizarán facilitando una trazabilidad y explicabilidad adecuadas, haciendo que las personas sean conscientes de que se comunican o interactúan con un sistema de IA, informando debidamente a los usuarios sobre las capacidades y limitaciones de dicho sistema de IA e informando a las personas afectadas de sus derechos.»

15. «e) “Diversidad, no discriminación y equidad”: los sistemas de IA se desarrollarán y utilizarán incluyendo a diversos agentes y promoviendo la igualdad de acceso, la igualdad de género y la diversidad cultural, evitando al mismo tiempo los efectos discriminatorios y los sesgos injustos prohibidos por el Derecho nacional o de la Unión.»

16. «f) “Bienestar social y medioambiental”: los sistemas de IA se desarrollarán y utilizarán de manera sostenible y respetuosa con el medio ambiente, así como en beneficio de todos los seres humanos, al tiempo que se supervisan y evalúan los efectos a largo plazo en las personas, la sociedad y la democracia.»

obligaciones en virtud del presente Reglamento».¹⁷ No obstante, se afirmaba que los principios debían inspirar los procesos de normalización y las orientaciones técnicas.¹⁸

Finalmente, el RIA no incluye principios y, para los sistemas que no son de alto riesgo hay que seguir esencialmente el artículo 95. Es importante mencionar que el Convenio de IA del Consejo de Europa de 17 de mayo de 2024¹⁹ sí lo ha hecho. El Capítulo III de este Convenio trata sobre «Principios relacionados con las actividades dentro del ciclo de vida de los sistemas de inteligencia artificial» y «establece los principios generales comunes que cada Parte deberá aplicar [...] de manera adecuada a su ordenamiento jurídico interno».²⁰ Aunque con cierta laxitud, los principios regulados en ocho artículos tendrán una proyección general para todo tipo de sistema IA.²¹ Obviamente, para los Estados y partes que suscriban el Convenio y una vez entre en vigor.

-
17. Apartado 2: «El apartado 1 se entenderá sin perjuicio de las obligaciones establecidas por el Derecho de la Unión y nacional en vigor. En el caso de los sistemas de IA de alto riesgo, los principios generales son aplicados y cumplidos por los proveedores o implementadores mediante los requisitos establecidos en los artículos 8 a 15 del presente Reglamento, así como las obligaciones pertinentes establecidas en el capítulo 3 del título III del presente Reglamento. En el caso de los modelos fundacionales, los principios generales son aplicados y cumplidos por los proveedores o implementadores mediante los requisitos establecidos en los artículos 28 a 28 ter del presente Reglamento. Para todos los sistemas de IA, la aplicación de los principios a los que se hace referencia en el apartado 1 puede conseguirse, según proceda, a través de las disposiciones del artículo 28 o el artículo 52 o de la aplicación de las normas armonizadas, especificaciones técnicas y códigos de conducta a los que hace referencia el artículo 69, sin crear nuevas obligaciones en virtud del presente Reglamento».
 18. 3. La Comisión y la Oficina de IA incorporarán estos principios rectores en las peticiones de normalización, así como en las recomendaciones consistentes en orientaciones técnicas destinadas a prestar asistencia a proveedores e implementadores en cuanto al modo de desarrollar y utilizar los sistemas de IA. Las organizaciones europeas de normalización tendrán en cuenta los principios generales a que se refiere el apartado 1 del presente artículo como objetivos basados en los resultados cuando elaboren las correspondientes normas armonizadas para los sistemas de IA de alto riesgo a que se refiere el artículo 40, apartado 2 ter.
 19. Convenio Marco del Consejo de Europa sobre Inteligencia Artificial y Derechos Humanos, Democracia y Estado de Derecho, adoptado en el 133º período de sesiones del Comité de Ministros, Estrasburgo, <https://search.coe.int/cm?i=0900001680afb11f> Sobre el mismo, Cotino Hueso, L. «El Convenio sobre inteligencia artificial, derechos humanos, democracia y Estado de Derecho del Consejo de Europa», *Revista Administración & Ciudadanía*, EGAP, 2024.
 20. Así se afirma en la versión de 18 de diciembre como nota explicativa.
 21. Se sigue por la versión de diciembre de 2023, ahí se regulan cho artículos (arts. 6 a 13) en los que esencialmente se expresan y afirman tales «principios»: dignidad humana y autonomía individual (art. 6), transparencia y supervisión (art. 7), rendición de cuentas y responsabilidad (art. 8), igualdad y no discriminación (art. 9), privacidad y protección de datos personales (art. 10), preservación de la salud [y del medio ambiente] (art. 11), fiabilidad y confianza (art. 12), innovación segura (art. 13).

III. LA INSERCIÓN DE LA INTELIGENCIA ARTIFICIAL EN LOS MODELOS DE CERTIFICACIÓN Y SELLOS TECNOLÓGICOS EN LA UE. ¿UN SELLO ESPAÑOL DE INTELIGENCIA ARTIFICIAL?

El RIA incorpora la IA en el ecosistema de certificación voluntaria, sellos y códigos de conducta impulsados por la UE en el ámbito tecnológico. Estos modelos suponen definir estándares claros que las organizaciones deben cumplir para obtener la certificación, verificando su cumplimiento a través de organismos acreditados. Estos instrumentos facilitan demostrar el cumplimiento con estándares de calidad, seguridad y ética, incrementando la confianza de los consumidores y usuarios.

La normativa de la UE respalda estos modelos para dotarlos de credibilidad y reconocimiento oficial, especialmente en sectores como la ciberseguridad, regulados por el Reglamento (UE) 2019/881. También son importantes los códigos de conducta y sellos que demuestran el cumplimiento con el RGPD, como el esquema RGPD-CARPA de la CNPD de Luxemburgo.²² Asimismo, pueden mencionarse también los certificados de servicios de confianza electrónica emitidos conforme al Reglamento eIDAS (Reglamento (UE) N.º 910/2014) que garantizan la autenticidad e integridad de las transacciones electrónicas en la Unión Europea.

Carlos Galán propuso en 2019 la creación de un Esquema Europeo de Certificación para regular el desarrollo y despliegue de tecnologías de IA.²³ Obviamente era pronto para pensar en todo el sistema regulatorio del RIA.

En España, la Estrategia Nacional de IA (ENIA) de 2020²⁴ incluyó el desarrollo de un código de conducta o «sello» como medida para crear confianza en la IA. Esta era la primera medida dentro de la Línea de actuación 6.1, «Crear confianza en la IA», en concreto, la Medida 26 «Desarrollo de un sello nacional de calidad IA y la elaboración de catálogo de medidas suplementarias a la Certificación en IA a nivel europeo». Para ello, el Gobierno delineó acciones para su implementación mediante el contrato «Sello IA del Gobierno Español»²⁵. Siguiendo el pliego técnico de este contrato, las acciones básicas son las siguientes: 1. Desarrollo de las normas técnicas del sello o certificación; 2. Propuesta del marco de acreditación: esquema de certificación y proceso de acreditación; 3. Manuales de buenas prácticas de implantación del

22. Al respecto, CNPD, «The certification scheme GDPR CARPA», en <https://cnpd.public.lu/en/professionnels/outils-conformite/certification/gdpr-carpa.html>

CNPD, GDPR-CARPA, *GDPR-Certified assurance report-based processing activities*, Commission nationale pour la protection des données, Luxemburgo, 2022, <https://cnpd.public.lu/content/dam/cnpd/fr/professionnels/certification/lu-gdpr-carpa-certificationscheme.pdf>

23. Galán, C., «The Certification...» cit. Este esquema habría de estar respaldado por una normativa europea que señalarla los estándares y especificaciones técnicas, la independencia de las entidades evaluadoras y que el sistema incluya procesos de evaluación continua y actualizaciones periódicas.

24. SEDIA, *Estrategia Nacional de Inteligencia Artificial*, noviembre 2020, <https://www.la-moncloa.gob.es/presidente/actividades/Documents/2020/ENIAResumen2B.pdf>

25. Servicios para el desarrollo de planes de impacto de la Inteligencia Artificial, desarrollo de un sello y servicios de estudio relativos a entornos de experimentación de sistemas de IA. <https://planderecuperacion.gob.es/como-acceder-a-los-fondos/convocatorias/PLC/11383932/servicios-para-el-desarrollo-de-planes-de-impacto-de-la-inteligenciaartificial-desarrollo-de-un-sello-y-servicios-de-estudio-relativos-a-entornos-de-experimentacion-de-sistemas-de-ia>

Sello IA y 4. Desarrollo de una herramienta software para la auto-evaluación del cumplimiento de los requerimientos.

Así, en primer término, el objetivo es establecer requisitos técnicos alineados con los estándares europeos, abarcar aspectos de seguridad y protección de datos específicos de la IA. En segundo lugar, para el desarrollo del Sello español se pretende desarrollar un marco de acreditación en colaboración con entidades como AENOR, UNE o Adigital. Además, se considera importante ofrecer caminos alternativos de certificación para PYMES y establecer procedimientos claros para el mantenimiento y retiro de la acreditación.

En tercer lugar, se está previsto el desarrollo de manuales de Buenas Prácticas que expliquen las normas técnicas y legislativas aplicables. Estos manuales abordarán la gobernanza, trazabilidad, entrenamiento y modelado de algoritmos, transparencia, explicabilidad, gestión de datasets y riesgos, y evaluación de impactos. Finalmente, en cuarto lugar, se prevé desarrollar una herramienta software para la autoevaluación del cumplimiento de los requisitos del Sello, que automatice y facilite la autoevaluación, garantice la persistencia y seguridad de la información, y ofrezca informes visuales sobre el nivel de madurez de IA de la empresa.

El plan trazado en España estaba en ejecución, pero el cambio de Gobierno en 2023 parece haber influido en esta cuestión. La nueva Estrategia Nacional de Inteligencia Artificial adoptada en mayo de 2024²⁶ omite casi por completo referencias a sellos o certificados no vinculados al ámbito de la sostenibilidad.²⁷

Ya por cuanto a la experiencia de desarrollo de herramientas o sistemas de certificación en España, destaca la iniciativa de *Adigital* con su «Certificación de Transparencia Algorítmica», www.transparenciaalgoritmica.es, lanzada en enero de 2024, que evalúa la transparencia y explicabilidad en el uso de algoritmos por parte de las empresas en España. Se puede acceder a diversas concreciones en la plataforma.²⁸ Los sistemas evaluados son de alto riesgo y el cliente debe presentar evidencias (documentación e información) que justifiquen la valoración. No se utilizan herramientas software en el proceso, y la evaluación puede ser iterativa hasta alcanzar una puntuación determinada que permita obtener el certificado de transparencia. En el verano de 2023, se realizó un piloto con empresas como Adevinta (InfoJobs), Holaluz y Shakers para probar el sello en entornos de alto impacto como la empleabilidad y la infraestructura crítica.²⁹ La evaluación se centra en el producto y no en los procesos ni en el sistema de gestión. En principio no entran a valorar el cumplimiento legal (como el RGPD). A finales de octubre de 2023, presentaron el certificado en Bruselas junto con las tres empresas del piloto.

26. ENIA, mayo 2024, https://portal.mineco.gob.es/es-es/digitalizacionIA/Documents/Estrategia_IA_2024.pdf

27. Así, en la «Palanca 2: Generar Capacidades de Almacenamiento en Condiciones de Sostenibilidad», en concreto, «Iniciativa 2.3. Sello y ecosistema en torno a la IA sostenible». Sin mayor concreción se hace referencia a la importancia de la AESIA para los códigos en el concreto ámbito de la IA generativa.

28. https://www.adigital.org/media/policy-brief_ai-transparency-and-ethics-certifications.pdf

29. <https://www.adigital.org/actualidad/adigital-arranca-su-certificacion-de-transparencia-algoritmica-con-las-tres-primeras-empresas-acreditadas/>

El modelo de certificación de la *Fundación Éticas*,³⁰ dedicada a la ética en la inteligencia artificial, evalúa la implementación de principios éticos en los sistemas de IA.³¹ Este marco está dirigido a empresas que desean certificar que sus procesos y productos cumplen con criterios éticos estrictos, incluyendo transparencia en algoritmos, equidad en resultados, protección de datos personales, responsabilidad en decisiones automatizadas y capacidad de explicación de procesos de IA. Tiene un enfoque inicial en Europa, pero abierto a organizaciones de todo el mundo. Pese a no estar disponible información concreta del sistema, la certificación incluye elementos sobre la transparencia en los algoritmos, la equidad en los resultados, la protección de los datos personales, la responsabilidad en la toma de decisiones automatizadas y la capacidad de explicación de los procesos de IA. Además, se requiere la implementación de mecanismos para la supervisión continua y la evaluación del impacto ético de los sistemas.

IV. LAS VARIADAS INICIATIVAS Y HERRAMIENTAS DE CERTIFICACIÓN O SELLOS DE INTELIGENCIA ARTIFICIAL EUROPEAS E INTERNACIONALES

Las posibilidades de certificación de IA en los próximos años son numerosas y diversas. Podrán coexistir en el mercado modelos de certificación impulsados por gobiernos, colaboraciones público-privadas y modelos privados. Estos sistemas se centrarán en sectores específicos como el empresarial, público, educativo, salud, medioambiental, inclusión, medios de comunicación e IA generativa. También se desarrollarán productos centrados en el cumplimiento de trazabilidad, transparencia, supervisión, calidad y gobernanza de datos, con sellos de diferentes niveles de exigencia.

Existen ya variadas soluciones y herramientas de certificación. La OCDE tiene un repertorio completo que, en mayo de 2024,³² incluye más de 30 modelos de certificación de IA.³³ Ahora se señalan sólo algunos que en su momento han destacado, si bien no es sencillo saber el grado de actualización.

30. <https://eticas.ai/guide-to-algorithmic-auditing>

31. <https://eticas.org/>

32. <https://oecd.ai/en/catalogue/tools?approachIds=3&approachIds=2&approachIds=1&toolTypeIds=20&toolTypeIds=21&orderBy=dateDesc&toXLS=null&page=1>

33. Los resultados en dicha base de datos son 32 en mayo 2024: AI Trust Standard & Label; AIShield AI Security Product; Algorithmic Transparency Certification for Artificial Intelligence Systems; Building Trust in Artificial Intelligence; CounterGen; D-Seal; Digital Trust Label; Ethical Problem Solving; Evaluate Library and Evaluation on the Hub (Hugging Face); FRR Quality Mark for (AI Based) Robotics; Giskard; GRACE; Holistic AI Audits; Holistic AI Bias Audits; Holistic AI Governance, Risk and Compliance Platform; Holistic AI Open Source Library; Holistic AI risk mitigation roadmaps; Human-Computer Trust Scale (HCTS); IEEE CertifAIEd; KomplyAi; Model Cards; Naaia; OneTrust AI Governance; Orthrus; SAIF CHECK; Saimple; SECure: A Social and Environmental Certificate for AI Systems; The Certification as a Mechanism for Control of Artificial Intelligence in Europe; The Citrusx Platform; TÜV for Artificial Intelligence; Zupervise

Así, además de las iniciativas de ISO o NIST, de la máxima relevancia, se mencionan ahora: el modelo *ALTAI*, iniciativa *Future-AI*, *capAI*, *AI Safety Institute*, *Ada Lovelace Institute*, iniciativa *RIAL*, Certificado de *Fairly Trained*, *German AI Association*, *AI Cloud Service Compliance Criteria Catalogue (AIC4)*, *IEEE CertifAIEd*, sello de *Aprobación de la IA de la KI Bundesverband*, modelo de certificación *Towards Auditable AI Systems*, *Denmark's new labelling program for IT security and responsible use of data* o el *Responsible Artificial Intelligence Institute (RIA Institute)*.

Cabe destacar el modelo *ALTAI*, conocido por las Directrices éticas para una IA fiable del grupo de expertos de la UE, con su exhaustiva lista de evaluación de un modelo de ética en el diseño³⁴. Este modelo, desarrollado por el vicepresidente del AI HLEG y su equipo en el *Insight Center for Data Analytics* de la University College Cork³⁵ guía a los desarrolladores e implementadores de IA mediante una lista de verificación accesible y dinámica, centrada en siete requisitos clave: agencia humana y supervisión, robustez técnica y seguridad, privacidad y gobernanza de datos, transparencia, diversidad, no discriminación y equidad, bienestar ambiental y social, y responsabilidad.³⁶

Paralelamente al sistema *ALTAI*, en el ámbito de la salud destaca la iniciativa *Future-ai*³⁷. El mismo incluye y desarrolla un sistema de *checklist* de evaluación ética de la IA para salud, con preguntas concretas y acciones que abarcan siete etapas del desarrollo de IA: conceptualización clínica, recopilación de requisitos, diseño técnico, selección y preparación de datos, implementación y optimización de IA, evaluación y despliegue y monitoreo de IA. Incluye elementos sobre la justicia, universalidad, trazabilidad, usabilidad, robustez y explicabilidad, proporcionando ejemplos de medidas de mitigación para minimizar los riesgos de los algoritmos de IA en salud³⁸. Se ofrece un marco integral que ayuda a los desarrolladores y clínicos a crear y evaluar herramientas de IA médica de manera sistemática. Todo ellos, en colaboración entre investigadores, desarrolladores y profesionales médicos para abordar los desafíos éticos y técnicos en la IA médica.

Investigadores de la Universidad de Oxford con Floridi han desarrollado *capAI*,³⁹ una herramienta diseñada para realizar la evaluación de conformidad de los sistemas de IA según el Reglamento IA. *CapAI* ofrece directrices prácticas para convertir principios éticos en criterios verificables, facilitando el diseño, desarrollo, implementación y uso ético de la IA. Los requisitos de *capAI* incluyen evaluación de riesgos, protección de datos, transparencia en los algoritmos y responsabilidad en decisiones automatizadas, con un enfoque en la capacidad de explicación de los procesos de IA y mecanismos de supervisión continua. *CapAI* se está validando con empresas.

34. HLEG-Comisión Europea, *Directrices éticas para una IA fiable*, 2019, *Directrices éticas para una IA fiable*, 2019, <https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1> en especial Capítulo III y listado, pp. 33-41.

35. <https://www.ucc.ie/en/compsci/research/insight/>

36. <https://futurium.ec.europa.eu/en/european-ai-alliance/pages/welcome-altai-portal>

37. <https://future-ai.eu/>

38. Un detalle de especificaciones y elementos en <https://future-ai.eu/checklist/>

39. Floridi, L. y otros *capAI — A Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act*, marzo, 2022, <http://dx.doi.org/10.2139/ssrn.4064091>

Es bien conocida la Recomendación sobre la ética de la inteligencia artificial de la UNESCO de noviembre de 2021.⁴⁰ Pues bien, cabe destacar en 2023 la difusión de una metodología que incluye indicadores cualitativos y cuantitativos agrupados en diversas dimensiones que permiten una evaluación comprensiva del estado de preparación de cada país para la implementación ética de la IA.⁴¹ Se incluye la evaluación de diversas dimensiones: la jurídica-regulatoria evalúa la capacidad de los Estados para implementar marcos regulatorios que aseguren la protección de datos, privacidad, y la igualdad de género, entre otros aspectos. La social y cultural aborda la inclusión, diversidad, y la confianza pública en la IA, así como las políticas medioambientales y de sostenibilidad. La científica-educativa examina el nivel de investigación y desarrollo en IA, incluyendo la formación y las oportunidades educativas. La dimensión económica evalúa el tamaño y la fuerza del ecosistema de IA en el país, incluyendo el mercado laboral y la inversión en tecnologías de IA y la técnica e infraestructural analiza la infraestructura técnica y la conectividad necesaria para el desarrollo y la aplicación de la IA. Este documento también define la composición del equipo nacional evaluador y detalla cómo debe culminar la evaluación en un informe nacional y una hoja de ruta para desarrollar capacidades y mejorar políticas y marcos regulatorios.

La norma *ISO/IEC 42001:2023 — Information technology Artificial intelligence Management system* es una norma internacional desarrollada por el Comité Técnico ISO/IEC JTC 1/SC 42.⁴² Este estándar, el primero de su clase a nivel mundial, proporciona directrices detalladas para gestionar riesgos y seguridad en el desarrollo e implementación de sistemas de IA. Dirigida a organizaciones de cualquier tamaño, que estén involucradas en el desarrollo, implementación y gestión de sistemas de IA, ayuda a gestionar sistemas de IA de manera segura y eficiente, cumpliendo con los más altos estándares de calidad y transparencia. Esta norma está dirigida a organizaciones de cualquier tamaño. Con un ámbito de aplicación internacional, similar a otras normas ISO, ISO/IEC 42001 incluye directrices sobre la gestión de riesgos, la seguridad de la información y el cumplimiento de estándares de calidad en la implementación de sistemas de IA. Las normas ISO de sistemas de gestión son reconocidas a nivel global por su rigurosidad y contribución a la mejora continua de las organizaciones.

Hay que hacer un seguimiento cercano de los desarrollos desde el NIST de Estados Unidos. El 29 de abril de 2024, el NIST presentó un nuevo borrador del *Marco de Gestión de Riesgos de IA (AI RMF)*.⁴³ Este marco ayuda a las organizaciones a identificar, evaluar y gestionar riesgos asociados con los sistemas de IA. Participaron más de 2500 miembros en el grupo de trabajo público de IA generativa, destacando 12 riesgos y más de 400 acciones que los desarrolladores pueden implementar. En 2024, se difundió un *Borrador del Perfil de IA Generativa* para la identificación y gestión de riesgos de la IA generativa.

40. <https://www.unesco.org/es/artificial-intelligence/recommendation-ethics>

41. UNESCO, *Metodología de evaluación del estado de preparación: una herramienta de la Recomendación sobre la Ética de la Inteligencia Artificial*, UNESCO, 2023, https://unesdoc.unesco.org/ark:/48223/pf0000385198_spa

42. <https://www.iso.org/standard/81230.html>

43. <https://www.nist.gov/itl/ai-risk-management-framework>
<https://doi.org/10.6028/NIST.AI.100-1>

Desde el Reino Unido destacan los desarrollos desde el *AI Safety Institute*⁴⁴, Instituto de Seguridad de la IA es la primera organización respaldada por el gobierno del Reino Unido dedicada a la seguridad de la inteligencia artificial, en fase de desarrollo. En aquel país, *Ada Lovelace Institute* es una de las más destacadas organizaciones de la inteligencia artificial y las tecnologías emergentes.⁴⁵ Su modelo de certificación⁴⁶ evalúa diversas dimensiones de los sistemas de IA, incluyendo la transparencia, la equidad, la privacidad y la rendición de cuentas. La certificación está dirigida a empresas y organizaciones con un enfoque inicial en Europa, pero accesible a organizaciones de todo el mundo. Los criterios de certificación incluyen la transparencia en los algoritmos, la equidad en los resultados, la protección de los datos personales, la responsabilidad en la toma de decisiones automatizadas y la capacidad de explicar los procesos de IA. Además, se requiere la implementación de mecanismos para la supervisión continua y la evaluación del impacto ético de los sistemas.

La iniciativa *RIAL*, generada por un equipo internacional desde 2019,⁴⁷ fomenta la adopción de restricciones de uso en las licencias para mitigar riesgos y daños causados por la IA en la industria. Las licencias *RIAL*⁴⁸ incluyen cláusulas de uso conductual que restringen y controlan aplicaciones tecnológicas de IA. Entre ellas, la Licencia de Código Fuente *RIAL* permite compartir el código bajo términos responsables; la Licencia de Modelo *RIAL* establece limitaciones en el uso y distribución de modelos de IA; y la Licencia de Datos *RIAL* asegura un uso ético y responsable de los conjuntos de datos.⁴⁹

El programa y sello *IEEE CertifAIEd* ayuda a las organizaciones a abordar aspectos esenciales de transparencia, responsabilidad, sesgo algorítmico y privacidad en sus sistemas de IA. Se establecen estándares⁵⁰ y criterios éticos que incluyen: transparencia y valores integrados en el diseño del sistema; responsabilidad y autonomía del sistema con capacidades de aprendizaje; prevención de errores sistemáticos y comportamientos indeseados; y protección de la privacidad. Además, el programa proporciona un «ecosistema de formadores, evaluadores y certificadores».

El Certificado *Fairly Trained*, otorgado por una organización europea sin fines de lucro, se enfoca en modelos de IA generativa con ámbito de aplicación internacional. Evalúa y certifica productos de inteligencia artificial para garantizar que sus modelos de entrenamiento de datos respeten los derechos de los creadores y se obtengan de manera justa. La plataforma proporciona información detallada sobre el acceso al código. La etiqueta *Fairly Trained* se otorga a las empresas que demuestran el uso de

44. <https://www.gov.uk/government/publications/ai-safety-institute-overview/introducing-the-ai-safety-institute#box-1>

45. <https://adalovelaceinstitute.org/>

46. Los detalles en https://www.adalovelaceinstitute.org/wp-content/uploads/2021/12/ADA_Technical-methods-regulatory-inspection_report.pdf

47. <https://www.licenses.ai/>

48. El marco teórico se puede seguir en FAccT de ACM 2022 *Behavioural-use Licensing for Responsible AI*, y la necesidad de estandarización en *On the Standardization of Behavioral Use Clauses and Their Adoption for Responsible Licensing of AI*

49. <https://www.licenses.ai/ai-licenses>

50. Las especificaciones ontológicas en <https://engagestandards.ieee.org/ieeecertifai.html>

datos de entrenamiento éticos y respetuosos con los derechos de autor, promoviendo así la equidad en la IA. Sus requisitos clave se centran en asegurar que todos los datos se obtengan de manera justa y respetuosa con los derechos de los creadores.

En Alemania ha habido diversas iniciativas. La *German AI Association*,⁵¹ que incluye miembros como empresas y expertos en inteligencia artificial, es la responsable del *Sello de Aprobación de la IA de la KI Bundesverband*. Este sello evalúa la calidad y responsabilidad de los sistemas de IA desarrollados por sus miembros. El ámbito territorial del sello está enfocado a Alemania. Los requisitos clave del sello⁵² incluyen trabajar con criterios establecidos por la asociación, que abarcan la ética y la transparencia en los sistemas de IA.

El modelo de certificación *Towards Auditable AI Systems*⁵³ se presenta como un enfoque integral para evaluar y certificar sistemas de IA. Desarrollado por el *Fraunhofer Heinrich Hertz Institute* (HHI) junto a la *Asociación TÜV* y la *Oficina Federal de Seguridad de la Información* (BSI), el modelo *Towards Auditable AI Systems* ha producido dos documentos técnicos: una hoja de ruta en 2021 para examinar los modelos de IA a lo largo de su ciclo de vida⁵⁴ y una «Matriz de preparación para la certificación» (CRM) en 2022.⁵⁵

El programa va dirigido a desarrolladores y auditores de sistemas de IA con ámbito internacional e incluye la evaluación de procedimientos documentados, modelos de entrenamiento y prácticas de implementación. También en Alemania, el *AI Cloud Service Compliance Criteria Catalogue* (AIC4)⁵⁶ evalúa la seguridad y cumplimiento de los servicios en la nube que incorporan IA. Este catálogo, desarrollado por la Oficina Federal de Seguridad de la Información (BSI), establece criterios específicos que deben cumplir los proveedores de servicios en la nube para garantizar que sus soluciones de IA sean seguras, confiables y cumplan con las normativas vigentes. Su ámbito de aplicación es principalmente en Alemania, pero puede ser adoptado por organizaciones internacionales que deseen cumplir con los altos estándares de seguridad y cumplimiento alemanes. El catálogo abarca la gestión de riesgos, protección de datos, seguridad de la información, transparencia en los procesos de IA y cumplimiento de normativas de seguridad y privacidad.

El (nuevo programa de etiquetado de Dinamarca para la seguridad de TI y el uso responsable de datos)⁵⁷ fue fundado por un consorcio independiente de partes interesadas en Dinamarca. Su propósito es evaluar si una empresa cumple con

51. <https://ki-verband.de/en/projects>

52. Algún detalle en https://www.am.ai/171feadfe65186a2f4d42891383a58d7/KIBV_Guetesiegel_190302_o.pdf

53. <https://www.hhi.fraunhofer.de/en/departments/ai/technologies-and-solutions/auditing-and-certification-of-ai-systems.html>

54. BSI, «Towards Auditable AI Systems: Current status and future directions» mayo de 2021. También, BSI, *Towards Auditable AI Systems AI Cloud Service*, https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/Towards_Auditable_AI_Systems_2022.pdf?__blob=publicationFile&v=4

55. «Towards Auditable AI Systems: From Principles to Practice» de mayo 2022.

56. BSI, *Compliance Criteria Catalogue* (AIC4), https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/CloudComputing/AIC4/AI-Cloud-Service-Compliance-Criteria-Catalogue_AIC4.pdf?__blob=publicationFile&v=4

57. <https://d-seal.eu/>

ciertos criterios de seguridad y uso responsable de datos, con el número de criterios variando según el nivel de etiqueta deseado.⁵⁸ El objetivo principal del programa es visibilizar si una empresa tiene una buena seguridad de TI y un uso responsable de los datos. Está dirigido a ciudadanos y consumidores de Dinamarca en su relación con las empresas, abarcando un ámbito territorial nacional. Los requisitos clave del programa incluyen un número de controles por criterio que varía y se actualiza regularmente. Además, las empresas deben registrarse y solicitar la aplicación del sello.

Finalmente, entre otras iniciativas, el *Responsible Artificial Intelligence Institute* (RIA Institute) ofrece un programa de certificación de IA responsable.⁵⁹ Inicialmente, este programa se centra en los sistemas de IA desarrollados en América del Norte, pero con un alcance potencialmente global. La certificación⁶⁰ evalúa seis dimensiones principales: operaciones del sistema, explicabilidad e interpretabilidad, responsabilidad, protección del consumidor, equidad y ausencia de sesgos, y robustez. El proceso de certificación incluye una revisión, desarrollo del marco de implementación, pruebas de evaluación y ajustes, formación y calibración.

V. PARA CONCLUIR

En este estudio se ha analizado la regulación del RIA respecto de los sistemas de IA que esencialmente no regula, es decir, los sistemas que no son de alto riesgo. En cualquier caso y como punto de partida, estos sistemas están sujetos a otras regulaciones aplicables, como la seguridad de productos o la protección de datos. Aunque el RIA se centra en los sistemas de IA de alto riesgo, imponiendo estrictas obligaciones, para los sistemas que no lo son impulsa la creación y desarrollo de un ecosistema de códigos de conducta, sellos y sistemas de certificación en el ámbito de la IA en la UE. Cabe señalar que esto se alinea con el resto del mundo, donde se priorizan los códigos de autorregulación y sistemas normativos más blandos que el RIA, como el Código de Hiroshima acordado en 2023 por el G7, en EEUU o el Reino Unido. En los próximos años, se desarrollarán certificaciones de variado alcance, origen, sector, naturaleza e intensidad de exigencias. El tiempo y el mercado determinarán la utilidad y éxito de estos instrumentos voluntarios para los sistemas que no son de alto riesgo. Estos códigos, sellos y certificaciones serán un sólido soporte para garantizar la calidad y seguridad de los sistemas, su conformidad con principios éticos fundamentales como la transparencia, la responsabilidad y la prevención de sesgos algorítmicos. Además, como apunta el RIA, pueden desempeñar un papel importante en la sostenibilidad, la inclusión y, en general, fomentar la confianza de los usuarios y consumidores en las tecnologías de IA.

Una vez expuesto el alcance del artículo 95 RIA, se ha señalado la oportunidad perdida que ha supuesto la no regulación de unos principios generales que propuso el Parlamento de la UE. Estos principios hubieran regido para todos los sistemas de IA, no sólo los de alto riesgo. Es bien sabido el importante papel que han jugado los principios del artículo 5 del RGPD. El reconocimiento normativo de estos principios,

58. Algún detalle en <https://d-seal.eu/criteria/>

59. <https://www.responsible.ai/how-we-help/#certification>

60. Los detalles en <https://www.responsible.ai/wp-content/uploads/2024/02/RIAI-Certification-Guidebook.pdf>

ya tan conocidos en el ámbito de la ética de la IA, podría haber tenido un gran potencial como principios y reglas jurídicas aplicables a todo tipo de IA.

Por cuanto al desarrollo de los sellos y certificaciones de IA, se trata de un terreno aún incipiente en el que el Gobierno español apostó pronto en su ENIA de 2020. Quizá demasiado pronto, al punto de que la última ENIA de mayo de 2024 parece haber olvidado la iniciativa de un Sello español de IA que se tenía prevista y, al menos en teoría, bien planificada para su ejecución.

El estudio ha analizado más de una treintena de iniciativas y herramientas de certificación de IA, tanto a nivel europeo como internacional. Entre ellas, se ha descrito más de una docena de las más relevantes o conocidas. La revisión de las mismas permite apreciar la diversidad y riqueza de enfoques disponibles para abordar los modelos de certificación voluntaria de la IA para los sistemas que no son de alto riesgo.

Estos esquemas de certificación y códigos de conducta de IA puede pensarse que quedarán a la sombra de las obligaciones *duras* para los sistemas de alto riesgo que establece el RIA y se han de desarrollar con criterios, normas armonizadas más concretas y especificaciones técnicas. No obstante, en el futuro va a ser fundamental continuar desarrollando y perfeccionando este ecosistema de las certificaciones voluntarias y es posible que acaben adquiriendo una gran presencia efectiva. Por ello, será importante fomentar la colaboración entre gobiernos, organizaciones privadas y la sociedad civil para asegurar que los sistemas de IA se desarrollen y desplieguen para hacer más efectiva la IA responsable y ética por la que apuesta la UE en su RIA, también para los sistemas que no de alto riesgo.

El artículo 50 del Reglamento y las obligaciones de transparencia de los proveedores y responsables del despliegue de determinados sistemas de inteligencia artificial

AGUSTÍ CERRILLO I MARTÍNEZ

Catedrático de Derecho Administrativo de la Universitat Oberta de Catalunya

I. LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL DE RIESGO LIMITADO

El RIA clasifica los sistemas de inteligencia artificial (IA) en función del riesgo que puedan entrañar a los intereses públicos y a los derechos fundamentales protegidos por el Derecho de la Unión. En efecto, como dispone el considerando 26 «Con el fin de establecer un conjunto proporcionado y eficaz de normas vinculantes para los sistemas de IA, es preciso aplicar un enfoque basado en los riesgos claramente definido, que adapte el tipo y contenido de las normas a la intensidad y el alcance de los riesgos que puedan generar los sistemas de IA en cuestión». A partir de este enfoque, el RIA prohíbe el uso de determinados sistemas de IA (artículo 5 RIA) y clasifica otros como sistemas de alto riesgo (artículo 6 RIA) por su impacto en los intereses públicos de la UE o a los derechos fundamentales.

Además, el RIA advierte que determinados sistemas de IA destinados a interactuar con personas físicas o a generar contenidos pueden generar otros riesgos específicos como la suplantación, el engaño o la manipulación de las personas. Como advierte Peguera, no nos encontramos en sentido estricto ante una categoría específica de riesgo¹, si bien el RIA prevé que las personas usuarias o destinatarias de los resultados de estos sistemas de IA deben poder ser conscientes de que están tratando con sistemas de IA o que los resultados obtenidos han sido generados de manera artificial. En esta dirección, el artículo 50 RIA prevé distintas obligaciones de transparencia que serán analizadas en este capítulo. De este modo, se persigue que cada persona que esté en contacto con estos sistemas o con los resultados que generan pueda conocer tales circunstancias, adoptar decisiones fundamentadas o evitar una situación determinada.

1. Peguera Poch, M., *La propuesta de Reglamento de IA: Una intervención legislativa insoslayable en contexto de incertidumbre*, en Peguera Poch, M., *Perspectivas regulatorias de la inteligencia artificial en la Unión Europea*, Reus, Madrid, 2023.

En las próximas páginas, en primer lugar, se expone la evolución que ha experimentado la regulación de las obligaciones de transparencia de determinados sistemas de IA desde la propuesta formulada por la Comisión Europea en 2021 hasta el texto finalmente publicado en el DOUE. A continuación, se describen los distintos sistemas de IA afectados y se analizan las obligaciones de transparencia previstas para cada uno de ellos. Finalmente, se concluye con unas reflexiones finales.

II. EVOLUCIÓN, TRAMITACIÓN Y CONTENIDO FINAL DEL ARTÍCULO 50 REGLAMENTO

La previsión de unas obligaciones de transparencia para determinados sistemas de IA ha estado prevista en el RIA desde la propuesta presentada por la Comisión Europea en 2021 [COM(2021) 206 final].

En efecto, ya en aquel primer texto se incluyó el título IV, conformado únicamente por el artículo 52, en el que se preveían explícitamente tres obligaciones cuya regulación no podía afectar a las obligaciones de transparencia previstas con carácter general en la regulación de los sistemas de alto riesgo (título III).

Según la explicación que acompañaba la propuesta, el artículo 52 se centraba en sistemas de IA que podían generar riesgos específicos de manipulación por lo que se disponían obligaciones específicas de transparencia que se concretaban en la obligación de informar sobre esta circunstancia con el fin de que la persona afectada pudiese adoptar decisiones fundamentadas o evitar una determinada situación.

En primer lugar, se establecía la obligación de los proveedores de garantizar que los sistemas de IA destinados a interactuar con personas físicas estuviesen diseñados de manera que pudiesen estar informadas de que estaban interactuando con un sistema de IA. En segundo lugar, se incluía la obligación de los usuarios de sistemas de reconocimiento de emociones o de categorización biométrica de informar sobre su funcionamiento a las personas físicas expuestas a ellos. Igualmente, en este caso, la propuesta incorporaba alguna excepción. En tercer lugar, se preveía la obligación de los usuarios de determinados sistemas que generasen o manipulasen el contenido de imágenes, sonidos o vídeos que pudiesen llevar erróneamente a pensar que son auténticos o verídicos de que habían sido generados de manera artificial o manipulada. Por último, los tres casos las obligaciones se establecían con algunas limitaciones y excepciones.

Una de las principales novedades que introdujo el Consejo, fue en relación con las excepciones relativas a la primera obligación². En la propuesta de la Comisión se preveía como limitación de la obligación de informar «en las situaciones en las que esto resulte evidente debido a las circunstancias y al contexto de utilización». En cambio, el Consejo lo concretó disponiendo que «en las situaciones en las que esto resulte evidente desde el punto de vista de una persona jurídica que esté razonablemente informada, observadora y circunspecta, dadas las circunstancias y el contexto de utilización». Esta ha sido la redacción finalmente adoptada. En relación con las excepciones a la primera obligación, el Consejo también incorporó que los sistemas

2. Según el documento aprobado el 25 de noviembre de 2022. Accesible en: <https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/es/pdf> (última consulta: febrero de 2024).

incluidos en ella —fines de detección, prevención, investigación o enjuiciamiento de infracciones penales— funcionasen con sujeción a las correspondientes salvaguardas de los derechos y libertades de terceros. En relación con la segunda obligación, el Consejo propuso distinguir en dos apartados la regulación relativa a los sistemas de reconocimiento de emociones y los de categorización biométrica, aunque, en términos generales, el alcance de la obligación era el ya previsto en la propuesta de la Comisión con relación a cada uno de estos sistemas añadiendo únicamente, como en el caso de la excepción a la primera obligación, que se deberían respetar los derechos y libertades de terceros. Por último, respecto a la tercera obligación, el Consejo únicamente propuso un cambio de redacción en relación con el alcance del derecho a la libertad de expresión. Asimismo, con carácter general, en relación con las tres obligaciones, se propuso por parte del Consejo que la información se facilitase de «modo claro y visible a más tardar con ocasión de la primera interacción o exposición». Finalmente, también se sugirió incluir en el reglamento que las obligaciones previstas en el artículo 52 no solo no afectarían a lo dispuesto en el título III tal y como ya contemplaba la propuesta de la Comisión, sino que tampoco afectarían a «otras obligaciones de transparencia para los usuarios de sistemas de IA establecidas en el Derecho de la Unión o nacional».

Por su parte, el Parlamento Europeo también presentó distintas enmiendas a la propuesta elaborada por la Comisión³. En relación con la primera obligación propuso concretar que se informase «de manera clara, inteligible y oportuna». Asimismo, las enmiendas del Parlamento sugirieron que «esta información también revelará qué funciones se encuentran habilitadas por la IA, si existe vigilancia humana y quién es responsable del proceso de toma de decisiones, así como los derechos y procesos existentes que, de conformidad con el Derecho de la Unión y nacional, permiten a las personas físicas o a sus representantes oponerse a que se les apliquen dichos sistemas y solicitar reparación judicial contra las decisiones adoptadas por los sistemas de IA o los perjuicios causados por ellos, incluido su derecho a solicitar una explicación» (enmienda 484). Por lo que se refiere a la segunda obligación, el Parlamento Europeo también propuso que la información fuese oportuna, clara e inteligible y que, en el caso de tratamiento de datos biométricos, se debería obtener el consentimiento de la persona física expuesta a él (enmienda 485). Respecto a la tercera obligación, el Parlamento Europeo sugirió que se previese la obligación de informar cuando fuese posible de la persona física o jurídica que generó o manipuló el contenido. También se propuso que se tuviese que etiquetar, de acuerdo con el estado de la técnica y las normas y especificaciones armonizadas pertinentes, el contenido no auténtico para que pueda resultar claramente visible (enmienda 486). Además, el Parlamento Europeo propuso que la tercera obligación no fuese exigible cuando la generación o manipulación estuviese autorizada por ley o, en el caso de que el contenido formase parte de una obra cinematográfica claramente creativa, satírica, artística o ficticia, de imágenes de videojuegos y de obras o formatos análogos, no obstaculizase la presentación de la obra (enmienda 487). Por último, se propuso que la información se facilitase

3. Según texto aprobado el 14 de junio de 2023. Accesible en: https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_ES.html (última consulta: febrero de 2024).

a las personas físicas de manera accesible con ocasión de la primera interacción o exposición (enmienda 488).

Como tendremos la oportunidad de analizar en detalle en las próximas secciones, el texto finalmente adoptado —que finalmente es el artículo 50 RIA— ha mantenido el espíritu de la propuesta de la Comisión, pero ha incluido la mayoría de las propuestas y enmiendas formuladas. Principalmente, las realizadas por el Consejo. Pero, tal vez, la principal novedad incorporada durante la tramitación responde a la decisión de regular los sistemas de inteligencia artificial de propósito general (GPAI) lo que se ha concretado en la inclusión de una obligación de transparencia específica respecto a aquellos sistemas que sean capaces de generar audios, imágenes, vídeos o textos sintéticos para que informen sobre ello.

III. EL ALCANCE DE LAS OBLIGACIONES DE TRANSPARENCIA PREVISTAS EN EL ARTÍCULO 50 REGLAMENTO

El artículo 50 RIA regula distintas obligaciones de transparencia que deben cumplir los proveedores y responsables del despliegue de determinados sistemas de IA. Estas obligaciones se aplican a un conjunto limitado de sistemas de IA cuyo uso si bien puede reportar numerosos beneficios también pueden entrañar algunos riesgos que potencialmente pueden tener un amplio impacto.

Como iremos constatando a lo largo de las próximas páginas, estos sistemas de IA inherentemente o de manera directa o exclusiva no generan riesgos contra intereses públicos o respecto la salud, la seguridad o los derechos fundamentales. Pero sí que pueden ser utilizados de manera que tengan una incidencia negativa en la sociedad por facilitar la desinformación, manipulación, el fraude, el engaño o, simplemente, la confusión.

El RIA ha optado por no prohibir ni restringir su uso sino por advertir a sus usuarios o a las personas que utilicen o conozcan sus resultados sobre la utilización de estos sistemas de IA. En relación con esta opción legislativa, algunos autores se han planteado si será suficientemente adecuada para evitar un impacto negativo en los intereses públicos o una violación de los derechos fundamentales. Por ello, han sugerido que el RIA debería haber establecido una regulación más estricta, por ejemplo, una evaluación de conformidad o de impacto sobre los derechos fundamentales⁴. Tal vez es pronto para saberlo y será necesario evaluar la efectividad de esta opción y, en su caso, actualizar la regulación para conseguir las finalidades previstas.

El RIA ha previsto que las obligaciones de transparencia aplicables a los sistemas de IA que serán descritos a continuación no afectan a la posible aplicación de los requisitos y obligaciones previstos en la propia norma para los sistemas de IA de alto riesgo (artículo 50 apartado 6). Asimismo, ha reconocido que tampoco afectan a otras obligaciones de transparencia que puedan preverse por los Estados Miembros o por la propia Unión Europea (artículo 50 apartado 6). Por ejemplo, las que puedan

4. Barkane, I., «*Questioning the EU proposal for an Artificial Intelligence Act: The need for prohibitions and a stricter approach to biometric surveillance 1*», *Information Polity*, núm 27 (2022).

derivarse de la legislación de transparencia cuando los sistemas de IA son utilizados por Administraciones públicas.

Por último, con carácter general y aplicable a los distintos supuestos que se analizan, el apartado 5 del artículo 50 RIA prevé que la información que faciliten los proveedores y responsables del despliegue en cumplimiento de las obligaciones de transparencia debe ser accesible y comprensible. Por este motivo debe facilitarse de manera clara y distinguible. Además, debe ser oportuna. De esta manera, es necesario que se facilite a más tardar con ocasión de la primera interacción o exposición. Por último, la información que se facilite deberá ajustarse a los requisitos de accesibilidad aplicables previstos en la Directiva (UE) 2016/2102 del Parlamento Europeo y del Consejo, de 26 de octubre de 2016, sobre la accesibilidad de los sitios web y aplicaciones para dispositivos móviles de los organismos del sector público y la Directiva (UE) 2019/882 del Parlamento Europeo y del Consejo, de 17 de abril de 2019, sobre los requisitos de accesibilidad de los productos y servicios.

IV. SISTEMAS DE INTELIGENCIA ARTIFICIAL QUE INTERACTÚEN DIRECTAMENTE CON PERSONAS FÍSICAS

1. SISTEMAS INCLUIDOS

El primer apartado del artículo 50 regula las obligaciones de transparencia de los sistemas de IA que interactúan directamente con personas físicas. Estos son sistemas de IA que permiten que las personas puedan relacionarse con dispositivos en lenguaje natural, oral o escrito, de manera que sean capaces de entender el contenido del mensaje y actuar en consecuencia⁵.

El RIA a lo largo de su articulado no incluye ninguna definición de estos sistemas ni tampoco determina las características que deben tener.

Tal vez la más extendida de los sistemas de IA que interactúan con personas son los robots conversacionales o chatbots utilizados por muchas empresas y Administraciones públicas y los asistentes virtuales incorporados en distintos dispositivos⁶. Pero más allá, existen otras aplicaciones de estos sistemas de IA, por ejemplo, para el control remoto de dispositivos aeroespaciales o de vehículos submarinos⁷.

A medida que la calidad de estos sistemas de IA se va incrementando, las personas que interactúan con ellos tienen mayores dificultades para saber si están relacionándose con una persona o con una máquina. En paralelo, se va ampliando la preocupación entre ellas hasta el extremo de generar distintos tipos de rechazo a su uso⁸.

5. Cerrillo I Martínez, A., «Robots, asistentes virtuales y automatización de las administraciones públicas», *Revista Galega de Administración Pública*, núm 61 (2021).
6. Adamopoulou, E. y Moussiades, L., «Chatbots: History, technology, and applications», *Machine Learning with Applications*, núm 2 (2020).
7. Sheridan, T. B., «Human-robot interaction: status and challenges», *Human factors*, núm 58 (2016).
8. Bartneck, C., Belpaeme, T., Eyssel, F., Kanda, T., Keijsers, M. y Šabanović, S., *Human-robot interaction: An introduction*, Cambridge University Press, Cambridge, 2020.

2. ALCANCE DE LAS OBLIGACIONES

En su primer apartado, el artículo 50 RIA establece una obligación de diseñar y desarrollar estos sistemas de manera que se pueda informar sobre el hecho de que la persona física pueda saber que está interactuando con un sistema de IA.

El sujeto obligado son los proveedores de los sistemas de AI que deberán diseñarlos de manera que se pueda cumplir la obligación de facilitar información. Como señalan Veale y Zuiderveen Borgesius, tal vez hubiese sido deseable que el RIA se refiriese no solo a proveedores sino también a los responsables del despliegue para garantizar que en cualquier caso la información llegue a la persona que interactúa con los sistemas de IA⁹. En particular, para aquellos casos en los que el sistema se integra en otro servicio que es el que finalmente recibe el usuario.

El RIA no indica cómo deberá diseñarse el sistema ni tampoco como debe facilitarse la información siendo el proveedor del sistema quién lo determine siempre y cuando se logre la finalidad prevista. Esta ausencia de criterios puede llevar a que sea cada proveedor quien acabe concretando el alcance de la información que facilita lo que eventualmente pueda incidir negativamente en la transparencia¹⁰.

La obligación se limita en los casos en que para una persona física razonablemente informada, atenta y perspicaz, teniendo en cuenta las circunstancias y el contexto de utilización resulte evidente que está relacionándose con un sistema de IA. Los calificativos introducidos en la regulación durante su tramitación subrayan la voluntad de garantizar que los proveedores de sistemas de IA sean especialmente cuidadosos a la hora de prever cómo la información podrá efectivamente llegar a las personas afectadas y no se les acabe trasladando a ellas la responsabilidad de localizar u obtener la información.

Por otro lado, la obligación prevista en el apartado 1 se exceptúa en el caso de los sistemas autorizados por la ley para detectar, prevenir, investigar y enjuiciar delitos a menos que dichos sistemas estén a disposición del público para denunciar infracciones penales. Al respecto no puede desconocerse que, como ya señaló la Comunicación *Generar confianza en la inteligencia artificial centrada en el ser humano* [COM(2019) 168 final] de 8 de abril de 2019, «La IA también puede ayudar a detectar el fraude y las amenazas de ciberseguridad y permite a los organismos encargados de hacer cumplir la ley luchar contra la delincuencia con más eficacia». Mas, también ha advertido la propia Comisión cómo «el uso conocido de tecnologías similares para fines de vigilancia, por parte de empresas públicas o privadas, puede suscitar preocupación y reducir la confianza en la economía digital entre particulares y organizaciones» (Comunicación *Hacia una economía de los datos próspera* COM(2014) 442 final de 2 de julio). En cualquier caso, esta excepción, que también se prevé con relación a otros sistemas de IA contemplados en el artículo 50, deberá sujetarse a las garantías adecuadas para los derechos y libertades de terceros.

9. Veale, M. y Zuiderveen Borgesius, F., «*Demystifying the Draft EU Artificial Intelligence Act-Analysing the good, the bad, and the unclear elements of the proposed approach*», *Computer Law Review International*, núm 22 (2021).

10. Stuurman, K. y Lachaud, E., «*Regulating AI. A label to complete the proposed Act on Artificial Intelligence*», *Computer Law & Security Review*, núm 44 (2022).

V. SISTEMAS DE INTELIGENCIA ARTIFICIAL QUE GENEREN CONTENIDOS SINTÉTICOS

1. SISTEMAS INCLUIDOS

Los sistemas de IA han experimentado una rápida evolución en los últimos tiempos hasta llegar a adquirir, entre otras capacidades, la de generar contenidos sintéticos, es decir, contenidos generados artificialmente. Estos contenidos son tan realistas que una persona no sea capaz de distinguir que han sido creados por un sistema de IA.

Los contenidos sintéticos pueden ser de distinto tipo. En particular, el apartado 2 del artículo 50 se refiere a los sistemas de IA que generen contenido sintético de audio, imagen, vídeo o texto.

Entre estos sistemas de IA que crean contenidos sintéticos se incluyen específicamente los sistemas de IA de uso general. Estos sistemas de IA tienen un gran potencial, pero también plantean numerosos riesgos que se han trasladado a las numerosas discusiones que se han producido a lo largo de la tramitación del RIA. En efecto, como titulaba el semanario Político en marzo de 2023 «ChatGPT rompió el plan de la UE para regular la inteligencia artificial»¹¹. De hecho, la propuesta de la Comisión europea no se refería a los sistemas de inteligencia artificial de uso general¹². Efectivamente, esta propuesta se centraba en los modelos de IA convencionales¹³. No fue hasta la presidencia eslovena en 2021 cuando se incluyó una primera mención que, posteriormente, fue profundizada por la presidencia francesa.

Al margen del análisis que se realiza en otros capítulos en este momento interesa traer a colación que estos sistemas de IA son muy complejos, se entrenan con millones de datos y están conformados con millones de parámetros. Pero su principal característica es que pueden realizar tareas muy distintas, algunas de ellas no previstas inicialmente. Para ello, consumen grandes volúmenes de capacidad de computación y por ende también mucha energía¹⁴.

A pesar de la calidad de los resultados que se pueden obtener y de los usos que se pueden dar a estos sistemas de IA —por ejemplo, en el ámbito de la medicina¹⁵, la planificación urbanística¹⁶, la educación¹⁷, o, incluso en la Administración

-
11. Accesible en: <https://www.politico.eu/article/eu-plan-regulate-chatgpt-openai-artificial-intelligence-act/> (última consulta: marzo de 2024).
 12. Moreira, N. A., Freitas, P. M. y Novais, P., *The AI Act Meets General Purpose AI: The Good, The Bad and The Uncertain*, en Moniz, N., Vale, Z., Cascalho, J., Silva, C. y Sebastião, R., *EPIA Conference on Artificial Intelligence*, Springer, Cham, 2023.
 13. Hacker, P., Engel, A. y Mauer, M., *Regulating ChatGPT and other large generative AI models*, 2023.
 14. OECD, *Measuring the environmental impacts of artificial intelligence compute and applications*, 2022.
 15. Sallam, M., *ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns*, MDPI, 2023.
 16. Wang, D., Lu, C.-T. y Fu, Y., «Towards automated urban planning: When generative and chatgpt-like ai meets urban planning», *arXiv preprint arXiv:2304.03892*, núm (2023).
 17. Grassini, S., «Shaping the future of education: exploring the potential and consequences of AI and ChatGPT in educational settings», *Education Sciences*, núm 13 (2023) <style face="italic">Education Sciences</style>, núm 13 (2023).

pública¹⁸, no dejan de ser textos, imágenes o vídeos generados a partir de una combinación de información basada en probabilidades. Tal y como se ha puesto de manifiesto, estos sistemas, no son capaces de entender el contenido generado (lo que se ha bautizado con la metáfora del *loro estocástico*)¹⁹. Asimismo, estos modelos pueden *alucinar*, es decir, ofrecer resultados muy creíbles o convincentes pero que no responden a los datos de entrenamiento del algoritmo por lo que pueden ser falsos²⁰. Además, estos sistemas de IA no se escapan de los riesgos que entraña en general la IA derivados de la calidad de los datos²¹. Por último, estos sistemas pueden vulnerar los derechos de propiedad intelectual²².

Junto a estos riesgos, como veremos posteriormente, la preocupación generada alrededor de estos sistemas de IA radica en el hecho de que pueden ser utilizados para generar contenidos que por su verosimilitud o hiperrealismo pueden llevar a la manipulación o la desinformación²³.

Por la existencia de todos estos riesgos, pero también por la posibilidad de que surjan otros nuevos a medida que vayan evolucionando los usos —los daños de cisne negro (*black swans damages*) a los que se refiere Kolt²⁴— es por lo que el RIA ha previsto las obligaciones de transparencia respecto a estos sistemas de IA.

2. ALCANCE DE LAS OBLIGACIONES

Las obligaciones de transparencia de los sistemas de IA que generen contenidos sintéticos se dirigen a los proveedores. Nuevamente en este caso, se ha planteado que hubiese sido oportuno prever también alguna obligación a cumplir por los responsables del despliegue²⁵. En efecto, si bien generalmente en los casos en que el responsable del despliegue modifique la finalidad prevista de un sistema de IA que ya haya sido introducido en el mercado o puesto en servicio, de tal manera que se convierta en un sistema de IA de alto riesgo (artículo 25.1.c RIA) deberá considerado como proveedor, en otros casos la modificación puede no ser tan sustancial pero a pesar de ello incidir en cómo el resultado finalmente llegue al usuario final.

18. Huang, J. y Huang, K., *ChatGPT in Government*, en Huang, K., Wang, Y., Zhu, F., Chen, X. y Xing, C., *Beyond AI: ChatGPT, Web3, and the Business Landscape of Tomorrow*, Springer Nature Switzerland, Cham, 2023.

19. Bender, E. M., Gebru, T., Mcmillan-Major, A. y Shmitchell, S., *On the dangers of stochastic parrots: Can language models be too big??*, 2021; Srivastava, V., «*When Stochastic Parrots Learn to Swim: The Regulation of General Purpose Artificial Intelligence in the EU*», núm (2023).

20. Triguero, I., Molina, D., Poyatos, J., Del Ser, J. y Herrera, F., «*General Purpose Artificial Intelligence Systems (GPAIS): Properties, definition, taxonomy, societal implications and responsible governance*», *Information Fusion*, núm 103 (2024).

21. Moreira, N. A., Freitas, P. M. y Novais, P., *The AI Act Meets General Purpose AI: The Good, The Bad and The Uncertain*, op.cit.

22. Lucchi, N., «*ChatGPT: a case study on copyright challenges for generative artificial intelligence systems*», *European Journal of Risk Regulation*, núm (2023).

23. Hacker, P., Engel, A. y Mauer, M., *Regulating ChatGPT and other large generative AI models*, 2023.

24. Kolt, N., «*Algorithmic black swans*», *Washington University Law Review*, núm 101 (2023).

25. Edwards, L., *Regulating AI in Europe: four problems and four solutions*, Ada Lovelace Institute, London, 2022.

Los proveedores del sistema de IA deben velar porque el resultado generado (información de salida del sistema de IA) incluya una marca y que permita detectar que dicho resultado ha sido generado o manipulado de manera artificial.

La marca debe ser legible por máquina, es decir, debe estar en un formato de archivo estructurado que permita a las aplicaciones informáticas identificar, reconocer y extraer con facilidad datos específicos, incluidas las declaraciones fácticas y su estructura interna²⁶. Para ello, el propio apartado prevé que los proveedores deberán utilizar soluciones técnicas que sean eficaces, interoperables, sólidas y fiables en la medida en que sea técnicamente viable. Asimismo, a la hora de definir las soluciones técnicas que utilizarán deberán tener en cuenta las particularidades y limitaciones inherentes a cada tipo de contenido creado (audio, imagen, vídeo o texto), los costes de aplicación y el estado actual de la técnica generalmente reconocido, según se refleje en las normas técnicas pertinentes.

Para ello, el apartado 7 del artículo 50 prevé que la Oficina de AI fomentará y facilitará la elaboración de códigos de prácticas a escala de la Unión para facilitar la aplicación efectiva de esta obligación. Asimismo, se faculta a la Comisión para que pueda adoptar actos de ejecución para aprobar estos códigos de prácticas de acuerdo con lo previsto en el artículo 56 RIA y, si considera que no es adecuado, poder adoptar un acto de ejecución que especifique las normas comunes para la aplicación de dichas obligaciones según lo dispuesto en el artículo 98 RIA.

La información que se facilite debe ser suficientemente clara como para que el destinatario del resultado pueda tener conocimiento de que el contenido ha sido generado o manipulado de manera artificial.

El apartado 2 del artículo 50 RIA prevé como límite a la obligación, aquellos casos en los que los sistemas de IA desempeñen una función de apoyo a la edición estándar o no alteren sustancialmente los datos de entrada facilitados por el responsable del despliegue o su semántica. En estas circunstancias los proveedores del sistema de IA no deberán velar porque la información de salida esté marcada ni que permita detectar que ha sido generada o manipulada de manera artificial.

Por último, como en el caso de los sistemas de IA que interactúen directamente con personas, en el caso de los sistemas que generen contenidos se prevé como excepción de la obligación de transparencia cuando estos sistemas estén autorizados por ley para detectar, prevenir, investigar o enjuiciar infracciones penales.

VI. SISTEMAS DE INTELIGENCIA ARTIFICIAL DE RECONOCIMIENTO DE EMOCIONES

1. SISTEMAS INCLUIDOS

En las últimas décadas, se ha ido avanzando en el desarrollo de sistemas de IA que son capaces de detectar de manera automatizada un elemento inherente a las personas como son las emociones.

26. Artículo 2.13 Directiva (UE) 2019/1024 del Parlamento Europeo y del Consejo, de 20 de junio de 2019, relativa a los datos abiertos y la reutilización de la información del sector público.

Los sistemas de IA de reconocimiento de emociones son un tipo de la llamada computación afectiva, es decir, ordenadores que tienen diversas capacidades relacionadas con las emociones como el reconocimiento, la expresión, la modelización, la comunicación o la reacción²⁷. Los sistemas de IA de reconocimiento de emociones persiguen que las máquinas puedan medir, valorar, predecir o reaccionar a los estados emocionales de las personas a partir de distintos datos extraídos de elementos físicos o fisiológicos que puedan estar en textos, voces, imágenes o vídeos o captarse a través de sensores biométricos²⁸. De este modo, los sistemas de IA de reconocimiento de emociones son capaces de convertir emociones en datos²⁹. No obstante, no permiten que los ordenadores sientan o expresen emociones por sí mismos por lo que son definidos como un tipo de IA débil³⁰.

El RIA define los sistemas de IA de reconocimiento de emociones como aquellos sistemas destinados a distinguir o inferir las emociones o las intenciones de las personas físicas a partir de sus datos biométricos (artículo 3, apartado 39). El RIA se refiere a sistemas de IA de reconocimiento de emociones y no simplemente a sistema de detección. De este modo, los sistemas de IA de detección de emociones que no impliquen el reconocimiento quedarán al margen de la obligación de transparencia previsto en el artículo 50.

Estos sistemas de IA persiguen asociar a una expresión facial, la cadencia al hablar o un movimiento corporal una determinada emoción (por ejemplo, miedo, tristeza, ira, alegría, sorpresa o asco). En esta dirección, el RIA hace referencia a emociones o intenciones como la felicidad, la tristeza, la indignación, la sorpresa, el asco, el apuro, el entusiasmo, la vergüenza, el desprecio, la satisfacción y la diversión. En cambio, excluye los estados físicos (por ejemplo, el dolor o el cansancio) (considerando 18).

Para ello, se examinan señales físicas (como expresiones faciales, el movimiento de ojos o del cuerpo, el discurso o el texto o posturas corporales) o fisiológicas (por ejemplo, la temperatura del cuerpo, el ritmo cardíaco o la cadencia de la respiración) captadas por diferentes sensores.

En la actualidad, son numerosas las numerosas aplicaciones que ya se están dando a estos sistemas de IA en el ámbito de la salud (por ejemplo, para detectar el dolor que sufre un paciente) y la salud mental (por ejemplo, para identificar el estado de ánimo). También se han ido extendiendo sus usos empresariales y comerciales (por ejemplo, en aplicaciones que recomiendan productos según los estados de ánimo de los clientes). Asimismo, el reconocimiento automatizado de emociones está siendo utilizado en la educación (por ejemplo, para la personalización del aprendizaje o en la identificación de las dificultades de aprendizaje). Igualmente,

27. Picard, R. W., *Affective computing*, MIT press, Boston, 2000.

28. Mcstay, A., «Emotional AI and EdTech: serving the public good?», *Learning, Media and Technology*, núm 45 (2020); Gremsl, T. y Hödl, E., «Emotional AI: Legal and ethical challenges 1», *Information Polity*, núm 27 (2022); Podoletz, L., «We have to talk about emotional AI and crime», *AI & SOCIETY*, núm 38 (2023).

29. Steindl, E., «Does the European Data Protection Framework Adequately Protect Our Emotions? *Emotion Tech in Light of the Draft AI Act and Its Interplay with the GDPR*», *Eur. Data Prot. L. Rev.*, núm 8 (2022) <style face="italic">Eur. Data Prot. L. Rev.</style>, núm 8 (2022).

30. Mcstay, A., «Emotional AI and EdTech: serving the public good?», *Learning, Media and Technology*, núm 45 (2020).

en la seguridad ciudadana se han impulsado proyectos como Avatar —el Agente Virtual Automatizado para la Evaluación de la Verdad en Tiempo Real desarrollado en los EUA para el control de fronteras— que analiza el lenguaje verbal y no verbal de los viajeros que quieren entrar en el país³¹; o iBorderCtrl, el proyecto de control de fronteras financiado por la Comisión Europea³². Por último, también se utilizan estos sistemas de IA en el reconocimiento de emociones en otras aplicaciones como las que incorporan algunos vehículos para detectar si un conductor al volante se está durmiendo,³³ o las que utilizan algunas empresas para hacer el seguimiento y el control de la actividad desempeñada en el puesto de trabajo³⁴. De todos modos, en relación con estas últimas aplicaciones, cabe tener presente que el considerando 18 RIA no considera incluidos entre los sistemas de reconocimiento de emociones a aquellos sistemas utilizados para detectar el cansancio de los pilotos o conductores profesionales con el fin de evitar accidentes.

La extensión de los dispositivos que tienen algún tipo de sistema de IA de reconocimiento de emociones es grande calculándose que se extiende a un 10% de los dispositivos y se ha previsto que pueda alcanzar un valor de 37 billones de dólares en 2026³⁵.

A pesar de los avances que se han ido produciendo, no se puede desconocer que no existe un consenso académico sobre la relación entre las emociones y su expresión física o fisiológica³⁶. En los últimos años distintas voces han manifestado que el reconocimiento de emociones no está científicamente probado³⁷. También han indicado que las expresiones de las emociones, por ejemplo, a través del rostro, no son las mismas según el contexto o la cultura³⁸. De este modo, han destacado que estos sistemas de IA con frecuencia no logran obtener los resultados esperados³⁹.

Tampoco se puede menospreciar el impacto que el uso de estos sistemas puede tener en los derechos fundamentales. De hecho, determinados sistemas de IA de reconocimiento de emociones son considerados en el RIA como sistemas de alto riesgo (apartado 1.c anexo III).

31. Cotino Hueso, L., «Sistemas de inteligencia artificial con reconocimiento facial y datos biométricos. Mejor regular bien que prohibir mal», *El Cronista del Estado Social y Democrático de Derecho*, núm 100 (2022)
32. Romano, A., «Drets fonamentals i intel·ligència artificial emocional en iBorderCtrl: reptes de l'automatització en l'àmbit migratori», *Revista Catalana de Dret Públic*, núm (2023).
33. Mcstay, A. y Urquhart, L., «In cars (are we really safest of all?): interior sensing and emotional opacity», *International Review of Law, Computers & Technology*, núm 36 (2022).
34. Kumar, M., Aijaz, A., Chattar, O., Shukla, J. y Mutharaju, R., «Opacity, Transparency, and the Ethics of Affective Computing», núm (2024).
35. Crawford, K., «Time to regulate AI that interprets human emotions», *Nature*, núm 592 (2021).
36. Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M. y Pollak, S. D., «Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements», *Psychological science in the public interest*, núm 20 (2019).
37. Barkane, I., «Questioning the EU proposal for an Artificial Intelligence Act: The need for prohibitions and a stricter approach to biometric surveillance 1», *Information Polity*, núm 27 (2022).
38. Heaven, D., «Why faces don't always tell the truth about feelings», *Nature*, núm 578 (2020).
39. Katirai, A., «Ethical considerations in emotion recognition technologies: a review of the literature», *AI and Ethics*, núm (2023).

En particular, la utilización de los sistemas de IA de reconocimiento de emociones puede tener un impacto en la privacidad y la protección de datos personales⁴⁰. En efecto, no podemos desconocer la sensibilidad de los datos sobre emociones (*emotional data*) que en determinados casos pueden ser considerados como datos personales en la medida en que permitan llegar a identificar a una persona⁴¹. Incluso los datos sobre emociones también pueden llegar a ser catalogados como datos biométricos que exigen una protección especial⁴². Si bien el RGPD no se refiere explícitamente a ello, es evidente que a la vista de la definición de los datos biométricos que incluye el artículo 4.14 en numerosas ocasiones los datos sobre emociones podrán considerarse como datos biométricos con las consecuencias que de ello se puedan derivar como la prohibición de que sean objeto de tratamiento si no concurre ninguno de los supuestos previstos en el artículo 9.2 RGPD.

De hecho, la definición del sistema de IA de reconocimiento de emociones se vincula a los datos biométricos. El artículo 3, apartado 30, RIA —de igual manera que el artículo 4.14 RGPD— define los datos biométricos como aquellos datos personales resultantes de un tratamiento técnico específico relativo a las características físicas, fisiológicas o de comportamiento de una persona física, como las imágenes faciales o los datos dactiloscópicos y, por lo tanto, los datos sobre emociones pueden llegar a ser datos biométricos. Por ello, en la medida en que los sistemas de IA de reconocimiento de emociones impliquen un tratamiento de datos personales no solo entrarán en el ámbito de aplicación del artículo 50 RIA sino también será de aplicación lo dispuesto en el RGPD o, en su caso, el Reglamento (UE) 2018/1725 del Parlamento Europeo y del Consejo, de 23 de octubre de 2018, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por las instituciones, órganos y organismos de la Unión, y a la libre circulación de esos datos o la Directiva (UE) 2016/680 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por parte de las autoridades competentes para fines de prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de sanciones penales, y a la libre circulación de dichos datos.

Además, en la medida en que los datos sobre emociones tratados por el sistema de IA puedan considerarse datos biométricos estos sistemas de IA también serán considerados como sistemas de IA de alto riesgo según se dispone en el anexo III RIA por remisión del artículo 6, apartado 2, RIA. Precisamente por ello se ha criticado que el RIA califique a los sistemas de IA de reconocimiento de emociones como sistemas de riesgo limitado por considerar que es insuficiente para dar respuesta a los riesgos que eventualmente pueda generar y para informar adecuadamente a los usuarios del impacto que pueden tener o de la intrusión que puedan suponer⁴³.

40. Podoletz, L., «We have to talk about emotional AI and crime», *AI & SOCIETY*, núm 38 (2023).

41. Gremsl, T. y Hödl, E., «Emotional AI: Legal and ethical challenges 1», *Information Polity*, núm 27 (2022).

42. Romano, A., «Drets fonamentals i intel·ligència artificial emocional en iBorderCtrl: reptes de l'automatització en l'àmbit migratori», *Revista Catalana de Dret Públic*, núm (2023).

43. Steindl, E., «Does the European Data Protection Framework Adequately Protect Our Emotions? Emotion Tech in Light of the Draft AI Act and Its Interplay with the GDPR», *Eur. Data Prot. L. Rev.*, núm 8 (2022) <style face="italic">Eur. Data Prot. L. Rev.</style>.

En cualquier caso, como se ha indicado anteriormente, la aplicación de las obligaciones de transparencia previstas en el artículo 50 no afectará a los requisitos y obligaciones exigibles a los sistemas de alto riesgo.

2. ALCANCE DE LAS OBLIGACIONES

En los últimos años ha ido surgiendo un creciente temor hacia la vigilancia emocional (*emotiveillance*)⁴⁴ que ha llevado a que desde algunas instancias se haya propuesto la regulación de su uso⁴⁵, y, hasta entonces, su prohibición⁴⁶.

El RIA más allá de su posible catalogación como sistemas de IA de alto riesgo ha optado en el apartado 3 artículo 50 por establecer unas obligaciones de transparencia de los sistemas de IA de reconocimiento de emociones que deberán cumplir los responsables del despliegue.

A diferencia de los sistemas de IA analizados en los apartados anteriores, en este caso no son los proveedores de los sistemas de IA sino sus usuarios, los responsables del despliegue, quienes deberán informar del funcionamiento del sistema.

En efecto, de acuerdo con el apartado 3 del artículo 50, los responsables del despliegue deberán informar a las personas físicas expuestas al sistema de IA de reconocimiento de emociones. De este modo, se persigue que las personas afectadas puedan ser fácilmente conscientes del uso de estos sistemas y de que sus emociones pueden ser automáticamente reconocidas a través de la IA.

La información debe referirse al funcionamiento del sistema, es decir, de acuerdo con el apartado 18 del artículo 3 RIA, a la capacidad del sistema para alcanzar su finalidad prevista, o sea, el reconocimiento de las emociones. Esta debería permitir no solo conocer la existencia del sistema de IA sino que la persona afectada pueda decidir si quiere ser objeto del reconocimiento de sus emociones de manera automatizada o de los resultados o las consecuencias que le puede acarrear.

La información que se facilite debe versar sobre el uso para el que se ha diseñado el sistema de IA, su contexto y condiciones de uso concretos. Para ello, el responsable del despliegue debe tener en cuenta la información facilitada por el proveedor en las instrucciones de uso, los materiales y las declaraciones de promoción y venta, y la documentación técnica (apartado 12 del artículo 3 RIA).

Para concretar el alcance de esta información podría tenerse en cuenta lo que dispone el artículo 13 respecto a los sistemas de alto riesgo cuando exige que la información que se facilite sea una «información concisa, completa, correcta y clara que sea pertinente, accesible y comprensible» sobre aspectos como la identidad y los datos de contacto del proveedor; las características, capacidades y limitaciones del funcionamiento del sistema de IA (finalidad prevista; nivel de precisión; cualquier

núm 8 (2022); Veale, M. y Zuiderveen Borgesius, F., «*Demystifying the Draft EU Artificial Intelligence Act-Analysing the good, the bad, and the unclear elements of the proposed approach*», *Computer Law Review International*, núm 22 (2021).

44. Steindl, E., «*Does the European Data Protection Framework Adequately Protect Our Emotions?*» *op.cit* <style face="italic">Eur. Data Prot. L. Rev.</style>, núm 8 (2022).

45. Crawford, K., «*Time to regulate AI that interprets human emotions*», *Nature*, núm 592 (2021).

46. AI Now Institute, *2019 Report*, 2019.

circunstancia conocida o previsible que pueda dar lugar a riesgos para la salud y la seguridad o los derechos fundamentales; prestaciones en relación con las personas a las que se vaya a utilizar el sistema; información que permita interpretar los resultados del sistema y utilizarlos adecuadamente); los cambios introducidos en el momento de efectuar la evaluación de la conformidad inicial; las medidas de supervisión humana; los recursos informáticos y de hardware necesarios, la vida útil prevista y las medidas de mantenimiento y cuidado necesarias; una descripción de los mecanismos incluidos en el sistema de IA que permitan a los usuarios recopilar, almacenar e interpretar adecuadamente los registros.

Más allá del contenido, debemos recordar la necesidad de que la información pueda llegar de manera adecuada a las personas afectadas facilitándose de manera clara y distinguible.

Por último, como en otros supuestos previstos en el artículo 50 RIA, en este apartado se prevé como excepción a la obligación de transparencia aquellos casos de sistemas de IA que estén permitidos por la ley para detectar, prevenir e investigar delitos penales siempre y cuando se den las garantías adecuadas para los derechos y libertades de terceros, y de conformidad con el Derecho de la Unión.

VII. SISTEMAS DE INTELIGENCIA ARTIFICIAL DE CATEGORIZACIÓN BIOMÉTRICA

1. SISTEMAS INCLUIDOS

La biometría es una de las aplicaciones de los sistemas de IA más extendidas y que pueden entrañar mayores riesgos para los derechos de las personas.

Generalmente, los datos biométricos son utilizados para establecer o autenticar la identidad de una persona a partir de elementos biológicos (por ejemplo, iris, cara, huellas dactilares o ADN), del comportamiento (los andares o la voz) o, incluso, adquiridos (marcas, tatuajes)⁴⁷. Pero también pueden ser utilizados para perfilar o clasificar a las personas en grupos⁴⁸. Así lo reconoce el RIA al afirmar que «Los datos biométricos pueden permitir la autenticación, la identificación o la categorización de las personas físicas y el reconocimiento de las emociones de las personas físicas.» (considerando 7).

De acuerdo con el RIA, los sistemas de categorización biométrica son «un sistema de IA destinado a incluir a las personas físicas en categorías específicas en función de sus datos biométricos, a menos que sea accesorio a otro servicio comercial y estrictamente necesario por razones técnicas objetivas» (artículo 3, apartado 40).

Los sistemas de IA incluidos en este apartado únicamente deben tener por finalidad la categorización. No en cambio, la identificación de una persona (es decir, el proceso de determinar la identidad de una persona comparando sus datos biométricos con los datos biométricos de personas almacenados en una base de datos)

47. De Keyser, A., Bart, Y., Gu, X., Liu, S. Q., Robinson, S. G. y Kannan, P., «Opportunities and challenges of using biometrics for business: Developing a research agenda», *Journal of Business Research*, núm 136 (2021).

48. Mobilio, G., «Your face is not new to me-Regulating the surveillance power of facial recognition technologies», *Internet Policy Review*, núm 12 (2023).

ni tampoco la verificación de su identidad (o sea, la autenticación de la identidad de las personas físicas mediante la comparación de sus datos biométricos con los datos biométricos facilitados previamente). De todos modos, como ha advertido entre otros el Parlamento Europeo, la distinción entre sistemas de identificación biométrica y los de categorización biométrica puede llegar a ser arbitraria en la medida en que la categorización puede utilizar datos que eventualmente pueden permitir la identificación⁴⁹.

De hecho, esta distinción es importante porque a la vista del RIA los sistemas que utilizan datos biométricos pueden clasificarse en tres categorías distintas con regulaciones bien diversas⁵⁰. En particular, se considera como una práctica de IA prohibida «la introducción en el mercado, la puesta en servicio para este fin específico o el uso de sistemas de categorización biométrica que clasifiquen individualmente a las personas físicas sobre la base de sus datos biométricos para deducir o inferir su raza, opiniones políticas, afiliación sindical, convicciones religiosas o filosóficas, vida sexual u orientación sexual; esta prohibición no abarca el etiquetado o filtrado de conjuntos de datos biométricos adquiridos legalmente, como imágenes, basado en datos biométricos ni la categorización de datos biométricos en el ámbito de la aplicación de la ley» (artículo 5.1.g RIA).

En segundo lugar, se incluye entre los sistemas de IA de alto riesgo, en la medida en que su uso esté permitido por el Derecho de la Unión o nacional aplicable, los «sistemas de IA destinados a ser utilizados para la categorización biométrica en función de atributos o características sensibles o protegidos basada en la inferencia de dichos atributos o características» (anexo III).

Por último, los sistemas que utilizan datos biométricos pueden ser considerados como sistemas de riesgo limitado si únicamente se utilizan para la categorización biométrica. En este caso, la decisión se basa en el menor impacto que pueden tener estos sistemas en los derechos fundamentales⁵¹.

Asimismo, el uso que se dé a los datos biométricos puede determinar el régimen jurídico aplicable. En particular, si los datos relativos a las emociones no permiten la identificación única de la persona no serán datos personales ni tampoco datos especialmente protegidos (artículos 14 apartado 1 y 9 RGPD). En caso contrario, sí lo serán y, como se desprende del apartado 3 del artículo 50 RIA, los responsables del despliegue de un sistema de categorización biométrica deberán tratarlos de conformidad con el RGPD el Reglamento (UE) 2018/1725 y con la Directiva (UE) 2016/680, según corresponda.

La categorización biométrica está siendo utilizada en el ámbito comercial para conocer las preferencias de los consumidores o personalizar las acciones de marketing. También en los departamentos de recursos humanos de las empresas durante el proceso de selección.

49. European Parliament, *Regulating facial recognition in the EU*, 2021.

50. Barkane, I., «*Questioning the EU proposal for an Artificial Intelligence Act: The need for prohibitions and a stricter approach to biometric surveillance 1*», *Information Polity*, núm 27 (2022).

51. Edwards, L., *The EU AI Act: a summary of its significance and scope*, Ada Lovelace Institute, 2022.

2. ALCANCE DE LAS OBLIGACIONES

Finalmente, en el RIA la regulación de la obligación de transparencia de los sistemas de categorización biométrica se ha realizado conjuntamente con la de los sistemas de IA de reconocimiento de emociones.

En los sistemas de IA de categorización biométrica, los sujetos obligados también son los responsables del despliegue (apartado 3 del artículo 50 RIA). Estos deben informar a las personas físicas expuestas a los sistemas de categorización biométrica del funcionamiento del sistema en los mismos términos que se han comentado en relación con los sistemas de reconocimiento de emociones.

Para que el responsable del despliegue pueda cumplir con sus obligaciones debe disponer de la información necesaria, entre otras, que pueda conocer que aquel sistema está llevando a cabo una categorización biométrica. Por ello, la obligación prevista en el artículo 50 debería poderse extender en determinados casos a los proveedores del sistema que serán los que en última instancia conozcan el diseño del sistema y puedan facilitar la información necesaria a los sujetos obligados de informar sobre la existencia de la categorización.

Por último, como en los sistemas expuestos en el apartado anterior, también en el caso de los sistemas de categorización biométrica se contempla como excepción de la obligación de transparencia cuando los sistemas sean para detectar, prevenir e investigar delitos penales, con sujeción a las garantías adecuadas para los derechos y libertades de terceros, y de conformidad con el Derecho de la Unión, en los supuestos permitidos por la ley.

VIII. SISTEMAS DE INTELIGENCIA ARTIFICIAL QUE GENEREN O MANIPULE CONTENIDOS QUE CONSTITUYAN UNA ULTRAFALSIFICACIÓN

1. SISTEMAS INCLUIDOS

Tal y como hemos visto anteriormente, la inteligencia artificial permite generar contenidos (imágenes, vídeos o voz) o manipular contenidos ya existentes. En ocasiones, estos contenidos pueden tener un aspecto muy realista o semejante a contenidos ya existentes que pueden llevar a las personas a pensar que son auténticos. La verosimilitud de los contenidos generados puede ser tan elevada que, como observan Seow et al. es necesario preguntarse si aún es válido el aforismo «ver para creer»⁵².

El RIA ha incluido entre los sistemas que deben cumplir obligaciones de transparencia aquellos sistemas de IA que generen o manipulen contenidos de imagen, audio o vídeo que constituyan una falsificación profunda.

El RIA define la ultrafalsificación como aquel contenido de imagen, audio o vídeo generado o manipulado por un sistema de IA que se asemeja a personas, objetos, lugares u otras entidades o sucesos reales y que puede inducir a una persona a pensar erróneamente que son auténticos o verídicos.

52. Seow, J. W., Lim, M. K., Phan, R. C. y Liu, J. K., «A comprehensive overview of Deepfake: Generation, detection, datasets, and opportunities», *Neurocomputing*, núm 513 (2022).

Las ultrafalsificaciones de imágenes o vídeos pueden consistir en la creación de una cara inexistente; la transferencia de la expresión facial o los movimientos del cuerpo de una persona otra; la manipulación de los atributos faciales (por ejemplo, el color de los ojos o de la piel) alterando la apariencia de una persona o el intercambio de caras manteniendo las expresiones originales⁵³. Estos contenidos tienen una gran apariencia de realidad pero que nunca han existido o sucedido⁵⁴.

La manipulación de imágenes o vídeos se está haciendo habitual a medida que van apareciendo nuevas aplicaciones con resultados más realistas y de mayor calidad y que cada vez hacen más difícil distinguir lo verdadero de lo falso. También porque están fácilmente accesibles para cualquier persona incluso sin conocimientos técnicos y ofreciendo resultados sorprendentes a partir de una única imagen. Además, algunas de estas aplicaciones están disponibles en fuentes abiertas facilitando su acceso⁵⁵.

La extensión de las falsificaciones profundas se está produciendo de la mano de la evolución de la inteligencia artificial pero también de la mayor disponibilidad de bases de datos con las que se entrenan los algoritmos. Todo ello, consume gran cantidad de recursos⁵⁶.

Las ultrafalsificaciones pueden generarse o utilizarse con distintas finalidades. Por ejemplo, se están extendiendo en la industria multimedia (por ejemplo, en la recreación de escenas en las películas; en la incorporación de efectos especiales; o para doblar a los actores a cualquier lengua), de los videojuegos (por ejemplo, creando dobles virtuales de jugadores). También se está utilizando en el ámbito de la educación, de la salud, de la asistencia personal o de la interpretación (por ejemplo, traduciendo un discurso y al mismo tiempo alterando los movimientos de los labios y las expresiones faciales para simular que todo el mundo habla el mismo idioma). Incluso, ya hay algunas aplicaciones para ayudar a gestionar el duelo o para permitir la interacción con personajes famosos fallecidos⁵⁷. En el ámbito empresarial también están teniendo muchas aplicaciones (por ejemplo, para la creación de campañas de marketing, embajadores de marca virtuales o modelos)⁵⁸.

No obstante, en los últimos años, se están multiplicando los usos de las ultrafalsificaciones con el fin de desinformar, defraudar o manipular. Este problema se ve incrementado significativamente por el uso de las redes sociales⁵⁹. Por ello,

53. Seow, J. W., Lim, M. K., Phan, R. C. y Liu, J. K., «A comprehensive overview of Deepfake: Generation, detection, datasets, and opportunities», *op.cit.*

54. Albahar, M. y Almalki, J., «Deepfakes: Threats and countermeasures systematic review», *Journal of Theoretical and Applied Information Technology*, núm 97 (2019).

55. Naitali, A., Ridouani, M., Salahdine, F. y Kaabouch, N., «Deepfake Attacks: Generation, Detection, Datasets, Challenges, and Research Directions», *Computers*, núm 12 (2023).

56. Seow, J. W., Lim, M. K., Phan, R. C. y Liu, J. K., «A comprehensive overview of Deepfake: Generation, detection, datasets, and opportunities», *op.cit.*

57. Caporusso, N., *Deepfakes for the good: A beneficial application of contentious artificial intelligence technology*, Springer, 2021.

58. Mustak, M., Salminen, J., Mäntymäki, M., Rahman, A. y Dwivedi, Y. K., «Deepfakes: Deceptions, mitigations, and opportunities», *Journal of Business Research*, núm 154 (2023).

59. Westerlund, M., «The emergence of deepfake technology: A review», *Technology innovation management review*, núm 9 (2019); Mustak, M., Salminen, J., Mäntymäki, M., Rahman, A. y Dwivedi, Y. K., «Deepfakes: Deceptions, mitigations, and opportunities», *op.cit.*; Masood, M., Nawaz, M., Malik, K. M., Javed, A., Irtaza, A. y Malik, H., «Deepfakes gene-

en la actualidad, se considera que las ultrafalsificaciones son una de las mayores amenazas de la sociedad⁶⁰, y ha llevado a numerosas instancias a impulsar medidas para hacer frente a la desinformación⁶¹.

El uso de ultrafalsificaciones puede tener finalidades muy diversas y se pueden producir tanto en el ámbito público como en el privado. En el ámbito público, el uso engañoso de ultrafalsificaciones puede perseguir influir la opinión pública o en resultados electorales⁶², minar la confianza pública en las instituciones⁶³, ampliar la polarización política o incluso dar apoyo a los discursos de grupos extremistas⁶⁴. Todo ello puede también afectar a la credibilidad de los medios de comunicación que tienen la tarea añadida de confirmar la veracidad de las ultrafalsificaciones.

En el ámbito privado, la generación y difusión de ultrafalsificaciones también se está utilizando, entre otros, para defraudar, acosar, extorsionar, vengarse de personas o para suplantar identidades⁶⁵. Algunas de estas acciones pueden comportar graves perjuicios para el prestigio o la credibilidad de las personas afectadas. También pueden generar confusión entre los consumidores y tener un impacto negativo en el mercado (por ejemplo, por la difusión de ultrafalsificaciones de imágenes o vídeos de directivos de una empresa en una situación comprometida o manipulando unas declaraciones)⁶⁶.

Además de los errores que derivarse las propias ultrafalsificaciones, otro problema vinculado a su generación es la dificultad para detectarlas. Los sistemas de IA permiten generar ultrafalsificaciones hiperrealistas así como manipular contenidos reduciendo o incluso suprimiendo las huellas o trazas que puedan permitir observar la manipulación.

Para dar respuesta a estos problemas se están promoviendo diversas medidas.

En primer lugar, se está avanzando en el desarrollo de técnicas que permitan evaluar la autenticidad de una imagen o un vídeo o para detectar las falsificaciones⁶⁷. En particular, se están desarrollando algoritmos que permiten buscar inconsistencias

ration and detection: state-of-the-art, open challenges, countermeasures, and way forward», Applied Intelligence, núm 53 (2023).

60. Caldwell, M., Andrews, J. T. A., Tanay, T. y Griffin, L. D., «AI-enabled future crime», *Crime Science*, núm 9 (2020).
61. Como muestra pueden traerse a colación, entre otros, los trabajos de la Unión Europea, recogidos, por ejemplo, en la Comunicación La lucha contra la desinformación en línea: un enfoque europeo [COM(2018) 236 final] y, más recientemente, en distintas medidas incluidas en el Reglamento (UE) 2022/2065 del Parlamento Europeo y del Consejo de 19 de octubre de 2022 relativo a un mercado único de servicios digitales y por el que se modifica la Directiva 2000/31/CE (Reglamento de Servicios Digitales).
62. Naitali, A., Ridouani, M., Salahdine, F. y Kaabouch, N., «Deepfake Attacks: Generation, Detection, Datasets, Challenges, and Research Directions», *Computers*, núm 12 (2023).
63. Seow, J. W., Lim, M. K., Phan, R. C. y Liu, J. K., «A comprehensive overview of Deepfake: Generation, detection, datasets, and opportunities», *op.cit.*
64. Europol, *Facing reality? Law enforcement and the challenge of deepfakes*, 2022.
65. Europol, *Facing reality? Law enforcement and the challenge of deepfakes*, 2022.
66. Mustak, M., Salminen, J., Mäntymäki, M., Rahman, A. y Dwivedi, Y. K., «Deepfakes: Deceptions, mitigations, and opportunities», *op.cit.*
67. Rana, M. S., Nobi, M. N., Murali, B. y Sung, A. H., «Deepfake detection: A systematic literature review», *IEEE access*, núm 10 (2022).

entre los fotogramas que conforman una imagen (por ejemplo, inconsistencias entre el discurso y el movimiento de labios, movimiento de los ojos o de los párpado, sobras, incoherencias en la iluminación de las distintas partes de la imagen o reflejos de la luz en los ojos)⁶⁸; o que analizan elementos físicos o fisiológicos de la imagen para valorar su verosimilitud (por ejemplo, analizando el color de la piel que genera la circulación de sangre por la cara). De todos modos, los sistemas de detección aún no son suficientemente fiables, entre otros aspectos, por la calidad y disponibilidad de datos para entrenar a los algoritmos⁶⁹, porque los vídeos a medida que son difundidos, comprimidos o reducidos se ven alterados⁷⁰, o porque aún no funcionan bien en tiempo real⁷¹. También se está promoviendo el uso de tecnologías, como las cadenas de bloques (blockchain), para verificar la legitimidad y el origen de los contenidos de una manera confiable, segura y descentralizada⁷².

En segundo lugar, se está avanzando en la promoción de un uso responsable y ético de estos sistemas de IA para evitar que los contenidos generados contribuyan a la desinformación o al desarrollo de actividades delictivas o perjudiciales⁷³. En esta dirección, una mayor sensibilización y formación pueden contribuir a evitar o minimizar los efectos nocivos que se puedan derivar del uso de los sistemas e IA que generen ultrafalsificaciones y de su difusión en las redes sociales con el fin de desinformar.

En tercer lugar, se está progresando en la regulación del uso de las ultrafalsificaciones. Para ello, se han propuesto distintas opciones. Una opción hubiese sido limitar o prohibir la circulación de ultrafalsificaciones. No obstante esta solución si bien puede prevenir o evitar determinados perjuicios también puede generar nuevos impactos⁷⁴. En efecto, no podemos desconocer que el uso de estos sistemas de IA puede ser una manifestación de la libertad de expresión o de la libertad de creación artística. Frente a estas opciones, y al margen de lo que se pueda derivar de la consideración de determinados sistemas como de alto impacto, el RIA ha optado por prever el cumplimiento de las obligaciones de transparencia que se analizan en el próximo epígrafe.

2. ALCANCE DE LAS OBLIGACIONES

El apartado 4 del artículo 50 RIA prevé la obligación de hacer público que los contenidos o imágenes han sido generados o manipulados de manera artificial.

68. Westerlund, M., «*The emergence of deepfake technology: A review*», op.cit.; Naitali, A., Ridouani, M., Salahdine, F. y Kaabouch, N., «*Deepfake Attacks: Generation, Detection, Datasets, Challenges, and Research Directions*», op.cit.

69. Naitali, A., Ridouani, M., Salahdine, F. y Kaabouch, N., «*Deepfake Attacks: Generation, Detection, Datasets, Challenges, and Research Directions*», op.cit.

70. Europol, *Facing reality? Law enforcement and the challenge of deepfakes*, 2022.

71. Naitali, A., Ridouani, M., Salahdine, F. y Kaabouch, N., «*Deepfake Attacks: Generation, Detection, Datasets, Challenges, and Research Directions*», op.cit.

72. Rana, M. S., Nobi, M. N., Murali, B. y Sung, A. H., «*Deepfake detection: A systematic literature review*», op.cit .

73. Naitali, A., Ridouani, M., Salahdine, F. y Kaabouch, N., «*Deepfake Attacks: Generation, Detection, Datasets, Challenges, and Research Directions*», op.cit.

74. Mustak, M., Salminen, J., Mäntymäki, M., Rahman, A. y Dwivedi, Y. K., «*Deepfakes: Deceptions, mitigations, and opportunities*», op.cit.

El sujeto obligado son los responsables del despliegue del sistema de IA que genere o manipule imágenes o contenidos de audio o vídeo que constituya una ultrafalsificación. Como ya hemos indicado anteriormente, en determinadas circunstancias los responsables del despliegue pueden desconocer hasta qué punto un determinado contenido ha sido creado por un sistema de IA. Ante este desconocimiento puede ser difícil que efectivamente puedan facilitar la información a los receptores finales de la imagen o el vídeo.

La obligación de transparencia consiste en informar de que el contenido ha sido generado o manipulado de manera artificial.

El RIA trata de encontrar un equilibrio entre la libertad de expresión y la prevención de la desinformación y la manipulación a través de la generación o difusión de ultrafalsificaciones. A tal efecto, se prevé que «Cuando el contenido forme parte de una obra o programa manifiestamente creativos, satíricos, artísticos o de ficción, las obligaciones de transparencia establecidas en el presente apartado se limitarán a la obligación de hacer pública la existencia de dicho contenido generado o manipulado artificialmente de una manera adecuada que no dificulte la exhibición o el disfrute de la obra».

Asimismo, el RIA persigue garantizar la libertad de prensa. Para ello, se dispone que «Los responsables del despliegue de un sistema de IA que genere o manipule texto que se publique con el fin de informar al público sobre asuntos de interés público divulgarán que el texto se ha generado o manipulado de manera artificial».

Como en otros supuestos, se prevé como excepción de esta obligación aquellos supuestos autorizados por la ley para detectar, prevenir, investigar o enjuiciar infracciones penales. Tampoco, cuando el contenido generado haya sido revisado por una persona o sometido control editorial y cuando una persona física o jurídica tenga la responsabilidad editorial por la publicación del contenido.

IX. RECAPITULACIÓN

El artículo 50 RIA regula unas obligaciones de transparencia destinadas a evitar que determinados sistemas de IA, que en sí mismos no han de comportar un riesgo para los intereses de la Unión Europea o los derechos fundamentales de las personas, puedan ser utilizados de tal manera que no generen confusión, engaños, manipulaciones o desinformación entre las personas que interactúen con dichos sistemas o destinatarias de los contenidos generados o manipulados.

Para ello, prevé que los proveedores o los responsables del despliegue según el caso, deben facilitar información a las personas usuarias o a las personas afectadas de manera que puedan ser conscientes o evitar una situación determinada o los resultados dañinos que eventualmente se puedan causar por el uso del sistema de IA.

En la práctica, las obligaciones de transparencia previstas en el artículo 50 RIA, orientadas a garantizar la comunicación a la persona usuaria o afectada de la existencia de un sistema de IA o de un resultado generado de manera artificial, serán en muchos casos complementarias a otras obligaciones de transparencia previstas en el RIA, por ejemplo, para los sistemas de IA de alto riesgo, cuya finalidad es garantizar la trazabilidad del funcionamiento del sistema de IA y la explicabilidad de los resultados obtenidos. Para lograr la finalidad prevista es fundamental que la

información que se facilite sea lo suficientemente clara para que pueda lograrse la finalidad prevista.

La evaluación que periódicamente realice la Comisión de acuerdo con lo previsto en el artículo 112.2 RIA permitirá conocer si las obligaciones previstas en el artículo 50 lo han logrado y determinará si es necesario modificar la relación de sistemas de IA que deban cumplir con las obligaciones de transparencia.

SANDBOX, GOBERNANZA, VIGILANCIA,
RÉGIMEN SANCIONADOR, DERECHOS Y
CONFIDENCIALIDAD EN EL REGLAMENTO

Sandbox, espacios controlados y pruebas en condiciones reales de sistemas de inteligencia artificial en el Reglamento. Medidas para PYMES, startups y micro empresas

LORENZO COTINO HUESO

Catedrático de Derecho Constitucional de la Universitat de València. Valgrai¹

I. LAS «MEDIDAS DE APOYO A LA INNOVACIÓN» DEL CAPÍTULO VI

El Capítulo VI, bajo el título «Medidas de apoyo a la innovación», incluye siete artículos de considerable extensión, sumando más de 5,500 palabras, que abordan diversos temas. Esencialmente, desde la propuesta inicial de la Comisión en 2021, este capítulo regula los «Espacios controlados de pruebas para la IA», que se abreviarán aquí como «sandbox». Además de regular su existencia y régimen (arts. 57-58), se legitima el posible tratamiento de datos personales por sistemas IA en estos sandbox (art. 59). Aunque no estaba en la propuesta inicial de la Comisión, el RIA incluye la regulación de las «Pruebas de sistemas de IA de alto riesgo en condiciones reales fuera de los espacios controlados de pruebas para la IA» (art. 60), junto con un artículo sobre el «consentimiento informado» de las personas afectadas por estas pruebas (art. 61). Desde el inicio, el capítulo VI también contenía un artículo sobre «Medidas dirigidas a proveedores y responsables del despliegue, en particular PYMES, incluidas las empresas emergentes» (art. 62), al que se ha añadido uno relativo a «Excepciones para operadores específicos» (art. 63).

1. cotino@uv.es. OdiseIA. El presente estudio es resultado de investigación de los siguientes proyectos: MICINN Proyecto «Derechos y garantías públicas frente a las decisiones automatizadas y el sesgo y discriminación algorítmicas» 2023-2025 (PID2022-136439OB-I00) financiado por MCIN/AEI/10.13039/501100011033/; «La regulación de la transformación digital ...» Generalitat Valenciana «Algorithmic law» (Prometeo/2021/009, 2021-24); «Algorithmic Decisions and the Law: Opening the Black Box» (TED2021-131472A-I00) y «Transición digital de las Administraciones públicas e inteligencia artificial» (TED2021-132191B-I00) del Plan de Recuperación, Transformación y Resiliencia. Estancia Generalitat Valenciana CIAEST/2022/1., Grupo de Investigación en Derecho Público y TIC Universidad Católica de Colombia; MICINN; Estancia Generalitat Valenciana CIAEST/2022/1, Convenio de Derechos Digitales-SEDIA Ámbito 5 (2023/C046/00228673) y Ámbito 6. (2023/C046/00229475).

Entre los cambios más notables desde la versión inicial, además de la incorporación de las pruebas en condiciones reales, se destaca la obligatoriedad de establecer un sandbox en cada Estado en los dos años posteriores a la entrada en vigor del RIA. Igualmente, se contempla la posibilidad de creación a nivel regional o local, conjuntamente con otros Estados, así como por la Comisión y el Supervisor Europeo de Protección de Datos. Asimismo, el Consejo introdujo las finalidades u objetivos de los sandbox (art. 57. 9º). El régimen de flexibilidad en el cumplimiento de la normativa por los participantes en un sandbox y su posible responsabilidad también ha ido variando. Se han añadido elementos como el plan específico a presentar, las consecuencias de la participación en un sandbox, la obligación de facilitar prueba escrita de las actividades realizadas por el participante, y el informe de salida. Es relevante que se haya vinculado la participación en un sandbox como medio para demostrar el cumplimiento del proceso de evaluación de la conformidad.

Por otro lado, en este capítulo VI, el Parlamento (Enmienda 516) propuso incluir un artículo relativo a la «Promoción de la investigación y el desarrollo de la IA en apoyo de resultados social y ambientalmente beneficiosos», con mandatos de promoción. Sin embargo, esta propuesta no tuvo éxito.²

II. ORIGEN Y CONCEPTO DE SANDBOX Y ESPACIOS CONTROLADOS

Montesquieu afirmaba que «a veces incluso es conveniente probar una ley antes de establecerla. Las constituciones de Roma y de Atenas eran muy sabias a este respecto: las decisiones del Senado tenían fuerza de ley durante un año, y sólo se hacían perpetuas por la voluntad del pueblo»³. Estados Unidos permitió los «estados-como-laboratorios». Como afirmó el juez Brandeis, «Uno de los incidentes felices de nuestro sistema federal es que un solo y valiente Estado puede, si sus ciudadanos así lo eligen, servir como un laboratorio y probar nuevos experimentos sociales y económicos sin riesgo para el resto del país» (voto particular del juez Brandeis formulado en el caso *New State Ice v. Liebmann* 285 U.S. 262, 310 [1932]).⁴

Pese a diversas experiencias históricas en la prueba de regulaciones y la innovación desde hace siglos, fue en la segunda mitad del siglo XX cuando se desarrollaron,

2. El mismo incluía un mandato de fomentar soluciones de IA que mejoren la accesibilidad para personas con discapacidad, reduzcan las desigualdades socioeconómicas y apoyen los objetivos de sostenibilidad y protección del medio ambiente. Para ello, se señalaban algunas medidas como proporcionar acceso prioritario; asignar financiación pública a proyectos de IA con impacto social y medioambiental positivo; organizar actividades de sensibilización sobre el RIA; financiación específica y procedimientos de solicitud, adaptadas a las necesidades y canales específicos de comunicación accesibles. Asimismo se estimulaba la participación de la sociedad civil respecto de IA para la sociedad y el medio ambiente.
3. Doménech Pascual, G., «Las regulaciones experimentales», *Anuario del buen gobierno y de la calidad de la regulación*, (monográfico sobre sandbox Ponce Solé, J. y Villoria Mendieta, M., coords.) n.º 1, 2022, pp. 103-146. Cita a Montesquieu, C.-L. de S. *Del espíritu de las leyes*, traducción de Blázquez y de Vega, Tecnos, Madrid 2000.
4. Voto particular del juez Brandeis formulado en el caso *New State Ice v. Liebmann* 285 U.S. 262,310 (1932). *Ibidem*.

acompañando el intervencionismo del Estado social en la vida social y económica.⁵ No obstante, el primer «sandbox» regulatorio formal y la propagación del concepto a nivel mundial se dio para probar la introducción en el mercado de productos *Fintech*.⁶ Así, desde 2014, por la FCA (*Financial Conduct Authority* de Reino Unido), extendiéndose luego a otros sectores regulados, como el sanitario (supervisado por la Care Quality Commission), el energético (OfGem), y de ahí a otros sectores.⁷ El éxito parece demostrado ya que, en julio de 2023, la OCDE afirmó que ya se han dado un centenar de iniciativas de «sandbox», incluidos los de *fintech* y privacidad.⁸

La terminología para referir realidades semejantes a los sandbox o espacios controlados de pruebas es muy variada: «Living laboratories», «innovation spaces», «regulatory test beds» o «real life experiments». Hace años, en su definición destacó el Consejo de la UE, que «8. percibe los “cajones de arena” regulatorios como marcos concretos que, al proporcionar un contexto estructurado para la experimentación, permiten, en su caso, en un entorno real, la prueba de tecnologías, productos, servicios o enfoques innovadores —en la actualidad, especialmente en el contexto de la digitalización— durante un tiempo limitado y en una parte limitada de un sector o ámbito bajo supervisión regulatoria, garantizando la existencia de las salvaguardias adecuadas». Asimismo, «9. entiende las cláusulas de experimentación [*experimental clauses*] como disposiciones legales que permiten a las autoridades encargadas de aplicar y hacer cumplir la legislación ejercer, caso por caso, cierto grado de flexibilidad en relación con la prueba de tecnologías, productos, servicios o enfoques innovadores».⁹

En Alemania, «los “cajones de arena” normativos son zonas de prueba establecidas por un tiempo limitado, que cubren un área limitada, en las que se pueden probar tecnologías y modelos de negocio innovadores en la vida real».¹⁰ Y las cláusulas de experimentación como un instrumento técnico regulatorio que permite hacer

5. BMWi, *Making Space for Innovation: The Handbook for Regulatory Sandboxes*, German Federal Ministry for Economic Affairs and Energy, 2019, p. 7 https://www.bmwk.de/Redaktion/EN/Publikationen/Digitale-Welt/handbook-regulatory-sandboxes.pdf?__blob=publicationFile&v=1
6. Sobre la regulación española en este ámbito, Huergo Lora, A. «Un “espacio controlado de pruebas» (regulatory sandbox) para las empresas financieras tecnológicamente innovadoras». El «Anteproyecto de Ley de Medidas para la transformación digital del sistema financiero», en *El Cronista del Estado Social y Democrático de Derecho*, n.º 76 (septiembre), 2018, pp. 48-59 y Hernández Peña, J. C., «La propuesta de un sandbox regulatorio para el sector financiero español: ¿más luces que sombras?», *Revista General de Derecho de los Sectores Regulados* n.º 2, 2018.
7. Al respecto, Truby, J., «Decarbonizing Bitcoin: Law and policy choices for reducing the energy consumption of Blockchain technologies and digital currencies», *Energy Research & Social Science*, Volume 44, 2018, pp. 399-410, <https://doi.org/10.1016/j.erss.2018.06.009>
8. OECD, *Regulatory sandboxes in artificial intelligence*, OECD Digital Economy Papers, July 2023 No. 356, 2023, p. 8 <https://read.oecd.org/10.1787/8f80a0e6-en?format=pdf>
9. Consejo de la Unión Europea, *Council Conclusions on Regulatory sandboxes and experimentation clauses as tools for an innovation-friendly, future-proof and resilient regulatory framework that masters disruptive challenges in the digital age*, Bruselas, 16 de noviembre de 2020 (OR. en), 13026/20 <https://data.consilium.europa.eu/doc/document/ST-13026-2020-INIT/en/pdf>
10. BMWi, *Making Space for Innovation...* cit. p. 7.

excepciones al marco jurídico general. Permiten, por tanto, adoptar nuevos enfoques, sin que sea posible predecir el resultado. Y ofrecen la oportunidad de aprender sobre las leyes y sus efectos.¹¹

Por su parte, la OCDE señala que «Los espacios aislados de regulación de la IA deben considerarse una de las diversas herramientas para la experimentación y la innovación en materia de regulación, junto con otras áreas complementarias: la normalización, los centros de innovación, otros espacios aislados como los de las tecnologías financieras y la privacidad, y las tecnologías de gobernanza».¹²

Pues bien, el apartado 55 del artículo 3. 1º del RIA define el «espacio controlado de pruebas para la IA» como «un marco controlado establecido por una autoridad competente que ofrece a los proveedores y proveedores potenciales de sistemas de IA la posibilidad de desarrollar, entrenar, validar y probar, en condiciones reales cuando proceda, un sistema de IA innovador, con arreglo a un plan del espacio controlado de pruebas y durante un tiempo limitado, bajo supervisión regulatoria».

III. EXPERIENCIAS DE SANDBOX DE INTELIGENCIA ARTIFICIAL

La experiencia de los sandbox vinculados a la IA está claramente asociada al cumplimiento de la normativa de protección de datos, especialmente en Europa (Reino Unido, Noruega y Francia), así como en Colombia. En otros contextos, los sandbox han estado particularmente ligados a experimentos regulatorios en el ámbito *Fintech*.

El ICO del Reino Unido inició en 2019 un «sandbox» «para apoyar a organizaciones que desarrollan productos o servicios particularmente innovadores que procesan datos personales»¹³. En el contexto de la IA, se centraba en «Innovaciones excepcionales», «tecnologías emergentes» (como tecnología de salud para el consumidor, dispositivos portátiles y aplicaciones de software que ayudan a las personas a evaluar su salud y bienestar; Internet de las cosas (IoT), tecnología inmersiva; finanzas descentralizadas: software que emplea tecnología *blockchain* para respaldar transacciones financieras entre pares) y sobre «Biometría». Se pueden seguir los informes de salida de todos los participantes desde entonces. En noviembre de 2021 publicaron un Informe beta sobre los aprendizajes del Sandbox.¹⁴ Como se expone posteriormente, es notable que no se reguló ninguna excepcionalidad o particularidad específica para los participantes.

En Noruega, la *Datatilsynet*, Autoridad Noruega de Protección de Datos, creó en 2021 un «Sandbox for Artificial Intelligence»¹⁵ inspirado en el del Reino Unido.

11. *Ibidem*, p. 81.

12. OECD, *Regulatory sandboxes... cit.* pp. 24 y ss.

13. <https://ico.org.uk/for-organisations/advice-and-services/regulatory-sandbox/>
Especialmente cabe seguir la Guía, ICO, Information Commissioner's Office, *The Guide to the Sandbox*, <https://ico.org.uk/for-organisations/regulatory-sandbox/the-guide-to-the-sandbox>

14. <https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2019/03/ico-opens-sandbox-beta-phase-to-enhance-data-protection-and-support-innovation/>

15. <https://www.datatilsynet.no/en/regulations-and-tools/sandbox-for-artificial-intelligence/>

Tampoco este sandbox regula un régimen especial para los participantes, como se aprecia después. Los participantes deben seguir las Directrices éticas de la IA responsable del alto grupo de expertos de IA de la UE.¹⁶ Recibió veinticinco solicitudes de múltiples organizaciones públicas y privadas y seleccionó cuatro proyectos para el sandbox, que comenzó en marzo de 2021. En 2023 se difundieron los informes de resultados.¹⁷

Francia cuenta con una amplia experiencia y regulación general e incluso constitucional de sandboxes.¹⁸ En IA, la autoridad de protección de datos (CNIL) ha tenido un claro liderazgo. En 2021, el sandbox estuvo dedicado a las aplicaciones sanitarias, con 10 proyectos.¹⁹ No se eximía del cumplimiento del RGPD, pero el sandbox tenía por objetivo facilitar dicho cumplimiento. La edición 2022 se dedicó a la tecnología educativa con cuatro proyectos.²⁰ En julio de 2023, el sandbox se centró en tres proyectos sobre inteligencia artificial en los servicios públicos.²¹ En 2023, la CNIL lanzó un plan de acción sobre IA.²²

Otros países en la UE no han centrado su actividad en la protección de datos. La estrategia de IA de Alemania incluyó laboratorios vivientes y bancos de pruebas de IA, creando nuevas cláusulas de experimentación como base jurídica. En el

16. HLEG-Comisión Europea, *Directrices éticas para una IA fiable*, 2019, *Directrices éticas para una IA fiable*, 2019, <https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1>

17. <https://www.datatilsynet.no/en/regulations-and-tools/sandbox-for-artificial-intelligence/reports/>

Sobre la transparencia de protección de datos en los sistemas IA. Se indica la importancia de la fijación de las finalidades de los tratamientos de datos de los sistemas (Informe Ruter). El Informe *Ahus* especialmente analiza la discriminación algorítmica de un algoritmo para predecir la insuficiencia cardiaca. El Informe *Simplifai* se centra en los usos de la Administración de IA para registrar y archivar correos electrónicos o la toma de decisiones (NVE). El Informe *Finterai* se centra en el acceso federado y limitado a los datos para el aprendizaje en la lucha contra el blanqueo de dinero y la financiación del terrorismo. El Informe *AVT* sobre evaluaciones individuales y educación adaptada con privacidad. También cabe tener en cuenta el Informe *NAV*, herramienta IA para predecir el desarrollo de las bajas por enfermedad a nivel individual.

18. Por todos, Conseil d'État, *Les expérimentations: comment innover dans la conduite des politiques publiques*, Conseil d'État, París, 2019, https://www.conseil-etat.fr/Media/actualites/documents/2019/10-octobre/etude-pm_experimentations_vdef

19. Resultados en <https://www.cnil.fr/en/digital-health-and-edtech-cnil-publishes-results-its-first-sandboxes#:~:text=its%20first%20%E2%80%9Csandboxes%E2%80%9D-,Digital%20health%20and%20EdTech%3A%20the%20CNIL%20publishes,results%20of%20its%20first%20%E2%80%9Csandboxes%E2%80%9D&text=The%20CNIL%20publishes%20the%20recommendations,health%20and%20educational%20digital%20tools>

20. <https://www.cnil.fr/en/edtech-sandbox-cnil-supports-10-innovative-projects> Proyectos Daylindo, Classroom, Francia Université Numérique y «nube personal» Academia de Rennes. Informe en https://www.cnil.fr/sites/cnil/files/2023-07/bilan_bac_a_sable_edtech.pdf

21. <https://www.cnil.fr/en/sandbox-cnil-launches-call-projects-artificial-intelligence-public-services>

22. <https://www.cnil.fr/en/artificial-intelligence-action-plan-cnil>

ámbito de la conducción automatizada se han desarrollado algunas pruebas, como el proyecto de sandbox en Renania del Norte-Westfalia²³. Un país digital como Estonia lanzó en 2022 un banco de pruebas (*AI Govstack Testbed*)²⁴ centrado en el desarrollo, gestión, análisis y etiquetado de datos. La Autoridad de Innovación Digital de Malta (*Malta Digital Innovation Authority*) creó en 2020 un MDIA-TAS (*Technology Assurance Sandbox*), un sandbox normativo centrado en tecnologías emergentes como la IA.²⁵ Este sandbox pretende ayudar a las empresas a cumplir las normas vigentes.

En Suiza, el Cantón de Zurich desarrolló el *Innovation Sandbox for Artificial Intelligence*²⁶ para colaborar en cuestiones regulatorias y permitir el uso de nuevas fuentes de datos. No parece tener cobertura normativa específica. A diferencia de otros sandboxes, «los proyectos seleccionados no sólo se están revisando, sino también poniéndose en práctica». Entre marzo y junio de 2022 se presentaron 21 proyectos de IA, de los cuales se seleccionaron cinco que actualmente están en fase de ejecución. Durará hasta abril de 2024 y habrá otra convocatoria entre marzo y mayo de 2024. Se han elegido cinco proyectos:²⁷ sistemas autónomos, como tractores o cortacéspedes autodirigidos en espacios públicos; mantenimiento de infraestructuras con drones; aplicaciones de IA en la educación; aparcamiento inteligente en ciudades; y buenas prácticas para la privacidad desde el diseño y traducciones automáticas para la administración pública.

En enero de 2024 se celebró el *AI Sandbox Summit en Zurich*,²⁸ la cumbre de Sandbox de IA, donde se reunieron iniciativas de Alemania, Bélgica, Noruega, Reino Unido, Francia y España. En la misma se subrayó la importancia de los sandboxes regulatorios y no se consideró crucial la dispersión terminológica o el consenso en las definiciones, sino la adopción de diferentes tipos según las necesidades de cada país. Se apostó por la colaboración internacional y se prevé la creación de una base de datos común de casos de uso relevantes en los diferentes sandboxes europeos para facilitar el intercambio de conocimientos.

En 2021, bajo la Ley Federal n.º 258-FZ, Rusia introdujo los «sandboxes» regulatorios para fomentar la innovación digital. Entre los ocho proyectos seleccionados había aplicaciones de IA en el ámbito del transporte, la sanidad y el turismo.²⁹

En Iberoamérica, en 2021, el gobierno de Chile publicó un documento sobre sandbox de IA.³⁰ El Gobierno de Colombia generó una guía de sandboxes regulatorios

23. Estado federado de Renania del Norte-Westfalia (NRW), donde se está llevando a cabo un amplio proyecto Digi-Sandbox.NRW.está en desarrollo. Su sitio web enumera varios reallabs en NRW, pero ninguno se centra en la protección de la privacidad y la inteligencia artificial.

24. https://e-estonia.com/ai-govstack-testbed_eng/

25. <https://www.mdia.gov.mt/technology-assurance-sandbox/>

26. <https://www.zh.ch/en/wirtschaft-arbeit/wirtschaftsstandort/innovation-sandbox.html> <https://innovation.zuerich/en/sandbox/>

27. Puede accederse a los dosieres en la referida web.

28. <https://www.zh.ch/en/wirtschaft-arbeit/wirtschaftsstandort/innovation-sandbox.html>

29. <https://a-ai.ru/en>

30. Guño, A., *Sandbox Regulatorio de Inteligencia Artificial en Chile. Documento para discusión*. CAF, banco de desarrollo de América Latina, agosto, 2021, <https://www.econo->

en IA en 2020³¹, y la Autoridad Colombiana de Protección de Datos Personales lanzó un sandbox regulatorio en privacidad desde el diseño y por defecto en proyectos de inteligencia artificial.³² En 2021 se seleccionaron dos proyectos (*NaaS Colombia S.A.S* —«Evolución Index Core»— y Alcaldía de Barranquilla —«*Chatbot*»—) y en 2022 el «*Diyosoy*» del *Wolman Group de Colombia Limitada*. Como recuerda Granero, en Argentina, para la Ciudad de Buenos Aires, la Ley 6491 sobre Espacio controlado de pruebas del 9 de diciembre de 2021 reguló el marco para realizar este tipo de pruebas.³³ En la provincia de Mendoza, la Ley n.º 9086 de 30 de julio de 2018,³⁴ en sus artículos 52 y siguientes, sobre el transporte urbano a través de aplicaciones móviles y plataformas, fue considerada por la Suprema Corte de Justicia de Mendoza como una regulación experimental.³⁵

En Asia, la *Monetary Authority of Singapore* lanzó en 2018 su sandbox regulatorio de Fintech, facilitando pruebas de aplicaciones IA.³⁶ Los principios publicados por esta autoridad se incorporaron a su Estrategia Nacional de IA. El 1 de abril de 2019 en Corea entró en vigor la Ley Especial de Asistencia a la Innovación Financiera (*SAAFI*) para apoyar el desarrollo de servicios financieros y aumentar los beneficios para los consumidores. Además de un sandbox financiero,³⁷ diversos ministerios de Corea crearon en 2019 un marco público de experimentación industrial con una exención normativa limitada a empresas para probar productos, servicios y modelos de negocio innovadores. Contó con siete proyectos AI+X.

A nivel global, desde 2020 comenzó el sandbox financiero de la *Global Financial Innovation Network* (GFIN) con más de 50 entidades financieras de todo el mundo y algunas lecciones aprendidas.³⁸ Sin embargo, ha sido difícil unirse al mismo por la difícil compatibilidad entre regímenes jurídicos. Así, de 38 solicitudes de empresas, sólo dos lo llevaron a cabo con éxito (*Banksysteme* y *Bedrock AI*).

IV. LAS VENTAJAS QUE IMPLICAN LOS SANDBOX DE INTELIGENCIA ARTIFICIAL DESDE LOS DIFERENTES PUNTOS DE VISTA

El artículo 57. 5º apunta el objetivo general que tiene un sandbox de IA: «proporcionarán un entorno controlado que fomente la innovación y facilite el desarrollo, el entrenamiento, la prueba y la validación de sistemas innovadores de IA durante un período limitado antes de su introducción en el mercado o su puesta en

mia.gob.cl/wp-content/uploads/2021/09/PaperSandboxIA.pdf

31. Superintendencia de Industria y Comercio, *Sandbox sobre privacidad desde el diseño y por defecto en proyectos de inteligencia artificial*, SIC, Colombia, 2021, <https://www.sic.gov.co/content/sandbox-sobre-privacidad-desde-el-disen%CC%83o-y-por-defecto-en-proyectos-de-inteligencia-artificial>

32. <https://www.sic.gov.co/sandbox-microsite> (no disponible).

33. https://documentosboletinoficial.buenosaires.gov.ar/publico/ck_PL-LEY-LCABA-LCBA-6491-22-6295.pdf

34. <https://www.mendoza.gov.ar/gobierno/wp-content/uploads/sites/19/2018/10/Ley-de-Movilidad-N%C2%BA-9086.pdf>

35. Granero Horacio R., «La imperiosa necesidad de regular —bien— la inteligencia artificial», ponencia, FACA, *EIDial*, 2023.

36. <https://www.mas.gov.sg/development/fintech/regulatory-sandbox>

37. <https://sandbox.fintech.or.kr/?lang=en>

38. Puede accederse a los informes en <https://www.thegfin.com/crossborder-testing>

servicio». Asimismo, el artículo 57. 9º RIA señala los posibles objetivos de un sandbox: fomentar la innovación y la competitividad, facilitar el desarrollo de un ecosistema de IA, «facilitar y acelerar el acceso al mercado», en particular de PYMEs y startups, «mejorar la seguridad jurídica y contribuir a la puesta en común de las mejores prácticas mediante la cooperación con las autoridades». Además, se menciona «el intercambio de mejores prácticas mediante la cooperación con las autoridades» y «mejorar la seguridad jurídica para lograr el cumplimiento del presente Reglamento». En la versión del Consejo también se afirmaba «contribuir a la aplicación uniforme y eficaz del RIA» y «contribuir a la elaboración o a la actualización de las normas armonizadas». En sus versiones anteriores a la adoptada, el artículo 53 1 a) RIA señalaba que «El recinto de seguridad reglamentario de la IA permitirá y facilitará la participación de los organismos notificados, los organismos de normalización y otras partes interesadas pertinentes cuando proceda».

Desde la OCDE,³⁹ para el caso de la IA, se entiende que los sandboxes son especialmente adecuados para probar la preparación de los productos o servicios relacionados con la IA para su comercialización a la luz de los estándares y normas. Se subraya la interdependencia entre los estándares y la regulación normativa de la IA, especialmente en los enfoques basados en el riesgo que siguen los estándares, precisamente para aplicar la regulación. Así, los datos que se recogen en un sandbox pueden usarse para detectar patrones e identificar la necesidad de un estándar en un área específica. Y, a la inversa, los procesos de estandarización pueden servir para las pruebas a realizar en los sandboxes de IA.

La cooperación multidisciplinar y entre múltiples partes interesadas es fundamental para un sandbox de IA debido a la intensa transversalidad de esta tecnología, que ha concurrido en particular con las tecnologías financieras y con el ámbito de protección de datos. Esta cooperación exige evitar «compartimentos estancos»⁴⁰ y fomentar la participación entre las distintas autoridades regulatorias interesadas: «autoridades de competencia, oficinas de propiedad intelectual, organismos nacionales de normalización y autoridades de protección de datos». A este respecto, se mencionan algunos ejemplos de cooperación en Corea, Alemania y Brasil.⁴¹ La coordinación con los agentes del mercado también es importante, incluyendo a las empresas.

Las ventajas de los sandbox —y en particular de IA— son variadas desde diferentes puntos de vista. Desde los intereses y finalidades públicas, hay ventajas en términos de innovación, para las autoridades y la mejora e implantación de la regulación. Desde el punto de vista de la economía y el mercado, también hay diversas ventajas. Si bien hay numerosos intereses públicos o colectivos, también

39. OECD, *Regulatory sandboxes...* cit. p. 24.

40. *Ibidem*, pp. 12-13 y en especial p. 19. La coordinación de las respuestas regulatorias entre los organismos nacionales es fundamental, Brummer, C. and Y. Yadav (2019), «Fintech and the Innovation Trilemma», *The Georgetown Law Journal* 107, <https://scholarship.law.vanderbilt.edu/faculty-publications/1084/>

41. *Ibidem*, pp. 19 y ss. El sandbox de Corea cuenta con participación interministerial, y diversos sectores en Alemania utilizan marcos de sandbox flexibles y genéricos. En Brasil, la Comisión de Valores y Bolsa y el Banco Central crearon un comité interno que interactúa con universidades, investigadores, asociaciones y representantes del sector para evaluar las solicitudes de sandbox.

son variados los intereses de las entidades participantes en un sandbox de IA, especialmente si se trata de pequeñas y medianas empresas y startups.

Desde el punto de vista de la innovación, un sandbox de IA implica la promoción de la innovación mediante el suministro de datos,⁴² la transferencia de *know-how* y habilitación de nuevos proyectos, y la gobernanza compartida en colaboración del sector privado con la ciencia, la universidad y el sector regulatorio. El modelo sandbox fomenta asociaciones de investigación y política, la búsqueda de consensos y soluciones normativas. En un sandbox se dan enfoques multidisciplinares bajo contextos controlados y bajo la autoridad de una burocracia profesionalizada, reduciendo las asimetrías de información. La OCDE ha subrayado que los espacios aislados no deben utilizarse únicamente para validar las expectativas de la legislación vinculante, sino también para apoyar la innovación y los centros de innovación. En esta línea, el RIA apunta al apoyo técnico y científico desde los centros de innovación en el ecosistema de la IA (Consid. 139).

Para las autoridades y la mejora regulatoria, la práctica del sandbox aumenta la velocidad de las autorizaciones a las empresas,⁴³ mejora la comunicación entre reguladores y empresas,⁴⁴ permite evaluar la eficacia de la regulación y las políticas en condiciones reales, y genera datos empíricos utilizables para una mejor toma de decisiones regulatorias u otras finalidades de interés público. Un sandbox puede posibilitar una regulación más adaptable y dinámica.⁴⁵ Como se apunta en Suiza, permiten «proporcionar claridad regulatoria». Así, un sandbox permite aprender cómo aplicar la normativa por las autoridades, cómo regular nuevos tipos de servicios en colaboración con actores privados, y reunir más información. Permite sugerir o recomendar ajustes regulatorios, fijar criterios que faciliten el cumplimiento de la regulación y señalar los procedimientos para su aplicación. También facilita que el cumplimiento normativo sea un componente esencial del diseño y puesta en marcha de proyectos de inteligencia artificial.⁴⁶

42. Se sigue de Canton of Zurich, *Innovation Sandbox for Artificial Intelligence (AI)*, <https://www.zh.ch/en/wirtschaft-arbeit/wirtschaftsstandort/innovation-sandbox.html>

43. *Ibidem*. Así, se señala que desde su lanzamiento en 2015, el sandbox de la FCA británica ha prestado apoyo a más de 700 empresas y ha aumentado su velocidad media de comercialización en un 40 % en comparación con el tiempo de autorización estándar del regulador. Se remite a Truby, J. y otros, «A Sandbox Approach to Regulating High-Risk Artificial Intelligence Applications», *European Journal of Risk Regulation*, 2021 <https://www.cambridge.org/core/journals/european-journal-of-risk-regulation/article/sandbox-approach-to-regulating-highrisk-artificial-intelligence-applications/C350EADFB379465E7F4A95B973A4977D>

44. Entre otras fuentes, se sigue Ranchordás, S., «Experimental Regulations for AI: Sandboxes for Morals and Mores», *Morals & Machines*, 1(1), pp. 86-100. <https://doi.org/10.5771/2747-2021-1-86> y Superintendencia de Industria y Comercio, *Sandbox sobre privacidad ... cit.*

https://www.sic.gov.co/sites/default/files/normatividad/112020/031120_Sandbox-sobre-privacidad-desde-el-diseno-y-por-defecto.pdf

45. Ranchordás, S., «Experimental Regulations for AI ...» *cit.*; Guño, A., *Sandbox Regulatorio ... cit.* p. 19 y Superintendencia de Industria y Comercio, *Sandbox sobre privacidad...cit.*

46. Así, se apunta que uno de los proyectos de la quinta cohorte de fintech sandbox de la FCA británica dio lugar a modificaciones normativas. En concreto se apuntan

Desde la perspectiva del mercado y la economía, los sandbox muestran al país que los realiza como una economía innovadora. También mejoran la competitividad en el panorama de la IA para desarrollar nuevas aplicaciones, fomentar la experimentación o incluso desarrollar tecnologías que puedan usarse fuera del país. Los sandbox facilitan la prueba de nuevos productos que, de otro modo, no tendrían acceso a los mercados. La realización de la prueba supone una atracción de inversión, como se ha demostrado en el caso de los fintech y Reino Unido.⁴⁷ La disponibilidad de las autoridades para conocer las nuevas tecnologías es en sí un elemento de atracción de inversión.

Desde el punto de vista de las entidades participantes, el sandbox supone un régimen jurídico beneficioso durante un tiempo limitado. Además, las ventajas de participar en estos contextos de experimentación han sido señaladas por el ICO⁴⁸ y otras entidades: acceso a la experiencia, facilitar el aprendizaje por las partes, mayor confianza en la conformidad de su producto o servicio terminado, una mejor comprensión de la futura normativa RIA y cómo afecta a su empresa o entidad, ser considerado responsable y proactivo en su enfoque de la futura normativa IA por parte de los clientes, otras organizaciones y las autoridades regulatorias responsables, lo que conduce a una mayor confianza del consumidor en su organización y contribuye al desarrollo de productos y servicios que puedan demostrar su valor para el público. También desde la OCDE⁴⁹ se afirma que el 40% de las empresas que completaron el programa sandbox financiero inaugural de la FCA del Reino Unido recibieron posteriormente financiación y la inversión en tecnología financiera fue 6,6 veces mayor. Asimismo, la participación en un sandbox genera dinámicas internas positivas dentro de las organizaciones participantes.

En el caso de las pequeñas y medianas empresas y los startups hay intereses particulares y públicos en juego. El coste de la responsabilidad por daños e impactos de la tecnología y las incertidumbres por el marco jurídico aplicable suponen un coste y riesgo que puede asfixiar la innovación y que no pueden permitirse. La flexibilización de la responsabilidad estricta en el contexto de un sandbox puede ser relevante.⁵⁰

avances en este sentido en la *UK FCA Policy Statement PS19/22* Orientación sobre criptoactivos. También adaptación a medidas de control contra el blanqueo de capitales y la financiación del terrorismo para las iniciativas de incorporación de clientes a distancia por la Autoridad Monetaria de Hong Kong 2020. También, el Documento de Consulta Conjunta 2019 21-402 de los Administradores de Valores Canadienses y la Organización Reguladora de la Industria de Inversión de Canadá: Marco propuesto para las plataformas de negociación de criptoactivos.

47. En el Reino Unido el 30% de las empresas de riesgo que participaron en el sandbox regulatorio recibieron inversión de riesgo, y el monto promedio de inversión aumentó 6,6 veces. Se sigue por Guño, A., *Sandbox Regulatorio ... cit.*
48. ICO, *The Guide to the Sandbox...* cit. También, Truby, J. y otros «A Sandbox Approach...» cit.
49. Se remite a estudios UK FCA, *Regulatory sandbox lessons learned report*, 2017, p. 22, <https://www.fca.org.uk/publication/research-and-data/regulatory-sandbox-lessons-learned-report.pdf> y Goo, J. and J. Heo, «The Impact of the Regulatory Sandbox on the Fintech Industry, with a Discussion on the Relation between Regulatory Sandboxes and Open Innovation», 6 J. *Open Innov. Technol. Mark. Complex*, 2020p. 19 <https://www.mdpi.com/2199-8531/6/2/>
50. Al respecto Truby, J. y otros «A Sandbox Approach to Regulating High-Risk Artificial Intelligence Applications», *European Journal of Risk Regulation*. 2022; 13(2), pp. 270-

Se ha señalado que los costes totales de cumplimiento suponen un 17% del total de los costes de inversión en IA, aunque es probable que esta cifra sea mayor para las PYME que para las grandes empresas debido a las economías de escala.⁵¹ Las PYME también enfrentan barreras de entrada específicas como los procedimientos de estándares o de calidad. Es por ello que el Consejo de la UE «subraya que los “cajones de arena” reglamentarios pueden ofrecer importantes oportunidades, en particular para innovar y crecer, a todas las empresas, especialmente a las PYME, incluidas las microempresas y las empresas de nueva creación, en la industria, los servicios y otros sectores».⁵²

En esta dirección, el RIA afirma el acceso gratuito a sandbox (art. 58.2 d),⁵³ y el artículo 62 concreta que se dará «un acceso prioritario a los espacios controlados de pruebas para la IA, siempre que cumplan las condiciones de admisibilidad y los criterios de selección». Asimismo, en su artículo 58.3^o, recuerda que los actos de ejecución de la Comisión sobre sandbox «ofrecerán a los proveedores potenciales que participen en los espacios controlados de pruebas para la IA, en particular a las pymes y las empresas emergentes, cuando proceda, servicios previos al despliegue, como orientaciones sobre la aplicación del presente Reglamento, otros servicios que aportan valor añadido, como ayuda con los documentos de normalización y la certificación, y acceso a las instalaciones de ensayo y experimentación, los centros europeos de innovación digital y los centros de excelencia».

V. EL MARCO NORMATIVO DE UN SANDBOX DE INTELIGENCIA ARTIFICIAL BAJO EL REGLAMENTO

El RIA pretende evitar la fragmentación de la regulación de la IA entre los Estados miembros, especialmente en el ámbito de los sandboxes. Se había cuestionado que pudiera haber regulación nacional dispersa: «cómo un regulador nacional puede participar plenamente en un sandbox regulatorio cuando el área de regulación cae parcial o totalmente bajo las competencias de la UE».⁵⁴ Entre otras cosas, era necesario evitar que los desarrolladores de IA eligieran Estados de la UE con regímenes de sandbox menos estrictos, así como la dispersión en los métodos de recogida de datos experimentales y los límites. Por ello, se apostó por una regulación común de la UE combinada con la de los Estados.⁵⁵ Sin embargo, se señala que la regulación propuesta podría crear confusión en el mercado, ya que permite que las autoridades competentes de los Estados miembros acuerden una aplicación y un marco comunes que combinen las normas de la UE y de los Estados miembros respecto de un sandbox.

294. doi:10.1017/err.2021.52

51. Se sigue por Truby, J. y otros «A Sandbox Approach...» *cit.* nota 59, 155, 160, 166.

52. Consejo de la Unión Europea, *Council Conclusions...* *cit.*

53. «sin perjuicio de los costes excepcionales que las autoridades nacionales competentes puedan recuperar de una forma justa y proporcionada»

54. Yordanova, K., «The shifting sands of regulatory sandboxes for AI» (KU Leuven, Centre for IT&IP Law, 2019) <https://www.law.kuleuven.be/citip/blog/the-shifting-sands-of-regulatory-sandboxes-for-ai/> Sigo por Truby, J. y otros «A Sandbox Approach...» *cit.*

55. *Ibidem.*

Así, el artículo 53. 1º RIA de la propuesta de la Comisión mencionaba el «cumplimiento de los requisitos establecidos en el presente Reglamento y, en su caso, en otras legislaciones de la Unión y de los Estados miembros supervisadas en el marco del espacio controlado de pruebas». No obstante, esta referencia ha desaparecido y la tendencia a la homogeneidad de la regulación de los sandboxes de IA es clara. Por ejemplo, el artículo 57 con sus 17 apartados implica un régimen común aplicable a los sandboxes de IA en la UE. Especialmente, se ha introducido el artículo 58. 1º: «A fin de evitar que se produzca una fragmentación en la Unión, la Comisión adoptará actos de ejecución que incluirán principios comunes y garantizarán toda una serie de elementos». A lo largo de trece párrafos, se detallan tales elementos comunes que garantizarán los actos de la Comisión Europea. Se afirma que la ordenación de los sandboxes de IA debe garantizar la «flexibilidad para establecer y gestionar sus espacios controlados de pruebas para la IA» (art. 58. 2º c).

Así pues, en la UE todo sandbox de IA debe partir de la regulación de los sandboxes en el propio RIA y de los futuros actos de ejecución de la Comisión (art. 58. 1º). La regulación nacional existente no podrá ser contraria al RIA. Ahora bien, el artículo 57. 4º dispone que «El presente artículo no afectará a otros espacios controlados de pruebas establecidos en virtud del Derecho de la Unión o nacional». Ello implica que otros sandboxes cuya finalidad u objeto esencial no sea la IA se regirán por la normativa nacional general de sandboxes y, en su caso, por su normativa específica. En todo caso, se aconseja que «cuando proceda, las autoridades competentes pertinentes encargadas de esos otros espacios controlados de pruebas deben ponderar las ventajas de utilizarlos también con el fin de garantizar el cumplimiento del presente Reglamento por parte de los sistemas de IA» (Consid. 139). En estos casos, se entiende que los sandboxes sí quedarían bajo el régimen del RIA.

Es importante considerar *el establecimiento y el marco normativo de un sandbox de IA*. En España, no ha habido una regulación general de sandboxes o espacios controlados de prueba hasta el artículo 16 de la Ley 28/2022, de 21 de diciembre, de fomento del ecosistema de las empresas emergentes. Bien es cierto que esta regulación general se da para el ámbito de las empresas emergentes, pero puede entenderse como régimen general de sandbox en España. En el ámbito municipal, destaca la pionera Ordenanza Municipal reguladora del Sandbox Urbano de la Ciudad de València de abril de 2024.⁵⁶ Existen leyes dispersas sectoriales de pruebas, pilotos y sandbox en ámbitos de telecomunicaciones⁵⁷, sector financiero,⁵⁸ ámbito energético,⁵⁹ para el sector público

56. El anteproyecto disponible en https://sede.valencia.es/sede/descarga/doc/DOCUMENT_1_20230005159164 Tramitación en <https://sede.valencia.es/sede/ordenanzas/detalle/MzE2NjQ.AvPAIt3D.AvOvTok>

57. Así, el art. 61.f) de la Ley 9/2014, de 9 de mayo, General de Telecomunicaciones regulaba las autorizaciones para utilizar el dominio público radioeléctrico con fines experimentales, ahora en el artículo 86 f) Ley 11/2022, de 28 de junio, General de Telecomunicaciones.

58. En todo caso, fue pionero el sector financiero y hay que tener en cuenta los artículos 4 a 18 de la Ley 7/2020, de 13 de noviembre, para la transformación digital del sistema financiero.

59. En el ámbito energético, la disposición adicional 23.ª de la Ley 24/2013, de 26 de diciembre, del Sector Eléctrico (según el Real Decreto-ley 23/2020, de 23 de junio) y el Real Decreto 568/2022, de 11 de julio, por el que se establece el marco general del

en la ley de ciencia, la tecnología y la innovación,⁶⁰ evaluación de políticas públicas⁶¹ o sobre convenios con entidades para pruebas piloto en Cataluña.⁶²

Como se ha adelantado, la versión final del RIA obliga a cada Estado a realizar al menos un sandbox de IA de nivel nacional en los 24 meses posteriores a la entrada en vigor (art. 57. 1º). El establecimiento de un sandbox específico de IA en España debe realizarse reglamentariamente,⁶³ a partir de la cobertura general legal reciente. En el ámbito de IA en España, existe el Real Decreto 817/2023, de 8 de noviembre,⁶⁴ que teóricamente estableció un sandbox de IA. Sin embargo, parece un instrumento fallido o abandonado. Una vez adoptado el RIA, el sandbox de IA en España solo podría ser reactivado con una reforma del Real Decreto 817/2023, que habría que cumplir con la regulación del RIA. De lo contrario, tampoco valdría como el sandbox obligatorio que hay que establecer en los primeros 24 meses.

El reglamento que establezca el sandbox de IA en España fijará el marco regulatorio particular del sandbox. Además de dicho reglamento, no será extraño que se dicten bases o condiciones generales de la convocatoria. Las bases, en general, son un acto administrativo general con vigencia definida.⁶⁵ Este instrumento es idóneo para establecer las reglas del sandbox que, al tiempo, inician el proceso de acceso y selección de participantes. Otra fórmula para la regulación concreta de un sandbox es la adhesión, aceptación y suscripción voluntaria por el participante a actos administrativos unilaterales de la Administración en los que se establecen términos, condiciones y régimen de participación en el sandbox. Así sucede con los «Protocolos» en el sandbox financiero.⁶⁶ El RIA impone la existencia de un «plan

banco de pruebas regulatorio para el fomento de la investigación y la innovación en el sector eléctrico.

60. También hay que tener en cuenta la Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación en su redacción dada por la Ley 17/2022, de 5 de septiembre. La misma en su artículo 33.1 prevé medidas innovadoras a través de aceleradoras, incubadoras y centros demostradores; los espacios de experimentación y diseminación; la compra pública de innovación; y los acuerdos marco de servicios para el desarrollo de soluciones que impliquen la introducción de tecnologías disruptivas en la Administración (Art. 33.1.k).
61. Cabe tener en cuenta también la Ley 27/2022, de 20 de diciembre, de institucionalización de la evaluación de políticas públicas en la Administración General del Estado.
62. En el caso de Cataluña hay que tener en cuenta el artículo 64.4 de la Ley 19/2014, de 29 de diciembre, de transparencia que remite a convenios con entidades para pruebas piloto.
63. Art. 16.1º Ley 28/2022, de 21 de diciembre: «Los poderes públicos promoverán, reglamentariamente, la creación de entornos controlados».
64. «que establece un entorno controlado de pruebas para el ensayo del cumplimiento de la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial».
65. En el contexto de convocatorias de personal público son actos administrativos (Real Decreto Legislativo 5/2015, de 30 de octubre, por el que se aprueba el texto refundido de la Ley del Estatuto Básico del Empleado Público). En el caso de las subvenciones, sí que se regulan como disposiciones (artículo 17 de la Ley 38/2003, de 17 de noviembre, General de Subvenciones).
66. El artículo 3 de la Ley 7/2020, de 13 de noviembre, para la transformación digital del sistema financiero regula el «Protocolo», que es el documento en el que se incluyen los términos en los que se realizarán las pruebas. Se suscribirá por el promotor y la

específico acordado entre los proveedores o proveedores potenciales y la autoridad competente» (art. 57. 5º). Estos protocolos o planes específicos están vinculados a las bases de la convocatoria y detallan el régimen específico de obligaciones.

VI. EXCEPCIONALIDAD DEL RÉGIMEN JURÍDICO Y RESPONSABILIDAD DE LOS PARTICIPANTES. TEORÍA, REALIDAD Y REGLAMENTO

Como se expone a continuación, existe una compleja interacción entre la teoría y la realidad respecto a las excepciones o singularidades que implica un sandbox, especialmente uno de IA. Sobre esta base, se analiza la regulación en el RIA. En teoría, *un sandbox regulatorio implica diversas fórmulas de excepcionalidad en la vigencia, aplicación o exigibilidad del Derecho vigente*.⁶⁷ Esta cuestión ha merecido la regulación constitucional en Francia (art. 72), así como jurisprudencia del TC alemán⁶⁸ y del TJUE para la Unión en relación con los sandboxes.⁶⁹ Por ello, debe estar regulado qué disposiciones legislativas pueden ser inaplicadas, excepcionadas, no perseguidas o derogadas, y en su caso, suplidas por decisiones de las autoridades del sandbox. Se debe concretar la naturaleza de las responsabilidades involucradas (civiles, administrativas, etc.) y si son sancionadoras o penales.

Desde el principio de competencia, debe aclararse si el sandbox implica una alteración de competencias o la habilitación, atribución o delegación específica de una autoridad u órgano regulador, que debe estar establecida por una norma del mismo rango que regule la competencia alterada.⁷⁰ También debe concretarse

autoridad o autoridades supervisoras que resulten competentes por razón de la materia del proyecto.

67. Como recuerda BMWi, *Making Space for Innovation... cit.* pp. 82 y ss. Ver Cap. 3. Asimismo, cabe seguir la tipología jurídico-normativa en Conseil d'État, *Les expérimentations... cit.*
68. En particular por la regulación constitucional y las decisiones del Consejo Constitucional en Francia y la importante acción del Consejo de Estado.
En Alemania, además de BMWi, *Making Space for Innovation... cit.*, cabe tener en cuenta las sentencias del Tribunal Constitucional Federal. En primer término les ha dado rango constitucional en su decisión de 16 de junio de 1981 (la llamada tercera decisión sobre radiodifusión), también, la decisión del Tribunal Constitucional Federal de 24 de marzo de 1987, la llamada quinta decisión sobre radiodifusión, «fünfte Rundfunkentscheidung»).
69. En particular, TJUE, Conclusiones Abogado General en el caso C-127/07. Se afirma que la necesidad inherente de diferenciación de un sandbox es compatible con el principio de igualdad de trato siempre que las leyes experimentales tengan un carácter transitorio y el juicio se realice según criterios objetivos.
70. Ver Guño, A., *Sandbox Regulatorio ... cit.* p. 20. existen disposiciones técnicas y procedimentales en las que la ley dispone que sean las mismas autoridades las que definan algunas reglas en la materia. No es extraño ver en distintas regulaciones que el órgano legislativo brinde a las agencias o entes reguladores la potestad para generar algunas reglas específicas, considerando el conocimiento y experiencia que tienen en la materia. De esta forma, la experimentación podrá darse con la generación de una nueva circular u otra documentación al efecto para crear nuevas disposiciones que la autoridad defina, detentando la potestad legal para hacerlo. En este caso es importante considerar el fenómeno de discrecionalidad dentro de las agencias regulatorias que ha sido largamente analizado en la academia norteamericana. Asimismo, dada la imposibilidad que el Congreso brinde todos los elementos específicos que se

a qué autoridades afectarían las excepciones o particularidades (protección de datos, inteligencia artificial, sectoriales, etc.). Ranchordás⁷¹ advirtió que debía ser el propio Derecho de la UE el que regule los márgenes de actuación del Derecho nacional respecto de estas exenciones para evitar la fragmentación. En principio, la excepcionalidad de una norma debe estar establecida en una norma de superior o igual rango.⁷²

Si bien en teoría debe regularse claramente las excepciones o singularidades que implica un sandbox, *en la práctica los sandbox existentes no abordan con claridad las exenciones sancionatorias*. En las experiencias conocidas, esta cuestión no se afronta con claridad. Por ejemplo, el sandbox financiero del Reino Unido de 2015 era consciente de la falta de cobertura normativa del Derecho de la UE para las exenciones, y se afirmaba que «El gobierno podría considerar cambiar las condiciones de exención en FSMA para que sea más fácil para la FCA renunciar a las reglas para una empresa dentro de la caja de arena».⁷³

En España es difícil encontrar una cobertura legal para una exención o inaplicación normativa. El artículo 15.4 de la Ley 7/2020, de 13 de noviembre, que regula el sandbox financiero, establece que, si el participante sigue la ley del sandbox y los protocolos, hay exención o exoneración, pero solo respecto de su participación en el sandbox. Sin embargo, las autoridades mantienen todas sus competencias y las reglas de responsabilidad por daños.⁷⁴

requieren para aplicar una norma, quedan espacios para la interpretación e incluso para decidir cuándo aplicar o no una norma.

71. Ranchordás, S., «Experimental Regulations for AI ...» *cit.*

72. Guño, A., *Sandbox Regulatorio ... cit.* «se pueden generar normas específicas que permitan la experimentación de otras normas que respeten las jerarquías normativas correspondientes. De esta forma, si lo que se quiere experimentar es lo señalado en una ley, es porque otra ley así lo permite. En términos generales, no sería admisible que una norma de menor jerarquía permitiera experimentar el contenido de una ley de rango superior.»

73. Financial Conduct Authority, *Regulatory sandbox*, 2015, pp. 14-15, <https://www.fca.org.uk/publication/research/regulatory-sandbox.pdf>

74. En el sandbox financiero, durante la realización de las pruebas la exención principal es que los promotores no están sometidos a la normativa financiera que les impone una autorización administrativa (art. 4. 2, Ley 7/2020). Por cuanto al cumplimiento del régimen jurídico, hay una «exoneración» y «exención» respecto de actividades específicas por la participación en el sandbox, y «dentro de los límites del proyecto piloto» (art. 4.3, Ley 7/2020).

Es decir, se parte de que si se sigue la ley del sandbox y los protocolos, sí que hay exención o exoneración. Si no siguen protocolos, sí que se aplica la normativa y específicamente se subraya en estos casos la responsabilidad a quienes «infrinjan además normas de ordenación o disciplina» (art. 15).

Ahora bien, entre quienes sí que sigan la ley y los protocolos del sandbox, su exención o exoneración está delimitada a sus actividades relativas a su participación concreta en el sandbox. Pero «En ningún caso, esta exención se extenderá a las actividades ordinarias fuera del espacio controlado de pruebas» (art. 4.3). Ahora bien, en estos supuestos aunque no hay exención sí que se prevé una «ponderación del principio de proporcionalidad» (Art. 4.3 y 19).

Ahora bien, se mantiene la obligatoriedad y la responsabilidad sancionadora a quienes «infrinjan además normas de ordenación o disciplina» (art. 15).

En el ámbito de la IA, no se ha detectado ninguna norma que prevea una exoneración general. Pese a que la esencia de un sandbox es que implique una regulación especial y exención, las instituciones recelan de ver limitada su potestad de control y sancionadora, especialmente si no participan en el sandbox. En el caso español del sandbox de IA, la protección de datos parece evidente y el RGPD no contempla la posibilidad de exoneraciones. El Real Decreto 817/2023, de 8 de noviembre, no contempla excepcionalidad alguna (art. 4 sobre «Régimen jurídico»).

En los casos de los sandbox de IA del Reino Unido, Francia o Noruega, se afirma que no es posible exonerar del cumplimiento de la normativa de protección de datos. Informalmente, en sus guías, webs y publicaciones, se usan fórmulas eclécticas para transmitir que no se perseguirán irregularidades, sino todo lo contrario. Resulta de especial interés la solución informal del ICO,⁷⁵ que señala la obligación de comunicación de posibles irregularidades por la entidad participante, al tiempo que indica la escasa posibilidad de una actuación sancionadora. Por cuanto a la posibilidad de un incumplimiento de legislación sectorial, afirma que no realizará una acción proactiva para comunicar posibles incumplimientos. Cabe recordar en este sentido que las guías de participación, información del sandbox, pese a no tener carácter normativo, sí que puede adquirir algún valor jurídico. Así, pueden servir para la valoración de la culpabilidad concreta de la entidad participante de un sandbox. Quien sigue las indicaciones expresamente formuladas por los diversos canales y puede probarlo, difícilmente puede considerarse que ha realizado una conducta con culpabilidad. Los eufemismos se dan también en Noruega, donde el *Datatilsynet* en 2021 afirmaba que «El sandbox no puede otorgar exenciones de las regulaciones. La Autoridad de Protección de Datos no tiene intención de iniciar medidas

Asimismo cabe recordar que las autoridades mantienen sus competencias (art. 2). Se mantienen en general reglas de responsabilidad por daños (art. 12).

Asimismo, los «monitores» no asumen responsabilidad por los incumplimientos de los participantes (art. 3).

75. <https://ico.org.uk/media/for-organisations/documents/2618111/sandbox-terms-and-conditions.pdf> El ICO británico en sus «Términos y condiciones» señala que «Usted acepta que sigue siendo responsable de su cumplimiento, y del cumplimiento de su Innovación propuesta, con todas las obligaciones legales y reglamentarias, ya sea con respecto a la ley de protección de datos o de otra manera.» (1.6). Y que «El hecho de ser aceptado en el recinto de seguridad no impide la adopción de medidas reglamentarias por nuestra parte o por cualquier otra autoridad competente en materia de protección de datos o por cualquier otro organismo o autoridad reguladora. Los comentarios no afectan a los derechos conferidos a terceros (como sus clientes), ni vinculan a ningún tribunal, y pueden no reflejar las opiniones de ninguna otra autoridad de protección de datos.» (1.10). Ahora bien, en el mismo documento, dando respuesta a «¿Qué ocurre si nos encontramos con una violación de datos personales mientras nuestro producto está en el recinto de seguridad?» además de que se señala que «esperamos que lo comunique a la OIC en un plazo de 72 horas, de acuerdo con el requisito del GDPR del Reino Unido» se afirma expresamente que «Aunque la OIC considerará la infracción de acuerdo con nuestros procedimientos estándar, es muy poco probable que emprendamos una acción coercitiva si está cumpliendo con los términos de su carta de ingreso al sandbox». Asimismo se afirma que «equipo del entorno de pruebas no evaluará de forma proactiva la conformidad de su organización o sus procesos en general. Si identificamos una infracción notificable durante el transcurso del Sandbox [...] le aconsejaremos que lo comunique a la OIC».

*correctivas durante la participación de una organización en el sandbox. El enfoque estará en ayudar a los participantes a cumplir con las regulaciones existentes.»*⁷⁶ Más comedida es la CNIL de Francia. En su web afirma que «Este sandbox no puede conducir al levantamiento de las restricciones regulatorias, ni siquiera temporalmente, porque los textos europeos sobre protección de datos (GDPR) no prevén una exención por este motivo. Sin embargo, sí tiene una vocación experimental, de prueba, para resolver una dificultad o una incertidumbre, identificada en colaboración con el líder del proyecto».⁷⁷

Analizando la cuestión a la luz del RIA, se observa que ha variado considerablemente en su tramitación. La primera versión y la última aprobada son contrarias respecto a la posible exoneración o exención de responsabilidad sancionadora. La versión inicial (art. 53.3) negaba expresamente una exoneración o falta de actuación de las autoridades competentes.⁷⁸ Sin embargo, en la versión final, el punto de partida es que los sandbox «no afectarán a las facultades de supervisión o correctoras de las autoridades competentes que supervisan los espacios controlados de pruebas» (art. 57. 11^o). No obstante, se admite una exención de responsabilidades sancionadoras siempre que los proveedores «respeten el plan específico y las condiciones de su participación y sigan de buena fe las orientaciones dadas por la autoridad nacional competente, las autoridades no impondrán multas administrativas por infracciones del presente Reglamento» (art. 57. 12^o). Esta exención no se extiende a la aplicación de otro Derecho sectorial que sea de aplicación, como la protección de datos. No obstante, se puede eximir del cumplimiento de protección de datos si «otras autoridades competentes responsables de otra legislación hayan participado activamente en la supervisión del sistema de IA en el espacio aislado y hayan proporcionado orientaciones para su cumplimiento» (art. 57. 12^o).

Además del régimen excepcional que pueda implicar el sandbox, es necesario centrar la atención en el *régimen particular de responsabilidad por daños*. Durante la fase de prueba de un proyecto piloto, es posible que se produzcan daños a terceros. Por lo tanto, es necesario definir cómo se asignarán responsabilidades entre los diversos participantes en el proceso. Actualmente, en la Unión Europea, no existe una legislación específica que regule la responsabilidad por los daños causados por sistemas de inteligencia artificial.⁷⁹ El RIA afirma la responsabilidad de los

76. <https://www.datatilsynet.no/en/regulations-and-tools/sandbox-for-artificial-intelligence/framework-for-the-regulatory-sandbox/what-are-the-relevant-regulations/>

77. <https://www.cnil.fr/fr/un-bac-sable-rgpd-pour-accompagner-des-projets-innovants-dans-le-domaine-de-la-sante-numerique>

78. «3. Los espacios controlados de pruebas para la IA no afectarán a las facultades de supervisión y correctoras de las autoridades competentes. Cualquier riesgo significativo para la salud, la seguridad y los derechos fundamentales detectado durante el proceso de desarrollo y prueba de estos sistemas implicará la mitigación inmediata y, en su defecto, la suspensión del proceso de desarrollo y prueba hasta que se produzca dicha mitigación.»

79. Al respecto, Grupo de Expertos sobre responsabilidad y nuevas tecnologías de la Comisión Europea, *Liability for artificial intelligence and other emerging digital technologies*, 2019. <https://op.europa.eu/en/publication-detail/-/publication/1c5e30be-1197-11ea-8c1f-01aa75ed71a1/language-en> De igual modo, desde 2022 hay una Directiva sobre responsabilidad en materia de IA, COM/2022/496 final en trami-

proveedores (art. 57. 12º) «con arreglo al Derecho de la Unión y nacional en materia de responsabilidad, de cualquier daño infligido a terceros como resultado de la experimentación realizada en el espacio controlado de pruebas». Esto sin perjuicio de la responsabilidad sancionadora ya señalada. En España, la responsabilidad por daños sufridos por los participantes está regulada, por ejemplo, en el artículo 12.1 de la Ley 7/2020, de 13 de noviembre, del Sandbox financiero,⁸⁰ o en la no aprobada Ley de Movilidad Sostenible.⁸¹ Además, la mayoría de los sandbox en España que están proyectados o regulados prevén una exclusión de la responsabilidad de las autoridades públicas que intervienen durante el desarrollo del piloto de pruebas,⁸² también en el malogrado sandbox de IA (art. 17 del Real Decreto 817/2023, de 8 de noviembre).⁸³ Pese a las exenciones reguladas, habría que estar al criterio jurisprudencial.⁸⁴ Otro elemento a regular es el régimen de garantías de los participantes, que deben presentarse antes de que comience el piloto de pruebas o sandbox. No obstante, el RIA no regula estas cuestiones.

VII. AUTORIDADES DEL SANDBOX, SELECCIÓN DE PARTICIPANTES, DURACIÓN, DESARROLLO Y OBTENCIÓN DE EVALUACIÓN DE CONFORMIDAD

1. LA AUTORIDAD COMPETENTE DEL SANDBOX Y SU COOPERACIÓN CON OTRAS AUTORIDADES NACIONALES Y EUROPEAS

En virtud del RIA, las autoridades competentes son las que «establecen» un sandbox para la IA (art. 57. 1º). Estas autoridades pueden ser de nivel nacional,

tación sobre la materia, respecto de la que la literatura es muy abundante, Atienza Navarro, M. L., *Daños causados por inteligencia artificial y responsabilidad civil*, Atelier, Madrid, 2022.

80. «La responsabilidad por los daños sufridos por los participantes como consecuencia de su participación en las pruebas será exclusivamente del promotor cuando se produzcan por un incumplimiento suyo del protocolo, se deriven de riesgos no informados por él o cuando medie culpa o negligencia por su parte. En caso de daños derivados de fallos técnicos o humanos durante el transcurso de las pruebas la responsabilidad será igualmente del promotor.»
81. La misma contiene una prolija regulación de la responsabilidad en su artículo 71. Se parte de la una presunción de responsabilidad del promotor, si bien se prevé expresamente la concurrencia de culpas y se afirma también al existencia de un régimen de garantías que se establezca en el protocolo.
82. Así, en el artículo 12.1, Ley 7/2020 sandbox financiero: «Las autoridades que intervengan durante el desarrollo de las pruebas no serán responsables de los posibles daños y perjuicios que pudieran originarse». O la Disposición Adicional 23, Ley 24/2013, de 26 de diciembre, del Sector Eléctrico o en el Anteproyecto de Ley de Movilidad Sostenible, artículo 71.5º.
83. «Tanto el proveedor IA participante como, en su caso, el usuario participante será responsable de los daños sufridos por cualquier persona como consecuencia de la aplicación del sistema de inteligencia artificial en el contexto del entorno controlado de pruebas, siempre que dichos daños deriven de un incumplimiento o cuando medie culpa, negligencia o dolo por su parte.»
84. Así, diversas sentencias consideran la responsabilidad patrimonial de la administración aplicando la *culpa in vigilando*, relacionada con supuestos de defectuosa «inspección o supervisión»: STSJ Aragón 15-2-1999; STS 25-1-1992; STC 112/2018; SAN 24-6-2019.

aunque también se contempla la posibilidad de sandbox «adicionales a escala regional o local o conjuntamente con las autoridades competentes de otros Estados miembros» (art. 57. 2º). El RIA no especifica quién debe ser la autoridad de un sandbox y no tiene por qué ser la autoridad de supervisión del mercado, a diferencia de lo que sucede con las pruebas en condiciones normales. En el contexto de la UE, estos sandboxes serían establecidos por el Supervisor Europeo de Protección de Datos (art. 57. 3º). Las autoridades que los establezcan son responsables de dotar de «recursos suficientes» para cumplir con el RIA.

El establecimiento de un sandbox de IA se realizaría bajo el marco del RIA y la legislación nacional, en España a través de un reglamento y concretado mediante bases específicas. Además, necesariamente habría un «Plan específico» acordado por las partes. La dotación suficiente de recursos debe ordenarse normativa e institucionalmente, así como dotarse presupuestariamente.

Las autoridades que establezcan el sandbox deben proporcionar «supervisión y apoyo dentro del espacio controlado» para determinar los riesgos, apoyo «a las pruebas y a las medidas de reducción y su eficacia» (art. 57. 6º) y dar «orientaciones sobre las expectativas en materia de regulación y la manera de cumplir los requisitos y obligaciones» del RIA (art. 57.7º). La supervisión de los sistemas de IA en el espacio controlado debe comprender su desarrollo, entrenamiento, prueba y validación antes de su introducción en el mercado o puesta en servicio, así como el concepto de «modificación sustancial» y su materialización (Consid. 139).

En la organización de un sandbox debe establecerse una gobernanza que permita incluir a las autoridades de protección de datos u otras, que queden «ligadas al funcionamiento» del sandbox (art. 57. 10º). Las autoridades competentes del sandbox «estarán facultadas para suspender temporal o permanentemente el proceso de prueba, o la participación en el espacio controlado de pruebas si no es posible una reducción efectiva, e informarán a la Oficina de IA de dicha decisión» (art. 57. 11º).

El RIA prevé una cooperación entre las autoridades nacionales de sandbox de IA y el marco europeo de gobernanza de IA. Los sandboxes «para la IA serán diseñados y puestos en práctica de tal manera que, cuando proceda, faciliten la cooperación transfronteriza entre las autoridades nacionales competentes» (art. 57. 13º). Además, «las autoridades nacionales competentes coordinarán sus actividades y cooperarán en el marco del Comité» (art. 57. 14º). Asimismo, «informarán a la Oficina de IA y al Comité del establecimiento de un espacio controlado de pruebas y podrán solicitarles apoyo y orientación» (art. 57. 15º).

Se prevé también un registro de sandbox por la Oficina de IA para facilitar la interacción y la cooperación (art. 57. 15º) y un sistema de informes anuales por las autoridades nacionales a la Oficina de IA y al Comité tras su realización, que serán públicos (art. 57.16º). Toda la información se gestionará a través de una «interfaz única y específica» coordinada por la Comisión (art. 57. 17º).

2. SELECCIÓN Y ADMISIÓN DE PARTICIPANTES Y DURACIÓN DEL SANDBOX

Un elemento relevante es la definición de quiénes pueden ser participantes en el sandbox y los criterios de admisión y el procedimiento básico. Es necesario que la normativa determine con precisión el sector implicado, objetivos y condiciones

de aplicación.⁸⁵ La OCDE recuerda que los sandboxes son selectivos debido a las limitaciones de recursos, y los participantes se seleccionan en función de criterios de elegibilidad. De los 63 solicitantes del sandbox de 2019 del ICO Reino Unido, solo diez fueron seleccionados basándose en criterios claramente determinados.⁸⁶

Los elementos de selección de un sandbox pueden ser similares a los procedimientos de contratación pública⁸⁷ u otros procesos para la atribución de ventajas o subvenciones. El RIA tiene en cuenta esta materia y los actos de ejecución que se adopten para evitar una fragmentación en la UE deberán garantizar que cualquier proveedor pueda acceder con criterios de admisibilidad y selección transparentes y equitativos (art. 58.2 a). Se permitirá «un acceso amplio e igualitario», con la posibilidad de presentarse «en asociación con responsables del despliegue y con otros terceros pertinentes» (art. 58.2 a). También se prevén ventajas igualitarias para PYMES y startups (art. 58.2 d) y acceso de «otros agentes pertinentes del ecosistema de la IA [...] para permitir y facilitar la cooperación con los sectores público y privado». Ejemplos incluyen «los organismos notificados y los organismos de normalización, las pymes, incluidas las empresas emergentes, las empresas, los agentes innovadores, las instalaciones de ensayo y experimentación, los laboratorios de investigación y experimentación y los centros europeos de innovación digital, los centros de excelencia y los investigadores» (art. 58.2 f).⁸⁸

Respecto a *la duración del sandbox*, es un elemento esencial y debe ser fijada por el acto que establece el sandbox.⁸⁹ Las posibles derogaciones, exenciones o no aplicaciones normativas deben ser transitorias⁹⁰ y su duración no puede quedar a la libre decisión de la Administración. Debe formularse la cuestión de «¿Cuánto tiempo se necesitará para alcanzar los objetivos del sandbox regulador?». ⁹¹ La duración debe ser adecuada a la naturaleza de la prueba sobre criterios objetivos, con tiempo suficiente para realizar pruebas representativas y válidas, pero no excesivamente larga o permanente. El plazo puede determinarse por fechas, meses desde el inicio,

85. Conseil d'État, *Les expérimentations...* cit. No obstante, la ley o el decreto deben definir con suficiente precisión el objetivo de la experimentación y las condiciones de su aplicación (CC, n.º 2004-503 DC de 12 de agosto de 2004 o CC, n.º 2019-778 DC de 21 de marzo de 2019).

86. OECD, *Regulatory sandboxes...* cit. p. 25. Así en UK ICO (2019), *Information Commissioner's Office Regulatory Sandbox*, <https://ico.org.uk/for-organisations/regulatory-sandbox/>

87. BMWi, *Making Space for Innovation...* cit.

88. Así se sitúa como ejemplos a «los organismos notificados y los organismos de normalización, las pymes, incluidas las empresas emergentes, las empresas, los agentes innovadores, las instalaciones de ensayo y experimentación, los laboratorios de investigación y experimentación y los centros europeos de innovación digital, los centros de excelencia y los investigadores».

89. El Consejo Constitucional francés ha afirmado que la limitación de su duración es inherente a la experimentación: debe ser fijada por el acto que la instituye. Cuando el legislador decide un experimento, no puede dejar que la autoridad reguladora fije el plazo (CC, n.º 2009-584 DC de 16 de julio de 2009) Ver BMWi, *Making Space for Innovation...* cit. apartado 3.

90. En particular, TJUE, Conclusiones Abogado General en el caso C-127/07, EU:C:2008:728). Se afirma que la necesidad inherente de diferenciación de un sandbox es compatible con el principio de igualdad de trato siempre que las leyes experimentales tengan un carácter transitorio y el juicio se realice según criterios objetivos.

91. BMWi, *Making Space for Innovation...* cit... Apartado 3 Diseño, pp. 80 y ss.

o circunstancias determinables. También puede ser determinable por la autoridad del sandbox a partir de ciertas circunstancias. No se debe excluir la posibilidad de regular posibles ampliaciones del plazo en función de decisiones de autoridades o de algunas circunstancias, dado el elemento de innovación y falta de conocimiento.⁹²

El RIA indica que los actos de ejecución de la Comisión que concreten normas y criterios de sandbox para toda la UE garantizarán «que la participación en el espacio controlado de pruebas para la IA se limite a un período que se ajuste a la complejidad y la escala del proyecto, y que podrá ser prorrogado por la autoridad nacional competente» (art. 58.2 h).

3. DESARROLLO, FINALIZACIÓN Y LOGRO DE UNA EVALUACIÓN DE CONFORMIDAD EN EL SANDBOX

Como se ha señalado, el desarrollo del sandbox implica que las «autoridades competentes proporcionarán, en su caso, orientación, supervisión y apoyo» (art. 57. 6º), y ofrecerán «orientaciones sobre las expectativas en materia de regulación y la manera de cumplir los requisitos» (art. 57.7º). La finalización del sandbox debe implicar acciones de evaluación y retroalimentación general, así como específica para cada participante. Esto puede articularse a través de informes finales, memorias o conclusiones. Es posible regular algunos elementos mínimos de estos documentos y, especialmente, el régimen de publicidad de los mismos. A este respecto, el RIA dispone que, si lo pide el participante del sandbox, «la autoridad competente aportará una prueba escrita de las actividades llevadas a cabo con éxito» y «proporcionará un informe de salida» y «resultados del aprendizaje correspondientes». Esto puede ser relevante para «demostrar su cumplimiento del presente Reglamento mediante el proceso de evaluación de la conformidad o las actividades de vigilancia del mercado pertinentes» (art. 57. 7º). Esto también se debe garantizar en los actos de ejecución de la Comisión (art. 58. 2º e).

El RIA también hace referencia a la importante cuestión de la confidencialidad, regulada con carácter general en el artículo 78. En particular, se señala que «la Comisión y el Comité estarán autorizados a acceder a los informes de salida y los tendrán en cuenta, según proceda, en el ejercicio de sus funciones en virtud del presente Reglamento». Para que los informes de salida de un sandbox sean públicos, el participante debe dar su consentimiento (art. 57. 8º).

VIII. LA PROTECCIÓN DATOS EN EL CONTEXTO DE UN SANDBOX DE INTELIGENCIA ARTIFICIAL

El RIA ha dedicado especial atención a la protección de datos en los sandboxes, en particular al «tratamiento ulterior de datos personales para el desarrollo de determinados sistemas de IA en favor del interés público en el espacio controlado de pruebas para la IA» (art. 59).

92. Algunas cláusulas de experimentación también permiten una ampliación posterior del plazo. La opción de ampliar el proyecto puede ser útil, sobre todo, en el caso de las cláusulas de experimentación con plazos cortos, para aumentar el grado de flexibilidad en la fase de prueba inicial.

Todo uso de inteligencia artificial que implique un tratamiento de datos debe cumplir la normativa de protección de datos y, entre otros aspectos, contar con una base de legitimación para tratar datos personales (art. 6 RGPD). Los sistemas de IA que aspiren a participar en el sandbox cuentan con datos personales para los que, en principio, tienen una legitimación a partir de consentimientos, ejecución de contratos, regulación legal suficiente, etc. No obstante, tratar datos con las finalidades del sandbox puede considerarse una finalidad incompatible, por lo que no podrían tratar datos en este contexto. Aquí es donde el RIA pretende actuar.

El RIA permite que en un sandbox y «únicamente con el objetivo de desarrollar, entrenar y probar determinados sistemas de IA» se traten datos recabados lícitamente para otros fines. El RIA pasa a ser la base legal para hacerlo, siempre que se «cumplan todas las condiciones siguientes». Se trata de diez párrafos de requisitos que deben cumplirse plenamente para legitimar el tratamiento de datos en el sandbox. Entre tales condicionantes se delimitan las finalidades de tratamiento (finalidades esenciales de interés público o que se traten datos para el cumplimiento de obligaciones del RIA), que existan mecanismos de supervisión eficaces, que los datos queden bajo el control de proveedores, que hayan técnicas y organizativas adecuadas y luego se eliminen, que haya un tratamiento de datos funcionalmente separado, aislado y protegido, que no puedan cederse ni salir del sandbox, que no se generen decisiones que afecten a los interesados, que haya un registro de logs, así como descripción completa y detallada del proceso o, finalmente, que se publique una breve síntesis del proyecto de IA. Se prevé que pueda haber alguna regulación específica «para desarrollar, probar o entrenar sistemas innovadores de IA o de cualquier otra base jurídica» (art. 59. 3º).

Pues bien, se trata de una normativa muy detallada y exigente, pero sólo se limita a legitimar tratamientos de datos de origen legales para facilitar el acceso al sandbox. Asimismo, implica normas específicas a seguir para el procesamiento de datos en este contexto. A pesar de la normativa detallada, el artículo 59 no facilita la participación en un sandbox desde la perspectiva de la protección de datos. Esto se debe a que puede generar un efecto inhibitorio para los participantes por temor a ser objeto de un control exhaustivo del cumplimiento de la normativa de protección de datos.

No hay que obviar las dificultades de garantizar un exhaustivo cumplimiento de la normativa en un contexto tan innovador. Buena parte de los diversos sistemas de alto riesgo de IA, son tratamientos de datos que no cuentan hoy día con una clara base de legitimación ni una regulación clara. No es fácil que quienes quieran participar puedan asegurar un cumplimiento exhaustivo de esta compleja normativa. En consecuencia, participar en un sandbox puede suponer *exponerse*, incluso *meterse en una ratonera* de cara a las autoridades de protección de datos. Y el RIA no aborda algunas cuestiones de protección de datos que pueden ser también relevantes.

La información sobre el cumplimiento de la normativa de protección de datos que debe facilitar el proveedor participante puede ser, también, una barrera para el acceso al sandbox. En algunos casos, la autoridad del sandbox puede no ser la de protección de datos, lo que añade complejidad. Una exigencia intensa de demostración del cumplimiento de protección de datos para el acceso y participación puede ser un claro inhibitorio. Y el foco del sandbox no tiene por qué ser el cumplimiento de esta normativa.

En el caso del sandbox de IA en España, el artículo 16 del Real Decreto 817/2023, de 8 de noviembre, afirma que «los proveedores IA y los usuarios participantes en el entorno controlado de pruebas deberán cumplir con lo previsto en» la normativa de protección de datos. En los anexos se incluye una «Declaración responsable de cumplimiento del principio de responsabilidad proactiva en materia de protección de datos» (Anexo IV). Se declara haber adoptado medidas de responsabilidad proactiva y se afirma que se les podría requerir documentación justificativa. También se establece que «el incumplimiento de esta normativa dará lugar al cese definitivo de las pruebas». El Anexo V concreta la documentación «que se podrá requerir» en diez puntos. En buena medida se percibe que la intención no es la de constituir un sistema de control de cumplimiento de normativa de protección de datos en el sandbox, pero sí curarse en salud de problemas que se puedan producir en este sentido y de un mínimo de responsabilidad en la materia.

Es importante señalar que el sandbox puede ser un lugar para identificar incumplimientos de protección de datos. Por principio, la competencia de las autoridades sectoriales no se altera, por lo que la autoridad de datos mantendrá todas sus capacidades de control sobre los participantes del sandbox (art. 57. 11º). No habría exención de sanciones si la autoridad de protección de datos no participa activamente en el sandbox (art. 57. 12º).

Finalmente, otro elemento que podrían tenerse en cuenta es que si se produce un incidente relevante de protección de datos en el sandbox, puede ser obligatorio comunicarlo a la autoridad de protección de datos, lo que puede ser otro un elemento inhibitorio para la participación. En el —malogrado— sandbox de IA en España, el artículo 15 del Real Decreto 817/2023, de 8 de noviembre, impone la obligación de comunicar a la autoridad del sandbox «cualquier incidente grave en los sistemas que pudiera constituir un incumplimiento de la legislación en vigor». Además, los sistemas IA que «estén sujetos a otra legislación específica, el órgano competente trasladará la comunicación a las Autoridades sectoriales competentes, siendo decisión de las autoridades sectoriales ejercer aquellas medidas que considere oportunas».

Estas cuestiones son muy relevantes en la práctica y no han sido reguladas en el RIA. La normativa nacional que establezca el sandbox podría atemperar algunos de los problemas mencionados.

IX. PRUEBAS EN CONDICIONES DE SISTEMAS INTELIGENCIA ARTIFICIAL DE ALTO RIESGO

En la propuesta inicial del RIA por la Comisión no se hacía referencia a las «Pruebas de sistemas de IA de alto riesgo en condiciones reales fuera de los espacios controlados de pruebas para la IA». Estas se introdujeron en las versiones internas del Consejo durante la Presidencia francesa en el primer semestre de 2022 y se han regulado finalmente en los artículos 61 y 62.

El Considerando 141 afirma que «es importante que los proveedores o proveedores potenciales de dichos sistemas también puedan beneficiarse de un régimen específico para probar dichos sistemas en condiciones reales, sin participar en un espacio controlado de pruebas para la IA.» Esto debe hacerse con «garantías y condiciones adecuadas y suficientes», como el «consentimiento informado de las personas físicas», que es distinto al consentimiento de protección de datos. También

se pretende «reducir al mínimo los riesgos y permitir la supervisión por parte de las autoridades competentes». Para ello, se exige presentar ante la autoridad «un plan de la prueba en condiciones reales» y que los proveedores registren la prueba en las secciones específicas de la base de datos de la UE. Asimismo, se requiere un «acuerdo por escrito que defina las funciones y responsabilidades de los proveedores potenciales y de los responsables del despliegue, y una supervisión eficaz por parte de personal competente que intervenga en la prueba en condiciones reales». Se prevén diversas «garantías adicionales»⁹³ y algunas relativas a la transferencia de datos.

Cabe recordar que las pruebas en condiciones reales también pueden realizarse dentro de un entorno controlado de pruebas (art. 57. 7º).⁹⁴ En estos casos, el régimen jurídico aplicable sería el del entorno controlado, pero las autoridades «acordarán específicamente las condiciones», incluyendo «garantías adecuadas con los participantes, con vistas a proteger los derechos fundamentales, la salud y la seguridad» (art. 58.4º).

Respecto de las limitaciones y condiciones de las pruebas en condiciones reales. Las pruebas en condiciones reales están limitadas a los «proveedores o proveedores potenciales de sistemas de IA de alto riesgo enumerados en el anexo III» (art. 60. 1º). En algunos casos, estas pruebas estarán cerca de la investigación. Esto puede sorprender, ya que el RIA no se aplica a «los sistemas o modelos de IA, incluidos sus resultados de salida, desarrollados y puestos en servicio específicamente con la investigación y el desarrollo científicos como única finalidad». Sin embargo, la investigación con miras al desarrollo de sistemas IA para su comercialización puede considerarse aplicable a los proveedores «potenciales» a los que se refiere el artículo 60. De ahí las referencias al consentimiento de los afectados y la superposición con «cualquier revisión ética» que se exija para estas pruebas, como en el ámbito de la investigación (art. 60. 3º).

Por cuanto a requisitos y responsabilidades de los proveedores. Se prevé una posible legislación específica para las pruebas de sistemas de IA de alto riesgo del anexo I, es decir, productos de cierta peligrosidad sometidos a evaluación de conformidad por un tercero que incorporen sistemas IA. El proveedor debe estar establecido en la Unión o haber designado un representante legal (art. 60. 4º d). Es posible organizar las pruebas «en cooperación con uno o varios responsables del despliegue». En estos casos, deben estar bien informados y existir un acuerdo entre encargado y responsables de protección de datos (art. 60. 4º h). El proveedor y los responsables del despliegue deben supervisar efectivamente las pruebas con personal cualificado, formado y con la autoridad necesaria (art. 60. 4º j). Igualmente, deben informar de cualquier incidente grave, adoptar medidas o suspender las pruebas, y tener un procedimiento para la rápida recuperación del sistema de IA (art. 60. 7º). Si se suspenden o terminan las pruebas, deben notificarlo a la autoridad (art. 60. 8º).

93. «Para asegurarse de que sea posible revertir efectivamente y descartar las predicciones, recomendaciones o decisiones del sistema de IA y de que los datos personales se protejan y se supriman cuando los sujetos retiren su consentimiento a participar en la prueba».

94. Así, el artículo 57.7º dispone que «Tales espacios controlados de pruebas podrán incluir pruebas en condiciones reales supervisadas dentro de ellos.»

«El proveedor o proveedor potencial será responsable [...] de cualquier daño» (art. 60. 8°).

Respecto del procedimiento y duración de las pruebas. Las pruebas se podrán realizar «en cualquier momento antes de la introducción en el mercado o la puesta en servicio del sistema de IA por cuenta propia o en asociación con uno o varios responsables del despliegue o responsables del despliegue potenciales.» (art. 60. 2°). La duración será la necesaria y no más de seis meses, prorrogables a otros seis con notificación y explicación a la autoridad (art. 60. 4° f).

El Plan de la prueba es esencial para establecer el régimen jurídico de las mismas, y la Comisión detalla estos planes en un acto de ejecución (art. 60. 1°). Quien pretende realizar la prueba presenta el plan a la autoridad de vigilancia del Estado, quien debe aprobarlo. En principio, se da por aprobado si no responde en treinta días (art. 60. 4° b). Las autoridades de vigilancia del mercado tendrán poderes efectivos para solicitar información y podrán realizar inspecciones sin previo aviso a distancia o *in situ* y controlar la realización de las pruebas (art. 60. 6°).

Por cuanto a la protección de las personas afectadas en las pruebas, el régimen aplicable puede coincidir con supuestos de investigación, de ahí cierta homogeneidad en su tratamiento. Se prevé una «protección adecuada» de las personas afectadas si son «colectivos vulnerables debido a su edad o a una discapacidad» (art. 60. 4° g). Debe asegurarse que «se pueden revertir y descartar de manera efectiva las predicciones, recomendaciones o decisiones del sistema de IA.» (art. 60. 4° k). Asimismo, los afectados pueden «abandonar las pruebas en cualquier momento retirando su consentimiento informado y solicitar la supresión inmediata y permanente de sus datos personales», sin perjuicio alguno (art. 60. 5°). Los sujetos a estas pruebas deben dar un consentimiento informado regulado en el artículo 61, distinto al consentimiento de protección de datos (Consid. 141). Este precepto detalla la información que debe proporcionarse sobre la naturaleza y los objetivos de las pruebas, las condiciones, duración, derechos y garantías, incluido el derecho a abandonar las pruebas, la posibilidad de solicitar la reversión o el descarte de las predicciones, recomendaciones o decisiones del sistema, e información del número de identificación único de las pruebas (art. 61. 1°). Respecto a los datos personales, solo se transferirán a terceros países si se cumplen las exigencias del RGPD.

X. PYMES, STARTUPS Y MICROEMPRESAS EN EL REGLAMENTO

Resta por último hacer referencia a los artículos 62 y 63, que abordan medidas dirigidas a PYMES y empresas emergentes, así como excepciones para microempresas. El RIA pretende considerar especialmente los intereses de las PYMES, incluidas las empresas emergentes, que sean proveedores o responsables del despliegue de sistemas de IA (Consid. 143). Cabe recordar que la Comisión debe aprobar «Directrices» sobre la aplicación del RIA y «prestará especial atención a las necesidades de las PYMES, incluidas las empresas emergentes, de las autoridades públicas locales y de los sectores más afectados por el presente Reglamento». También se considerará la especialidad de estas empresas en la aplicación del régimen sancionador (art. 99. 1° y 6°).

Como se ha mencionado anteriormente, se prevé una prioridad o facilidad de acceso a los sandbox. También el artículo 62 establece que los Estados miembros «organizarán

actividades de sensibilización y formación específicas sobre la aplicación del presente Reglamento adaptadas a las necesidades», «utilizarán canales específicos existentes y establecerán, en su caso, nuevos canales para la comunicación [...] a fin de proporcionar asesoramiento y responder a las dudas planteadas acerca de la aplicación del presente Reglamento» y «fomentarán la participación [...] en el proceso de desarrollo de la normalización». Al fijar las tasas para la evaluación de la conformidad, «se tendrán en cuenta los intereses y necesidades específicos». Destaca el «foro consultivo» regulado en el artículo 67 para integrar los intereses de las PYMEs.

Cabe recordar que las PYMEs y startups pueden cumplir de manera simplificada lo relativo a la documentación técnica (art. 11. 1º). En este precepto, aunque sin referencia explícita a PYMEs y startups, se dispone que la Oficina de la IA contará con modelos normalizados, una plataforma única de información, campañas de comunicación adecuadas y evaluará y fomentará la convergencia de las mejores prácticas en los procedimientos de contratación pública en relación con los sistemas de IA. Cabe recordar que, en virtud del artículo 95. 4º, «la Oficina de IA y los Estados miembros tendrán en cuenta los intereses y necesidades específicos de las PYMEs, incluidas las empresas emergentes, a la hora de fomentar y facilitar la elaboración de códigos de conducta».

Para el caso de «microempresas»⁹⁵ el Considerando 146 y el artículo 63 buscan flexibilizar el cumplimiento del reglamento, aunque de manera muy limitada. Básicamente, el trato de favor se limita a que podrán cumplir «de manera simplificada» algunos elementos del sistema de gestión de la calidad exigido por el artículo 17. Se prevé que «la Comisión debe elaborar directrices para especificar los elementos del sistema de gestión de la calidad que las microempresas deben cumplir de esta manera simplificada» (Consid. 146). No obstante, la excepcionalidad para las microempresas es muy limitada; para evitar cualquier duda, se establece que no se «exime a dichos operadores de cumplir cualquier otro requisito u obligación establecidos en el presente Reglamento, incluidos aquellos que figuran en los artículos 9, 10, 11, 12, 13, 14, 15, 72 y 73» (art. 63. 2º).

XI. CONCLUSIONES Y RECAPITULACIÓN

Este estudio se ha centrado en los sandboxes o espacios controlados, así como en las pruebas en condiciones reales de sistemas de inteligencia artificial. También se han abordado, por estar incluido en este Capítulo V del RIA, en las medidas para PYMES, startups y microempresas.

El RIA pretende facilitar la innovación mediante estas herramientas relativamente nuevas, que aún suscitan no pocos recelos, especialmente entre juristas. De hecho, el RIA impone la obligación a cada Estado miembro de establecer al menos un sandbox en los dos años siguientes a la entrada en vigor del RIA. Frente a la dispersión normativa

95. Las mismas se definen en el artículo 1 de la Recomendación de la Comisión, de 6 de mayo de 2003, sobre la definición de microempresas, pequeñas y medianas empresas: «Se considerará empresa toda entidad, independientemente de su forma jurídica, que ejerza una actividad económica. En particular, se considerarán empresas las entidades que ejerzan una actividad artesanal u otras actividades a título individual o familiar, las sociedades de personas y las asociaciones que ejerzan una actividad económica de forma regular.»

existente en torno a los sandboxes, el RIA establece un marco normativo homogéneo para toda la UE, permitiendo, no obstante, la flexibilidad y la experimentación controlada, cruciales para el desarrollo de tecnologías emergentes.

Se ha observado cómo los sandboxes, inicialmente vinculados a sectores como Fintech, han demostrado ser herramientas efectivas para probar y validar innovaciones tecnológicas en un entorno seguro y controlado. Hay cierta variedad terminológica que, en cierto modo, el RIA resuelve con su concepto de sandbox para sus propios efectos normativos. Se han analizado las experiencias internacionales de sandboxes de IA. Las de Reino Unido, Noruega y Francia muestran que su foco ha estado en el cumplimiento de la protección de datos, aunque el RIA va a impulsar otras perspectivas posiblemente de mayor interés, en todo caso compatibles con la protección de datos. La colaboración internacional y la creación de bases de datos comunes de casos de uso relevantes, contempladas por el RIA, son fundamentales para compartir conocimiento y mejorar las prácticas regulatorias en toda la UE. Asimismo, se han destacado las numerosas ventajas de los sandboxes desde diferentes perspectivas: fomentar la innovación y la competitividad, mejorar la seguridad jurídica y facilitar el acceso al mercado para PYMES y startups. Estos entornos permiten una regulación más adaptable y dinámica, atrayendo inversión y mejorando la competitividad global de la UE en el ámbito de la IA.

A partir de lo anterior, se ha centrado la atención en el marco normativo de un sandbox. Precisamente, el RIA regula este marco normativo para evitar la fragmentación regulatoria entre los Estados miembros. De hecho, la Comisión Europea adoptará actos de ejecución que incluirán elementos comunes que el artículo 58 detalla para la implementación de los sandboxes. Este marco de la UE no excluye que la normativa nacional pueda establecer sandboxes al margen de la IA. Asimismo, siempre que no vulneren el RIA los Estados mantienen la capacidad de regular los sandbox. Se ha expuesto cómo hay un marco legal básico en España y que cada sandbox específico se establece reglamentariamente (como el malogrado sandbox de IA previsto desde 2022 en España). A partir de esta normativa, caben las bases de la convocatoria, protocolos o acuerdos, así como el instrumento del «Plan» del sandbox que regula el propio RIA entre la autoridad y los participantes.

También se ha expuesto un elemento que teóricamente es esencial en un sandbox: la excepcionalidad del régimen jurídico y de la responsabilidad de los participantes. Teóricamente, en los sandboxes es posible eximir temporalmente a los participantes de ciertas obligaciones normativas y responsabilidades sancionadoras. No obstante, las experiencias de sandboxes de IA que ha habido han recurrido a diversos subterfugios para abordar la cuestión. El RIA establece que los proveedores que sigan de buena fe las orientaciones dadas por la autoridad competente no estarán sujetos a multas administrativas por infracciones del RIA durante su participación en el sandbox. Por otra parte, los proveedores siguen siendo responsables de cualquier daño causado a terceros durante la experimentación en el sandbox. En cualquier caso, las autoridades de datos, por ejemplo, podrán seguir aplicando su régimen normativo y sanciones, salvo que participen como autoridad del sandbox. A este respecto, se ha analizado la regulación del RIA respecto de las autoridades competentes a nivel nacional, pero también regional o local, que deben proporcionar supervisión y apoyo continuo a los participantes. También se han destacado las importantes cautelas y requisitos respecto a la admisión y selección de participantes, que deben basarse en criterios de admisibilidad transparentes y equitativos, con acceso prioritario de PYMES y startups. La duración de los sandboxes, en principio, está fijada para seis meses prorrogables. Debe haber un

seguimiento y aprendizaje colaborativo durante el sandbox y al finalizar, la autoridad competente debe proporcionar un informe de salida que documente las actividades realizadas y los aprendizajes obtenidos, lo cual, como novedad, puede ser relevante para demostrar el cumplimiento del RIA y facilitar la evaluación de conformidad.

El RIA también presta atención al tratamiento de datos en sandboxes de IA. La regulación se centra, quizá en exceso, en la legitimación del tratamiento de datos personales por proveedores e implantadores. Se imponen unos requisitos muy precisos, quizá excesivos, para considerar que participar en el sandbox no es un tratamiento incompatible. Sin embargo, no se abordan diversos aspectos de protección de datos que considero que generan un efecto inhibitorio para la participación en sandboxes, debido al temor al control exhaustivo de la normativa de protección de datos. La legislación nacional deberá atemperar estos efectos negativos para no desalentar la participación en sandboxes.

El Capítulo V también regula las innovadoras pruebas en condiciones reales de sistemas de IA de alto riesgo, fuera de los sandboxes. Se trata de una realidad que puede ser muy próxima a la investigación con IA, de ahí que las puedan realizar «proveedores potenciales». Estas pruebas en condiciones reales son fundamentales para evaluar la viabilidad y seguridad de sistemas de IA, pero deben realizarse bajo estrictas condiciones para proteger los derechos fundamentales de las personas afectadas. La regulación del RIA establece un marco claro, aunque complejo, para estas pruebas, destacando la necesidad de un plan de prueba detallado y una supervisión efectiva por parte de las autoridades competentes. Asimismo, y por la mencionada proximidad al ámbito de la investigación, se regula un consentimiento informado de los afectados con una serie de garantías.

Finalmente, se aborda la atención a las PYMES, startups y microempresas en el RIA. Se prevén medidas de apoyo, como el acceso prioritario a sandboxes y la posibilidad de cumplir con requisitos regulatorios de manera simplificada. No obstante, la flexibilidad del cumplimiento del RIA que tienen estas empresas es limitada y deben cumplir con la mayoría de las obligaciones del RIA. Las directrices de la Comisión Europea serán esenciales para especificar cómo pueden simplificar ciertos aspectos del sistema de gestión de la calidad.

Los sandboxes o espacios controlados y las pruebas en condiciones reales de sistemas de inteligencia artificial son herramientas cruciales para la innovación tecnológica y para aprender a la propia implantación del propio RIA.

El RIA ofrece un marco normativo homogéneo para toda la UE, permitiendo la flexibilidad necesaria para experimentar y desarrollar tecnologías emergentes. El RIA garantiza un entorno seguro y controlado para la experimentación. Decenas de experiencias de sandbox se van a producir en los próximos años y se ha de generar un aprendizaje común y una experiencia de la aplicación del propio RIA. Asimismo se verá si es adecuada la regulación de las pruebas en condiciones reales y permiten efectivamente la investigación y la innovación. De igual modo, también el tiempo señalará si es necesario establecer particularidades más concretas para las pequeñas empresas en la UE. El éxito de estas iniciativas dependerá de la colaboración continua entre autoridades nacionales y europeas, y de la capacidad de adaptar la normativa a las necesidades cambiantes del sector.

La gobernanza y vigilancia del Reglamento de inteligencia artificial: autoridades de vigilancia del mercado, Comisión y las diversas entidades

JUAN CARLOS HERNÁNDEZ PEÑA
Profesor Titular de Derecho administrativo
Universidad de Navarra

I. INTRODUCCIÓN

Para establecer un ecosistema de IA fiable, que aquilate seguridad, salud, protección de los derechos fundamentales e innovación responsable, es necesario establecer un sistema capilar y eficaz de gobernanza, así como definir los procedimientos a seguir en casos de incumplimientos.

Este capítulo profundiza en estas cuestiones, iniciando con el análisis del modelo de gobernanza multinivel delimitado por el Capítulo VII del Reglamento. A tales fines, se desarrollan las competencias atribuidas a la Comisión Europea y los Estados miembros, así como a los principales cuerpos administrativos de nueva institución: la Oficina de IA; el Comité Europeo de Inteligencia Artificial; el Foro consultivo; y, el Grupo de Expertos Científicos Independientes.

Posteriormente, nos referiremos a entes regulados que, si bien son comunes a otros ámbitos sectoriales, el Reglamento ordena su establecimiento o les atribuye competencias en el ámbito de la inteligencia artificial. Tal es el caso de las autoridades de vigilancia del mercado o los organismos notificados, previstos ya por la normativa general de seguridad de productos, así como el Centro Europeo para la Transparencia Algorítmica, creado al amparo del Reglamento de Servicios Digitales.

Dedicaremos los apartados finales de este capítulo a analizar las medidas de vigilancia del mercado para garantizar la protección del interés público en caso de duda sobre la conformidad de un sistema de IA. Se trata de los procedimientos de vigilancia del mercado y control de los sistemas de IA en la Unión, que recuperan en buena medida y adaptan los establecidos por el capítulo III del Reglamento 765/2008¹, así como el art. 19 y ss. del Reglamento (UE)2019/1020 (RVM)².

1. Reglamento (CE) N.º 765/2008, de 9 de julio de 2008, por el que se establecen los requisitos de acreditación y vigilancia del mercado relativos a la comercialización de los productos.
2. Reglamento (UE) N.º 2019/1020, de 20 de junio, relativo a la vigilancia del mercado y la conformidad de los productos.

II. EVOLUCIÓN, TRAMITACIÓN Y CONTENIDO FINAL

Respecto a la estructura de gobernanza, la propuesta de Reglamento de la Comisión Europea la Título VI³. Si bien establecía competencias para la Comisión y los Estados Miembros (EM), se centraba especialmente en la creación de un Comité Europeo de Inteligencia Artificial (CEIA), configurado como un ente de asesoramiento, asistencia y consulta técnica, que debía facilitar la aplicación armonizada y coherente, así como la cooperación con el resto de autoridades. En su seno se integrarían las autoridades nacionales de supervisión en materia de IA, el Supervisor Europeo de Protección de Datos y la propia Comisión Europea, a la que correspondería presidirlo y garantizar su funcionamiento. El esquema multinivel se completaba fundamentalmente con las autoridades nacionales competentes (autoridad nacional de supervisión, autoridad de vigilancia del mercado y autoridad notificante), a las que correspondía buena parte de las funciones de supervisión y ejecución del Reglamento⁴.

La posición común del Consejo Europeo (diciembre, 2022) proponía mantener el CEIA, como ente encargado de la implementación armonizada del Reglamento, y asesoría de la Comisión y los EM, pero ampliaba sus cometidos al exigirle estructurar fórmulas que permearan las posiciones e intereses de todos los *stakeholders* del ecosistema. La posición reforzaba el papel de los Estados Miembros en el Comité, proponiendo la designación de cualquier funcionario o miembro de entidades públicas, y reducía las funciones de la Comisión, a la que encargaba el apoyo administrativo y analítico, pero excluyéndola de participar en las votaciones del Comité. En cuanto a las autoridades nacionales, establecía únicamente la obligación de designar al menos una autoridad de vigilancia del mercado, y una autoridad notificante, acercándose más a la estructura establecida por la normativa de armonización de productos de la Unión.

Por su parte, el Parlamento Europeo, en las enmiendas aprobadas en junio de 2023, propuso modificaciones de entidad sustancial. La propuesta del Comité Europeo de IA se sustituía por una Oficina Europea de Inteligencia Artificial, con sede en Bruselas, personalidad jurídica propia y naturaleza de organismo independiente. Aunque la aplicación del Reglamento seguiría descansando en las autoridades nacionales, el apostaba por reforzar la independencia del ente de asesoramiento y coordinación, aunque sin modificar radicalmente sus atribuciones. Respecto a las autoridades nacionales, el Parlamento sigue una línea similar a la mantenida respecto a la autoridad comunitaria de coordinación y asesoramiento. No sólo recuperaba la designación autoridades nacionales de supervisión, sino que las preconfiguraba como administraciones independientes, al exigir que ejerzan sus poderes y desempeñen sus funciones de manera independiente, imparcial y objetiva, sin recibir instrucciones de ningún organismo público.

3. Una valoración general de la propuesta puede verse en Ebers, M., *et al.*, «The European Commission's Proposal for an Artificial Intelligence Act—A Critical Assessment by Members of the Robotics and AI Law Society (RIALS)», *J — Multidisciplinary Scientific Journal*, n.º 4, 2021 pp. 490 y ss.
4. Sobre la estructura de gobernanza de esta propuesta inicial, tuvimos la oportunidad de pronunciarnos en otro momento. Véase Hernández Peña, J.C., *El marco jurídico de la inteligencia artificial. Principios, procedimientos y estructuras de gobernanza*, Thomson-Reuters Aranzadi, Cizur Menor, 2022, pp. 173 y ss.

El texto final se aparta de la propuesta del Parlamento y establece, mediante una solución de compromiso, un sistema de gobernanza multinivel más atenuado donde los EM mantienen buena parte de las atribuciones de supervisión, con excepción de los modelos de propósito general cuyo control se comunitariza. Tales fines, se encargan a la Oficina de IA, aunque alejada de la figura de Agencia Europea, como veremos más adelante. La Comisión y los Estados Miembros mantienen un cúmulo importante de competencias que les permiten dirigir estratégicamente los usos de IA y asegurar un ecosistema fiable, mediante su participación en el CEIA y la designación de las autoridades nacionales competentes, que se asimilan a las que recoge el Reglamento de Vigilancia del Mercado. Finalmente, para asegurar la participación de todos los interesados y garantizar el debido asesoramiento, se crean estructuras de apoyo y consulta a las que se atribuyen diversas funciones (Foro consultivo, grupo de expertos, subgrupos permanentes).

En cuanto a las medidas y procedimientos de vigilancia del mercado y control de los sistemas de IA, más allá de algunas modificaciones menores, las variaciones entre las distintas propuestas se centraron fundamentalmente en la distribución de competencias en atención al modelo de gobernanza que se acogía o la enquistada discusión sobre determinados sistemas que se abordan en otros capítulos. Quizás el cambio de mayor trascendencia es el establecimiento en el texto final de un procedimiento para los sistemas de IA calificados como no de alto riesgo en aplicación del anexo III.

III. EL MODELO DE GOBERNANZA DE LA UE DEL REGLAMENTO

La política europea de IA, que tiene su mayor referente actual, en el RIA, pretende que el uso de esta familia de tecnologías sea ético⁵ y fiable, garantizando un ecosistema de confianza⁶, que refleje la apuesta por los insoslayables valores europeos y los derechos fundamentales⁷.

Para alcanzar este objetivo el diseño de un aquilatado sistema de gobernanza se erige como pilar esencial para la efectiva y armonizada implementación del Reglamento, y en consecuente correa de transmisión que proscriba transar derechos o establecer espacios de fragmentación que permitan arbitrajes regulatorios. En este sentido, como bien señaló el extinto Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial (AI-HLEG), el desarrollo de un ecosistema

5. Sobre esto, entre otros, véase Salazar, I., «El diseño ético de la inteligencia artificial para no discriminar ni lesionar derechos», en Balaguer Callejón, F. y Cotino Hueso, L. (Coords.), *Derecho Público de la Inteligencia Artificial*, Fundación Manuel Giménez Abad, Zaragoza, 2023, pp. 85 y ss.; Moreno Rebato, M., *Inteligencia artificial (Umbrales éticos, Derecho y administraciones públicas)*, Thomson Reuters Aranzadi, Cizur Menor, 2021, *in totum*; Cotino Hueso, L., «Ética en el diseño para el desarrollo de una inteligencia artificial, robótica y big data confiables y su utilidad desde el Derecho», *Revista Catalana de Dret Públic*, n.º 58, 2019, pp. 29 y ss.
6. Gamero Casado, E., «El enfoque europeo de inteligencia artificial», *Revista de Derecho Administrativo — CDA*, n.º 20, 2021.
7. Acerca del sentido y fines últimos del Reglamento ya se ha pronunciado la doctrina. Al respecto, véase Cotino Hueso, L., «Un análisis crítico constructivo de la Propuesta de Reglamento de la Unión Europea por el que se establecen normas armonizadas sobre la Inteligencia Artificial (Artificial Intelligence Act)», *Diario La Ley*, 2021.

de IA fiable sólo es posible con la intervención de mecanismos de supervisión independientes con verdadera capacidad de ampliar las capacidades de la Unión para dar respuesta a la incertidumbre que introduce la permeación generalizada de la IA⁸.

Pero antes de pasar a ello, interesa hacer alguna precisión para contextualizar el diseño recogido por el RIA. La intervención a nivel comunitario pretende un nivel elevado de protección de los intereses públicos, sin menoscabar la competitividad europea ni transar derechos fundamentales⁹. La atribución de funciones y competencias entre los distintos actores del sistema de gobernanza busca distribuir adecuadamente las responsabilidades bajo un paradigma de gobierno multinivel, pero intentando evitar la excesiva fragmentación haciendo uso para ello de una aplicación quirúrgica del principio de subsidiariedad, proporcionalidad y necesidad, mientras garantiza espacios de racionalización de las estructuras administrativas y de cooperación entre los distintos actores implicados.

La fórmula persigue un enfoque de ciclo de políticas que permita un seguimiento sistemático, que abarque desde el desarrollo y evaluación de un complejo programa normativo, hasta la evaluación constante de los sistemas de IA a lo largo de su ciclo de vida. Este modelo estructura de manera concéntrica el marco normativo y los entes de gobernanza. La aplicación y cumplimiento de la cascada regulatoria que tiene como cúspide el Reglamento, pero que incluye normas técnicas y éticas guían el diseño, modelado, puesta en funcionamiento y retirada de los sistemas de IA por parte de los proveedores, representa el primer círculo de protección. La aplicación de esta cascada normativa será verificada por los organismos notificados, abordados en otros capítulos. A su vez, estos órganos notificados han de estar previamente acreditados por los organismos notificantes, a los que corresponde comprobar que aquellos cuentan con los requisitos técnicos para realizar su labor, y reforzar el estándar de protección exigido a nivel comunitario. Esto constituye el segundo círculo. Finalmente, ya puestos en el mercado ha de garantizarse que los sistemas en funcionamiento se ajusten a la normativa, y no entrañen riesgos, contando con ello con un importante cúmulo de competencias para su control y supervisión *ex-post*, mediante procedimientos de seguimiento y control. Estas funciones se atribuyen a la autoridad de vigilancia del mercado o a la Oficina de IA, en atención a las características del sistema de IA concreto.

Como ya se mencionó, en un primer momento abordaremos las autoridades y entes de gobernanza, y más adelante los procedimientos de vigilancia y control

8. AI HLEG, *Policy and Investment Recommendations for Trustworthy AI*, 2019, p. 37.

9. Acerca del impacto de la IA en los derechos fundamentales pueden verse los interesantes estudios de Cotino Hueso, L., «Nuevo paradigma en las garantías de los derechos fundamentales y una nueva protección de datos frente al impacto social y colectivo de la inteligencia artificial» en Cotino Hueso, L. (Dir.), *Derechos y garantías ante la inteligencia artificial y las decisiones automatizadas*, Aranzadi, Cizur Menor, 2022, pp. 69 y ss.; Simón Castellanos, P., «Taxonomía de las garantías jurídicas en el empleo de los sistemas de inteligencia artificial», *Revista de Derecho político*, n.º 117, 2023, p. 155 y ss.; y Aba Catoira, A., «La garantía de los derechos como respuesta frente a los retos tecnológicos» en Balaguer Callejón, F. y Cotino Hueso, L. (coords.) *Derecho Público de la Inteligencia Artificial*, Fundación Manuel Giménez Abad, Zaragoza, 2023, pp. 57 y ss.

recogidos por el propio reglamento, así como por el marco general del nuevo enfoque armonizador comunitario¹⁰.

IV. COMISIÓN EUROPEA. ATRIBUCIONES Y FUNCIONES

Si bien el RIA recoge una pequeña galaxia de entes, estructuras administrativas y de apoyo para su implementación, atribuye a la Comisión un papel relevante para garantizar su aplicación armónica y coherente. Dado que la estructura y funcionamiento de esta institución comunitaria es harto conocida, pasaremos a hacer referencia al importante grupo de atribuciones normativas, de supervisión y control que tiene reservadas en el marco del RIA.

La Comisión desempeña una función fundamental respecto al entramado regulatorio. Le corresponde contribuir con el diseño final del programa normativo, así como velar por su adaptación en atención a los avances tecnológicos. A tales fines, el art. 97 RIA le habilita para modificar cuestiones no esenciales aprobando actos delegados¹¹, mientras que el cons. 175 y el articulado del texto le facultan para dictar de actos de ejecución¹² a fin de asegurar su aplicación uniforme.

Pues bien, con base en actos delegados se le encomienda, entre otras, establecer la metodología y lista de criterios para clasificar los sistemas independientes de alto riesgo (cons. 52); respecto a los modelos de propósito general, le corresponde especificar y ajustar los criterios e indicadores de comparación, incluyendo el umbral para calificarlos de riesgo sistémico en atención al anexo XIII, y designarlos (art. 52.4); o modificar los elementos mínimos de la documentación técnica que han de cumplimentar los proveedores de modelos de propósito general del anexo XI y las obligaciones de transparencia establecidas por el anexo XII (cons.101 y art. 53.5)¹³.

10. Álvarez-García, V. y Tahirí Moreno, J., «La regulación de la inteligencia artificial en Europa a través de la técnica armonizadora del nuevo enfoque», *Revista General de Derecho administrativo*, n.º 63, 2023, p. 5.

11. En consonancia con el art. 290 TFUE, con pleno respeto de los principios de proporcionalidad y subsidiariedad, siguiendo las directrices del acuerdo interinstitucional sobre mejora de la Legislación, *Interinstitutional Agreement Between the European Parliament, The Council of The European Union and The European Commission on Better Law-Making*, de 13 de abril de 2016. Sobre esto, Bradley, K. S. C., «Legislating in the European Union», en Barnard, C. y Peers, S., *European Union Law*, segunda edición, Oxford University Press, Nueva York, 2017, pp. 126 y ss.

12. De conformidad con el art. 291 TFUE. Estas facultades, basta señalar, ha de ejercerlas con respecto del principio de proporcionalidad y subsidiariedad, así como el Reglamento (UE) 182/2011 del Parlamento Europeo y del Consejo, de 16 de febrero de 2011, por el que se establecen las normas y los principios generales relativos a las modalidades de control por parte de los Estados miembros del ejercicio de las competencias de ejecución por la Comisión.

13. Como cabría esperar la lista de actos delegados no se agota con los mencionados. Sin ánimo de exhaustividad, también le compete modificar el listado de legislación de armonización de la Unión correspondiente tanto al antiguo como nuevo enfoque (recogida en el anexo I); adaptar los procedimientos de evaluación de conformidad; el contenido de la declaración UE de conformidad recogida por el anexo V, y la regulación de los procedimientos de evaluación de la conformidad basados en mecanismos de control interno y valoración de los sistemas de calidad

Por otra parte, mediante actos de ejecución se le faculta para aprobar códigos de buenas prácticas para el cumplimiento de obligaciones por parte de proveedores de modelos de propósito general (art. 56.6). También normas que regulen el etiquetado y detección de contenidos manipulados o generados artificialmente, contando para ello con la Oficina de IA (art. 50.7); que definen requisitos y modalidades de creación, funcionamiento y supervisión de espacios controlados de prueba (art. 58), así como las pruebas de sistemas de alto riesgo en tiempo real (art. 60.1)¹⁴.

Además de los poderes para dictar actos delegados y actos de ejecución, tiene competencias para dictar normas de *Soft Law*. Se le habilita para promover directrices relacionadas con los sistemas del anexo III que no se consideran de alto riesgo por excepción¹⁵, o la adopción de metodologías, indicadores de medición y evaluación comparativa para asegurar un nivel adecuado de precisión, robustez y ciberseguridad (art. 15.2). También puede dictar directrices que orienten sobre los elementos fundamentales de sistemas de gestión de la calidad simplificados para sistemas de IA de alto riesgo (art. 63.1) aplicables a PYMES, a fin de calibrar los costes de cumplimiento, y siempre que no deriven en la disminución del estándar de protección de los derechos fundamentales u otros fines imperiosos de interés general.

Además de las responsabilidades previamente mencionadas, la Comisión tiene atribuidas funciones de supervisión y control. Alcanzan una elevada intensidad en cuanto al cumplimiento de las obligaciones por parte de proveedores de modelos de propósito general, cuya supervisión se comunitariza (art. 88)¹⁶. No obstante, esta función debe delegarla en la Oficina de IA, por lo que volveremos sobre ello más adelante.

Si corresponde a la Comisión dictar decisiones individuales designando modelos de propósito general de riesgo sistémico¹⁷, así como reevaluarlos o calificarlos a partir de alertas cualificadas o cualquier otra vía (art. 51 y 52). A tales fines puede solicitar al proveedor la información necesaria para evaluar el sistema (art. 91), incluyendo el acceso al código fuente mediante interfaces de programación de aplicaciones (API) u otros medios técnicos (art. 68.3). También es competente para imponer multas a los proveedores de estos modelos por las infracciones recogidas en el art. 101.1.

de los anexos VI y VII, a efectos de asegurar que sean efectivos (cons. 173 y art. 47.5).

14. Además de lo anterior, mediante actos de ejecución se le habilita para crear el Grupo de Expertos Científicos Independientes (art. 68.1); regular la participación de expertos independientes en las evaluaciones de los modelos de propósito general (art. 92.6); o, impugnar la notificación de un organismo notificado que no cumpla los requisitos, pudiendo suspender o retirar la notificación si el EM notificante no adopta las medidas correctoras procedentes (art. 37.4).
15. Es cierto que esta calificación se realizará aplicando los criterios recogidos en el propio Reglamento (art. 6.3), pero la Comisión puede aprobar orientaciones que especifiquen cómo han de aplicarse, así como un listado que contribuya a identificarlos (art. 6.5).
16. Cons. 162. Con este enfoque se pretende maximizar y centralizar las capacidades de aglutinar experiencia y conocimientos especializados, asegurando una implementación y supervisión coherentes para responder a los riesgos que introducen estos sistemas.
17. Atendiendo a los criterios recogidos en el anexo XIII o equivalentes.

Las competencias de supervisión se extienden al ejercicio de atribuciones de los EM. Se pretende con ello asegurar un elevado de armonización, especialmente respecto a los espacios que confieren deferencia a los EM para completar el programa normativo. Interesa identificar divergencias que generen espacios de fragmentación del mercado común o tiendan a minorar el estándar de protección de los derechos fundamentales. Los sistemas de identificación biométrica en tiempo real, estudiados en otro capítulo, son ejemplo de ámbitos que pueden derivar en prácticas patológicas. De ahí que se atribuya a la Comisión competencias para recabar información sobre las normas nacionales aprobadas por los Estados Miembros (EM) para autorizar excepcionalmente su uso (5.4), y se exige que las autoridades de vigilancia del mercado y de protección de datos remitan un informe anual acerca del uso de dichos sistemas (art. 5.6).

Por otra parte, para asegurar el cumplimiento de obligaciones de transparencia y fomentar la aceptación de la IA, la Comisión debe establecer y mantener una base de datos en la que se registren los proveedores de sistema de alto riesgo distintos a aquellos sujetos a legislación armonizada, así como los del anexo III que por excepción no se consideren como tales (art. 71).

Finalmente, para cerrar el bloque de atribuciones dirigidas a garantizar la aplicación armonizada, corresponde a la Comisión evaluar y valorar el Reglamento, haciendo propuestas de actualización al Parlamento y al Consejo (art. 112).

Las competencias anteriores se refieren fundamentalmente al programa normativo o al ejercicio de poderes jurídicos. No obstante, a la Comisión también se le atribuyan funciones relacionadas con otros medios técnicos que también persiguen promover la confianza en la utilización de la IA. Así, por una parte, le corresponde desempeñar un papel de liderazgo en materia alfabetización de IA¹⁸, a fin de promover su aceptación entre los ciudadanos¹⁹. Para alcanzar este objetivo, el art. 4 recoge una serie de obligaciones, entendiéndose que el desarrollo de estas capacidades es crucial para que la cadena de valor del ecosistema de IA pueda dar cumplimiento adecuado a la normativa. A tales fines le compete, con apoyo del Comité de IA²⁰, desarrollar y promover herramientas de alfabetización que permita a usuarios y afectados comprender sus riesgos y beneficios, así como los derechos y obligaciones que derivan del Reglamento y todo el marco normativo que lo desarrolla. También está

18. Por tal se entiende, el desarrollo de capacidades, conocimientos y comprensión que permitan a proveedores, usuarios y afectados, teniendo en cuenta sus derechos y obligaciones, hacer un despliegue informado de los sistemas de IA, así como adquirir conciencia de las oportunidades y riesgos o posibles daños que esta tecnología puede causar (cons. 56), incluyendo a nuestro entender su impacto social, ético y medioambiental.

19. Esta atribución responde al compromiso asumido por los EM y la UE en distintos instrumentos. Las directrices éticas para una IA fiable la recogen como uno de los mecanismos no técnicos fundamentales para generar un ecosistema de IA fiable, complementando la respuesta regulatoria, códigos de conducta o normalización. AI HLEG, *Ethics Guidelines for Trustworthy AI*, pp. 22-23. En el mismo sentido, la Recomendación de la UNESCO sobre Ética de la Inteligencia Artificial, aprobada con el apoyo de los EM de la UE, la recoge entre sus principios, además de incorporarla entre las políticas públicas que se comprometen a implementar. UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, 2021, pp. 23 y 33.

20. Atribuye esta función el Art. 66.f RIA.

facultada para desarrollar, en colaboración con los EM y la Oficina de IA²¹, códigos de conducta voluntarios dirigidos a desarrolladores, implementadores o usuarios (cons. 20) con propósitos análogos.

También respecto a atribuciones relacionadas con otros instrumentos no técnicos, compete a la Comisión promover la adopción de normas técnicas que recojan el estado de la técnica y contribuyan a mantener un elevado nivel de protección de los ciudadanos²².

Si bien el desarrollo final de estas normas técnicas corresponde a las instituciones europeas de normalización (CEN, CENELEN y ETSI), a la Comisión corresponde adoptar el mandato y encargarles su desarrollo (art. 40.2). Más aún, en ausencia o ante la insuficiencia de dichas normas, está habilitada para dictar especificaciones comunes (art. 41), por lo que en último término ha de asegurar la existencia de instrumentos —incluso excepcionales²³— que permitan a los proveedores demostrar su conformidad, fomentando la innovación sin disminuir la protección de los ciudadanos.

V. OFICINA DE INTELIGENCIA ARTIFICIAL. NATURALEZA, ESTRUCTURA Y FUNCIONES

La creación de la Oficina de IA (OIA) fue objeto de debate en el marco de los trílogos. La propuesta inicial de la Comisión no la preveía, mientras que el mandato del Parlamento la establecía absorbiendo parte de las funciones del Comité de IA. Finalmente, el acuerdo de las negociaciones intergubernamentales estableció un escueto mandato a la Comisión para su institución.

La OIA se establece ahora por el art. 64, si bien diferenciando sus funciones de las atribuidas al Comité de IA. La Comisión dio cumplimiento al mandato, antes de aprobarse el texto final, mediante la Decisión C(2024)390 de enero de 2024 (en adelante la Decisión), entrando en funcionamiento el 21 de febrero del mismo año.

El Reglamento es escueto respecto a su naturaleza, estructura o integración en la Comisión, siendo la Decisión la que nos brinda estas coordenadas. Aunque a nivel comunitario la distinción entre oficinas especializadas y agencias puede resultar algo vidriosa, en nuestro caso la norma de creación de la OIA nos brinda cierta claridad. No nos encontramos ante una agencia, por lo que nos alejamos de las técnicas y características propias de estos cuerpos administrativos comunitarios típicos de la administración compuesta europea²⁴. Tampoco se trata de una entidad de la que pueda predicarse el reconocimiento de personalidad jurídica propia, tal como ocurre con las agencias especializadas.

21. En este sentido, art. 95.2.c RIA.

22. Sobre el sentido y funcionamiento de este tipo de normas, puede verse Álvarez García, V., *Derecho de la regulación económica. VIII. Industria*, Iustel, Madrid, 2010; Álvarez García, V., *Las normas técnicas armonizadas (una peculiar fuente del Derecho europeo)*, Iustel, Madrid, 2020.

23. Véase en este sentido el considerando 121.

24. Sobre esta cuestión, entre otros, Schmidt-Assmann, E., «La administración europea por las Agencias europeas», *Lex Social*, Vol. 3, n.º 2, 2013, pp. 5 y ss.

De hecho, la OIA se aparta de las entidades europeas en el sentido previsto por el Reglamento Financiero 2018/1026²⁵. Se constituye como parte integrante de la estructura administrativa de la Dirección General de Redes de Comunicación, Contenido y Tecnología (DG CONNECT), y se sujeta a su plan de gestión anual²⁶. Por esta razón ha de operar conforme a los procedimientos internos de la Comisión (cons. 7 de la Decisión) manteniendo una estrecha coordinación con los EM, autoridades nacionales competentes, así como otros órganos, agencias e instrumentos de apoyo especializados, como el Grupo de Expertos Científicos Independientes que le apoyará en el desarrollo de sus actividades (art. 68.3). En definitiva, se trata de una estructura integrada dentro de la Comisión que cuenta con cierta autonomía operativa²⁷. Esto acentúa una idea expresada previamente. El diseño institucional del Reglamento, reserva a la Comisión una intensa supervisión y el control estratégico de la IA, a fin de integrarla cohesivamente al resto de políticas comunitarias del mercado digital único.

Lo dicho se aplica especialmente a sus recursos financieros y humanos, si bien goza de cierta autonomía al reconocerle un capítulo de personal propio. Su plantilla está compuesta de miembros ya asignados o reasignados a DG CONNECT, aunque el art. 8 de la Decisión permite la incorporación de personal externo asumiendo los costes mediante la redistribución de partidas presupuestarias del Programa Europa Digital²⁸. Con cargo al mismo instrumento se sufragarán sus gastos operativos, según el objetivo específico recogido por el art. 5 del Reglamento 2018/1046. De esta forma, se asegura que la Oficina cuente con recursos y personal cualificado, con la experticia necesaria, pero sin recargar excesivamente el presupuesto de la Comisión.

El articulado del RIA, así como el art. 3 de la Decisión, recogen las atribuciones y funciones de la OIA. Persiguen, principalmente, que la Comisión y el resto de las estructuras de gobernanza cuenten con las capacidades y conocimientos especializados para asegurar la conformidad y mitigación de riesgos de los modelos de propósito general. A efectos de sistematización agruparemos estas atribuciones en tres bloques diferenciados, si bien conviene interpretarlas con perspectiva integradora.

En primer lugar, encontramos un numeroso grupo de funciones y competencias relacionadas con modelos de propósito general, a los que prestaremos mayor atención. Respecto a éstos, la OIA actúa como correa de transmisión de experticia para garantizar que la Comisión y el resto de autoridades responda proactiva y reactivamente frente a los riesgos de funcionamiento, explicabilidad y transparencia de estos modelos, puedan evaluarlos y determinar si se les aplica adecuadamente el

25. Reglamento (UE, Euratom) 2018/1046 del Parlamento Europeo y del Consejo, de 18 de julio de 2018. Esta norma, en su artículo 2.26 define las oficinas europeas como «una estructura administrativa establecida por la Comisión, o por la Comisión junto con una o varias de las demás instituciones de la Unión, para desempeñar funciones transversales específicas».

26. Cons. 6 y art. 1 de la Decisión C(2024)390 final.

27. Esta integración se recoge sin ambages en el art. 3.47 al definir la IA como «la función de la Comisión consistente en contribuir a la implantación, el seguimiento y la supervisión de los sistemas de IA y a la gobernanza de la IA».

28. Instrumento de financiación recogido por el Reglamento (UE) 2021/694 del Parlamento Europeo y del Consejo, de 29 de abril de 2021, por el que se establece el Programa Europa Digital.

programa normativo. En este sentido, le corresponde desarrollar las herramientas, metodologías e indicadores de referencia para su evaluación, especialmente de aquellos modelos que presentan riesgos sistémicos (art. 3.1.a de la Decisión)²⁹.

También le compete la supervisión a nivel de la Unión de los modelos de propósito general, atribución de especial significado dada la preocupación que han venido generando la regulación de estos modelos durante la negociación del Reglamento. En relación con esto, cabe recordar que se otorgan competencias exclusivas a la Comisión para supervisar el cumplimiento de las obligaciones de los proveedores (cons. 162 y art. 88.1 RIA); función que ha de confiar a la OIA, como exige el art. 88.1. Esta labor tiene una intensidad variable, especialmente elevada respecto a los sistemas que se basen en modelos de propósito general, y tanto modelo como sistema sean suministrados por el mismo proveedor. En este supuesto asume su vigilancia y control, pasando además a ejercer como Autoridad de Vigilancia del Mercado (AVM), contando con las competencias y funciones propias de estos entes (art. 75)³⁰.

Respecto al resto de sistemas de propósito general, el art. 89 faculta a la OIA para adoptar acciones y medidas de seguimiento que aseguren la adecuada implementación por parte de los proveedores³¹. En consonancia con lo anterior, también se le atribuye conducir investigaciones sobre posibles infracciones de proveedores de modelos y sistemas de IA de propósito general, incluyendo infracciones de no conformidad, tanto por iniciativa propia actuando como AVM como a petición de las AVM nacionales³².

-
29. La formulación de estos instrumentos es crucial para comprender las potencialidades y limitaciones inherentes a estos modelos y, en consecuencia, modular quirúrgicamente la intervención de los actores de la cadena de gobernanza. De ahí le especial importancia de revisar y actualizar las metodologías y umbrales de clasificación de modelos de propósito general con riesgos sistémicos.
 30. En ejercicio de estas atribuciones, la Oficina puede recopilar denuncias y reclamaciones de cualquier persona o entidad que cuente con fundamentos para considerar que se han cometido infracciones (art. 85). Esta misma labor se reconoce respecto a los proveedores intermedios, facultados para presentar denuncias por infracciones motivadas y dando cumplimiento a los extremos recogidos por el art. 89.2, así como de los representantes autorizados que pongan fin a su mandato ante el incumplimiento del proveedor de las obligaciones recogidas por el RIA (art. 54.4).
 31. Así, previa consulta al Comité de IA, le compete evaluar si la información suministrada a la Comisión es insuficiente o procede investigar posibles riesgos sistémicos a partir de informes cualificados del Grupo de Expertos Científicos Independientes mediando activación del sistema de alertas (art. 90.2), estando habilitada para establecer diálogos estructurados con los proveedores (art. 91.2); también recibir información y notificaciones de los proveedores en caso de activarse el procedimiento de clasificación de un modelo de propósito general con riesgos sistémicos, así como de incidentes graves relacionados con estos modelos y de las medidas correctivas adoptadas para mitigar su impacto (art. 55.1.c); y, solicitar la documentación técnica de modelos que no sean de uso libre y abierto (art. 53.1.a y 54.2), así como de la información necesaria para demostrar el cumplimiento de las obligaciones recogidas en el capítulo V. En cuanto a esto último, interesa recordar que los modelos puestos a disposición bajo licencia libre y abierta se encuentran exceptuados de esta obligación, salvo que se consideren de riesgo sistémico (art. 53.2).
 32. Así, art. 92, en consonancia con los considerandos 163 y 164.

Por último, siguiendo en este primer bloque de atribuciones, se le faculta para impulsar y colaborar en la aprobación de códigos de buenas prácticas a nivel de la UE dirigidas a modelos de propósito general con riesgo sistémico³³. Similar disposición se recoge respecto a las obligaciones relativas a la detección y etiquetado de contenidos generados o manipulados artificialmente, correspondiéndole en este caso la elaboración de códigos de buenas prácticas (art. 50.7).

Un segundo grupo de responsabilidades faculta a la Oficina para promover la coordinación eficaz entre los cuerpos administrativos del ecosistema de gobernanza, facilitando su asistencia o el intercambio de información. En este sentido, le corresponde apoyar la aplicación de las normas sobre prácticas prohibidas, así como a sistemas de alto riesgo, con miras a alcanzar un elevado grado de armonización y cohesión para no fragmentar el mercado único ni permitir arbitrajes regulatorios por los operadores. También se le reconocen algunas funciones relacionadas con la implementación de los espacios controlados de prueba³⁴; en el ámbito de las acciones de vigilancia del mercado³⁵; y el desarrollo de códigos, criterios, clausulados y plantillas que faciliten la aplicación del Reglamento por las autoridades nacionales y el resto de *stakeholders*, contribuyendo a gestionar la complejidad que conlleva su implementación y cumplimiento³⁶.

El último bloque recoge atribuciones relacionadas con la asistencia a la Comisión en la preparación de actos delegados, de implementación y decisiones, con el fin de

-
33. Corresponsiéndole en cooperación con el Comité de IA, evaluar la adherencia de los proveedores, identificar incumplimientos o inconsistencias en su aplicación y publicar informes acerca de la consecución de sus objetivos (art. 56).
 34. Si bien la implantación y desarrollo de los *sandboxes* compete a las autoridades nacionales (art. 57.1), debe informarse de ello a la OIA, facultada para brindar asesoramiento y orientación en caso de lo soliciten (art. 57.15). Además, dichas autoridades han de informar sobre suspensiones temporales o permanentes de los ensayos, y presentarle un informe anual acerca del funcionamiento de estos espacios (art. 57.11 y 16).
 35. Compete a la OIA prestar apoyo y coordinar el desarrollo de las investigaciones conjuntas (art. 74.11). Igualmente brindar apoyo en caso de investigaciones de modelos de propósito general a cargo de las autoridades nacionales de vigilancia del mercado, recurriendo al procedimiento de asistencia mutua transfronteriza recogido por el art. 22 y ss. del Reglamento 2019/1020.
 36. Así, le corresponde facilitar junto con los EM la elaboración de códigos de conducta voluntarios para proveedores de sistema que no sean considerados de alto riesgo (art. 95). También desarrollar modelos de clausulados contractuales voluntarios para proveedores de sistemas de alto riesgo y terceros que utilicen o integren dichos sistemas (25.4); proporcionar plantillas para: (i) recoger por los proveedores de modelos de propósito general un resumen público detallado sobre los datos utilizados para el entrenamiento (art. 53.1.d); (ii) desarrollar y mantener información unificada sobre los operadores de la Unión; (iii) sensibilizar acerca de las obligaciones establecidas por el Reglamento mediante campañas de comunicación; y, (iv) promover la convergencia de mejores prácticas en el ámbito de la contratación pública de sistemas de IA (art. 62.3.d). Igualmente, se encarga de preparar un cuestionario automatizado para las evaluaciones de impacto sobre derechos fundamentales de los sistemas de alto riesgo, destinado a los responsables del despliegue de estos sistemas, con el fin de facilitar la evaluación de su conformidad con las obligaciones del art. 27. Sobre estas evaluaciones con carácter general, véase Simón Castellanos, P., *La evaluación de impacto algorítmico en los derechos fundamentales*, Aranzadi, Cizur Menor, 2023.

asegurar la aplicación coherente del Reglamento. Esto incluye asistirle, e incluso instar, preparar y actualizar directrices (art. 96.2), así como desarrollar una metodología y orientar las revisiones de los criterios para evaluar los niveles de riesgo, y valorar la inclusión de nuevos sistemas en el anexo III, en la lista de prácticas prohibidas y de sistemas que requieren medidas de transparencia adicionales (art. 112.11).

VI. COMPETENCIAS DE LOS ESTADOS MIEMBROS Y AUTORIDADES NACIONALES COMPETENTES

El RIA asigna un rol fundamental a los EM en su arquitectura de gobernanza. Además de designar a las autoridades nacionales de vigilancia, que abordaremos más abajo, participan en los órganos de gobernanza comunitarios como el CEIA, designando un representante que actuará como punto único de contacto (art. 65.2). También se les atribuye competencia para desarrollar el régimen sancionador y de medidas coercitivas, así como implementarlo ajustándose a las directrices de la Comisión (art. 96 y 99).

Por otra parte, los EM son pieza clave para asegurar la modalidad de la red de gobernanza de vigilancia del mercado³⁷. En este sentido cabe recordar que el Reglamento, una vez aprobado, pasó a integrarse en la legislación armonizada de seguridad de los productos³⁸, por lo que se debe asegurarse la cooperación entre las distintas autoridades encargadas de estas funciones. El RIA no hace más que confirmarlo³⁹. También se les reserva un papel relevante para garantizar la aplicación del principio de innovación, puesto que les corresponde asegurar en último término que las autoridades designadas creen y pongan en funcionamiento al menos un *sandbox*, o garantizar esta obligación mediante cobertura nacional equivalente, como se explica en otro momento de esta obra.

Finalmente, les corresponden algunas competencias implícitas. Así, en aplicación del art. 13 del Reglamento (UE) 2019/1020, deben considerar e incorporar dentro de la estrategia nacional de vigilancia del mercado las prioridades para garantizar y supervisar eficazmente el cumplimiento del RIA. Igualmente, y dado que defensa y seguridad nacional son competencias exclusivas de los EM, les corresponderá

37. Sobre la modalidad como mecanismo de gobernanza digital, véase el clásico trabajo de Hood, C. y Margetts, H., *The Tools of Government in the Digital Age*, Palgrave, Hampshire, 2007, pp. 21 y ss.

38. Esto puede concluirse de la referencia y aplicación del Reglamento (UE) 2019/1020, que tal como señala la Guía Azul sobre la aplicación de la normativa europea relativa a los productos (2022/C 247/01), es una de las fórmulas para delimitar su ámbito de aplicación más allá de lo previsto por su art. 2 y anexo I del RVM. También es evidente de la estructura del RIA que recoge las disposiciones de referencia de la legislación comunitaria de armonización de productos establecida por el Anexo I de la Decisión N.º 768/2008/CE del Parlamento Europeo y del Consejo, de 9 de julio de 2008, sobre un marco común para la comercialización de los productos. No obstante, a efectos de no dejar lugar a dudas, el considerando 76 del RIA lo recoge expresamente.

39. En este sentido, han de facilitar la cooperación con las autoridades competentes en aplicación de la normativa de armonización del anexo II o los sistemas de alto riesgo del anexo III (art. 74.10). Se extiende también a las responsabilidades asignadas a la OIA, las tareas de revisión de la Comisión mencionadas en el artículo 112 y las autoridades nacionales encargadas de la protección de los derechos fundamentales, como la Agencia Española de Protección de Datos.

determinar y regular los sistemas de IA destinados a estos fines, ajustándose estrictamente al perímetro normativo de exclusión que establece el cons. 24 y art. 2.3.

Además de lo anterior, los EM completan la arquitectura de gobernanza a nivel nacional, puesto que designan las autoridades nacionales competentes encargadas de la supervisión y aplicación del Reglamento dentro de sus jurisdicciones. En relación con esto, el RIA replica la estructura prevista en la normativa sobre armonización de productos⁴⁰, fijando la obligación de designar al menos una autoridad notificante y una autoridad de vigilancia del mercado (art. 70.1)⁴¹. Se admite la concentración de todas estas autoridades en una única, si bien asumirá todas las funciones atribuidas. Con esta deferencia se habilita a los Estados para elegir la fórmula que consideren más adecuada según su organización interna, respetando el mandato del art. 5 del TUE. Pueden acudir a un sistema de reparto competencial funcional o geográfico⁴², siempre que se garantice la aplicación uniforme y la eficacia del Reglamento.

Bien se designe una o varias autoridades nacionales, debe garantizarse que actúen con imparcialidad y objetividad. También ha de garantizarse su autonomía, dotándolas de recursos financieros y humanos permanentes idóneos para el ejercicio de sus funciones⁴³.

Además, la versión final del RIA exige que tales autoridades actúen con independencia y e impermeabilidad respecto a cualquier actividad incompatible con sus funciones. Aunque no sea el lugar para pronunciarse con extensión sobre esta cuestión, a nuestro entender los términos recogidos por el Reglamento y el considerando 154 no son contundentes respecto a la exigencia de constituir o designar una autoridad independiente⁴⁴. Pese a esto, en el caso de España algunas de las administraciones que recoge la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público (LRJSP), pueden no ser la más idóneas para extremar el mandato de independencia exigido por el RIA. En este sentido, posiblemente la fórmula adoptada para la creación de la Agencia Española de Supervisión de Inteligencia

40. Específicamente los Reglamento (CE) 765/2008 y (UE) 2019/1020.

41. Esto en consonancia con el art. 10.1 del Reglamento 2019/1020, que atribuye la competencia exclusiva a los EM para asignar estas autoridades y supervisar el mercado en el ámbito de su territorio.

42. Deferencia que facilita el reparto de competencias entre los distintos niveles de gobierno en estados federales o descentralizados, al permitir distribuir la supervisión entre autoridades con distinta competencia territorial interna. Tal es, por ejemplo, el caso de España donde las competencias en materia de vigilancia del mercado están atribuidas a las Comunidades Autónomas. Esta cuestión la analiza con extensión Izquierdo Carrasco, M., *La seguridad de los productos industriales. Régimen jurídico-administrativo y protección de consumidores*, Marcial Pons, Madrid, 2000, pp. 65 y ss.

43. A tales efectos, debe tenerse en consideración la experiencia y conocimientos técnicos sobre tecnologías de inteligencia artificial y ciencia de datos, así como en cuestiones legales que permitan valorar los riesgos para los derechos fundamentales, la seguridad o la salud (art. 70.3). Además, ha de garantizarse el cumplimiento de las obligaciones de confidencialidad en sus actuaciones (art. 78).

44. En línea con lo dicho antes, el Reglamento otorga libertad a los Estados para designar cualquier autoridad siempre que se cumplan con los extremos señalados. Expresamente el considerando 153 establece que «los Estados miembros pueden decidir designar cualquier tipo de entidad pública para que desempeñe las tareas de las autoridades nacionales competentes en el sentido del presente Reglamento, de conformidad con sus características y necesidades organizativas nacionales específicas».

Artificial (AESIA), pese a ajustarse a los extremos exigidos por el Reglamento, sea deficitaria respecto al número de miembros con experticia técnica y un deseable refuerzo de su independencia⁴⁵.

La designación o establecimiento de autoridades nacionales ha de comunicarse a la Comisión para facilitar la colaboración y actuación conjunta con el resto de autoridades nacionales competentes, así como asegurar que cumplen y cuentan con las atribuciones necesarias para desempeñar sus funciones de supervisión. La Comisión está habilitada para exigir información, y los EM obligados a remitir un informe para debate y formulación de recomendaciones. Con esto se pretende fomentar una actuación armonizada, así como una coordinación incremental mediante fórmulas de gobernanza por indicadores⁴⁶ y revisión por pares⁴⁷ generalizadas en materia de seguridad de los productos⁴⁸.

Ahora bien, con independencia de la estructura institucional de cada EM, se encomiendan algunas atribuciones generales a las autoridades competentes. Así, entre otras, son responsables de ofrecer orientación y asesoramiento sobre la aplicación del RIA, especialmente a proveedores a pequeña escala y *start-ups* (art. 70.8). También, y respecto a sistemas de alto riesgo, se les atribuyen competencias para supervisar el cumplimiento de los requisitos horizontales; el ajuste normativo de los que no se consideren de alto riesgo por excepción; y les reconoce competencias para exigir a proveedores y otros actores obligaciones de información y comunicación.

Dicho esto, examinaremos a continuación de manera concisa las competencias y responsabilidades de las autoridades de vigilancia del mercado y autoridades notificantes, no sin antes mencionar que buena parte de éstas se explican con extensión en otros capítulos de este trabajo.

1. AUTORIDADES DE VIGILANCIA DEL MERCADO

La autoridad de vigilancia del mercado (AVM), es definida como la autoridad nacional a la que corresponden las funciones de vigilancia y la adopción de medidas previstas por el Reglamento 2019/1020⁴⁹. Cabe recordar que la vigilancia del mercado

45. Cabe señalar que AESIA se constituyó como Agencia Estatal al amparo de los art. 108 bis y ss. de la LRJSP. Si bien el Estatuto de la Agencia, aprobado por el Real Decreto 729/2003, reconoce cierta independencia de naturaleza técnica, no recoge garantías de independencia para el ejercicio de funciones de sus miembros. De hecho, la Presidencia de AESIA corresponde a la persona titular de la Secretaría de Estado de Digitalización e IA, que además ostenta la Presidencia de su Consejo Rector, cuya mayoría de miembros son altos cargos de la Administración General del Estado. Para subsanar estas deficiencias, la solución idónea sería aprobar mediante Ley una verdadera autoridad independiente de IA, de conformidad con el art. 110 LRJSP.

46. Acerca de la gobernanza por indicadores véase, entre otros, Davis, K., Kingsbury, B. y Merry, S., «Indicators as a Technology of Global Governance», *ILLJ Working Paper*, n.º 2010/2, 2010.

47. Sobre el *Peer Review* y el *Peer Pressure* como técnica de gobernanza, véase Baldwin, R., et al., *Understanding Regulation. Theory, Strategy, and Practice*, segunda edición, Oxford University Press, Nueva York, 2011, p. 431.

48. Así, por ejemplo, véase el art. 12 del Reglamento 2019/1020.

49. Cabe acotar que la excepción la representa el Supervisor Europeo de Protección de Datos, que según lo previsto por el art. 74.9, actuará como autoridad de vigilancia del

se estructura a nivel nacional conforme a las previsiones de este Reglamento. Por tanto, además de las funciones anteriores debe la AVM estar habilitada para desempeñar tareas de coordinación y cooperación a escala comunitaria, tal como exige el art. 10.4 RVM. De ahí que el Reglamento, la designe como punto de contacto único con la Comisión y el resto de autoridades (art. 70.2 RIA).

En relación con esto último, el articulado del RIA le atribuye un conjunto de poderes y obligaciones, que se completan con los recogidos por el art. 14 RVM. Sin ánimo de exhaustividad, debe informar a la Comisión, de manera periódica, sobre las actividades relacionadas con la supervisión del mercado y sobre información de potencial interés relacionada con el Derecho de la competencia⁵⁰. Por otra parte, y dado que la vigilancia del mercado se dirige a proteger a los consumidores, se le encomienda informar y sensibilizar a los ciudadanos y operadores, garantizando el principio de transparencia y los mandatos de la normativa nacional que los desarrolle⁵¹.

En cuanto a las funciones de vigilancia del mercado, son competentes para supervisar y controlar los requisitos de seguridad recogidos por el Reglamento. De manera concreta, se les otorga competencias para supervisar los sistemas de AI diseñados como componentes de seguridad de otros productos sujetos a la normativa «nuevo enfoque», confiriéndoles poderes intensos⁵². Finalmente, se les atribuye competencias vinculadas a algunos procedimientos de evaluación de la conformidad, así como respecto a procedimientos y medidas post-comercialización de vigilancia del mercado, cuestiones ya estudiadas por lo que remitimos a los apartados correspondientes.

2. AUTORIDAD NOTIFICANTE

La autoridad notificante es el ente encargado de designar y notificar a nivel de la Unión los organismos de evaluación de conformidad, en atención a la legislación armonizada y el RIA. En este sentido, es la responsable de establecer y aplicar los procedimientos relacionados con la evaluación, designación, notificación y supervisión de los organismos de evaluación de la conformidad (art. 3.19). Fundamentalmente se encargan de evaluar la capacidad y competencia técnica de los organismos notificados para implementar los procedimientos de evaluación de conformidad de los sistemas de IA, de forma íntegra, imparcial e independiente, y

mercado cuando se despliegan sistemas que entren en el ámbito de aplicación del RIA por parte de las instituciones, órganos y organismos de la UE, con la salvedad del TJUE en su función jurisdiccional.

50. En este supuesto, el art. 74.2 exige además que se comunique a las autoridades nacionales de competencia.
51. En el caso de España, la Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno.
52. Así, por destacar sólo algunos, pueden solicitar acceso a datos de diverso tipo, incluyendo los utilizados para entrenar y validar modelos, valiéndose incluso de medios técnicos que permitan el acceso a distancia, como APIs e interfaces de programación. También, previa solicitud y mediando justificación razonable, pueden acceder al código fuente de sistemas de alto riesgo, con la finalidad de evaluar el cumplimiento de los requisitos horizontales.

de informar su designación a la Comisión y resto de Estados Miembros a través de los sistemas correspondientes⁵³.

Aunque estos organismos y sus procedimientos de notificación se explican con detalle en otros capítulos de esta obra, interesa señalar que el papel de la autoridad notificante trasciende a la designación señalada, puesto que en último término asume la responsabilidad de la competencia técnica y capacidad de los organismos que notifique. Por otra parte, aunque es competencia de las autoridades notificantes establecer los procedimientos de designación y notificación, tanto la normativa general de seguridad de productos⁵⁴, como el Reglamento establecen lineamientos que guían esta labor. Así, han de ser proporcionados, evitar cargas innecesarias a los proveedores en aplicación del principio de innovación, y tener en consideración circunstancias como el tamaño de las empresas y el sistema de IA específico.

En cuanto a su actuación, del Reglamento se extraen una serie de principios que reflejan los mandatos de la normativa general⁵⁵. Así, han de ser imparciales y objetivos, evitar conflictos de interés tanto respecto a las actividades como a los participantes en los procesos de evaluación, y garantizar la competencia personal y técnica al ejercer sus funciones. Su régimen, como es menester, se completa con obligaciones de confidencialidad y sigilo acerca de la información a la que accedan; la prohibición de consecución de ánimo lucrativo; así como el principio de no competencia con los organismos notificados.

VII. COMITÉ EUROPEO DE INTELIGENCIA ARTIFICIAL Y LOS SUBGRUPOS PERMANENTES

El Comité Europeo de Inteligencia Artificial (CEIA) es otra pieza fundamental en la arquitectura institucional creada por el RIA, dado que sus atribuciones no sólo tienen por fin contribuir a su implementación armónica, sino que debe reflejar la diversidad de intereses del ecosistema de IA comunitario (cons. 149). En su seno, además, se crean algunos subgrupos permanentes a los que se encargan cuestiones específicas. Pasaremos a estudiarlos seguidamente.

1. COMITÉ EUROPEO DE INTELIGENCIA ARTIFICIAL. ESTRUCTURA Y ATRIBUCIONES

La composición del CEIA recae, fundamentalmente, en representantes de los EM. También participa el Supervisor Europeo de Protección de Datos, pero en calidad de observador. En versiones iniciales de la propuesta de Reglamento se recogía una participación más activa de la Comisión a la que correspondía presidirlo y gestionar sus sesiones. No obstante, la versión final establece que esté representada por la OIA, aunque sin derecho a voto.

Además de estos miembros naturales, podrán participar —previa invitación— otras autoridades nacionales, organismos o expertos nacionales o de la Unión si

53. Esto es, mediante el sistema informativo NANDO (New Approach Notified and Designated Organisations), administrado por la Comisión.

54. Especialmente la Decisión 768/2008/CE, ya citada.

55. Art. 4 del Reglamento (CE) 765/2008. En la normativa interna, véase el artículo 17 de la Ley 21/1992, de 16 de julio, de Industria.

los temas a tratar les resultan de relevancia. Resulta llamativo que el RIA excluya radicalmente la participación de expertos internacionales teniendo en consideración la función de asesoramiento que ha de desempeñar el Comité. Esta limitación potencialmente restringe el acceso a conocimientos sobre avances sofisticados, y priva a los reguladores de experiencias y perspectivas valiosas para la correcta regulación ética y jurídica de la IA⁵⁶.

Volviendo a los representantes de los EM, se designarán por un período de 3 años, renovable una vez. A diferencia de la propuesta inicial que asignaba la representación a las autoridades nacionales de supervisión, el RIA explicita que los Estados pueden designar con flexibilidad a personas vinculadas a cualquier ente público, siempre que cuente con competencias para coordinar su aplicación a nivel interno y tengan facultades para contribuir a desarrollar las funciones que corresponden al Comité⁵⁷.

El régimen de funcionamiento interno del CEIA lo aprobarán los representantes de los EM. Además, el RIA desapodera a la Comisión y blinda su Presidencia, correspondiendo a uno de dichos representantes elegido mediante el procedimiento y fórmulas acordadas. Ahora bien, el Estatuto que se apruebe en ejercicio de estas potestades de autoorganización ha de garantizar, en todo caso, que la actuación del Comité sea objetiva e imparcial. A la Comisión corresponde proveer la estructura administrativa para su funcionamiento a través de la OIA, brindando apoyo administrativo y técnico especializado, a fin de que las propuestas y recomendaciones sean sólidas y se realicen sobre la base de elementos objetivos.

El CEIA es, fundamentalmente, un órgano de asesoramiento, asistencia y consulta técnica de la Comisión y los EM. Se le atribuye un importante cúmulo de competencias y funciones por el art. 66, que tienen como fin último facilitar la aplicación armonizada y coherente del Reglamento, así como la cooperación con las autoridades de vigilancia del mercado. Al aglutinar autoridades nacionales y comunitarias, está llamado a brindar orientaciones sobre problemas emergentes de la IA que afecten el mercado único, así como recopilar y poner a disposición de los EM y la Comisión las mejores prácticas y conocimientos técnicos.

56. Elevar el Reglamento a estándar global, tal como pretende la UE, no debe construirse por la vía de la aplicación extraterritorial. Sobre esta cuestión, véase entre otros, López-Tarruella Martínez, A., «El reglamento de Inteligencia Artificial y las relaciones con terceros Estados», *Revista electrónica de estudios internacionales*, n.º 45, 2023. La permeación del efecto Bruselas, se asocia también al reconocimiento y legitimidad global de las normativas comunitarias. Por tanto, la participación de expertos —en las cuestiones que se consideren relevantes y de manera aquilataada— permitiría incluir preocupaciones o intereses dignos de protección que pueden estar fuera del radar de los actores europeos, y contribuir a aceptar el RIA como el «*Gold Standard*» de la IA.

57. En el caso de España, esta representación podría ejercerla la Secretaría de Estado de Digitalización e Inteligencia Artificial, bien a través de la persona titular de dicha Secretaría o de la Dirección General de Digitalización e Inteligencia Artificial, dadas las amplias competencias de representación ante instituciones europeas recogidas por el apdo. d) del art. 8.1 del Real Decreto 403/2020 de 25 de febrero, por el que se desarrolla la estructura orgánica básica del Ministerio de Asuntos Económicos y Transformación Digital, así el reconocimiento de competencias de coordinación y cooperación tanto a nivel interministerial como con otras administraciones públicas.

En el marco de su actividad asesora le compete emitir dictámenes, recomendaciones y directrices, así como informes relacionados con la implementación del Reglamento⁵⁸. Por otra parte, y dada la variable densidad normativa del Reglamento, en el complejo reparto competencial establecido, corresponda al Comité —al menos parcialmente— contribuir a completar y revisar el programa normativo. En relación con esto, le compete preparar los actos delegados y de ejecución, así como colaborar con la Comisión en las revisiones periódicas del Reglamento recogidas por el art. 112. Además, se le atribuye la aprobación de normas de *Soft Law* a fin de uniformar prácticas administrativas de los EM⁵⁹. Adicionalmente, el articulado del Reglamento le atribuye funciones relacionadas también con el asesoramiento⁶⁰, e incluso algunas de supervisión y control en corresponsabilidad con otros actores de la estructura de gobernanza⁶¹.

Finalmente, siguiendo el enfoque estratificado y la distribución multinivel de responsabilidades para la supervisión y control de la IA Generativa, el CEIA debe formular recomendaciones y orientación estratégica respecto a la implementación del Reglamento. La intensidad de su participación variará en atención a la categorización de dos niveles que se establece para este tipo de sistemas. Así, ejercerá funciones de orientación y asesoramiento respecto a los sistemas de uso general, pero para aquellos calificados o que presentes riesgos sistémicos le corresponderán actuaciones adicionales. En estos casos, será consultada por la OIA acerca de la necesidad de realizar evaluaciones de modelos específicos de propósito general; y, emitirá dictámenes a la Comisión respecto a las alertas cualificadas (art. 90 y 92).

Resta por señalar que el complejo entramado competencial señalado previamente no perfila completamente los mecanismos a través de los cuáles el Comité fomentará la coordinación, intercambio de información y cooperación entre los actores de la estructura de gobernanza, ni los instrumentos que permitan permear y valorar los intereses de los diferentes agentes de la cadena de valor de la IA. Esto se debe a que estas atribuciones se materializan mediante el Mandato que introdujo Consejo,

58. En este sentido, le corresponde pronunciarse sobre las capacidades técnicas y organizativas de los EM; informar y proponer recomendaciones sobre normas armonizadas y, de ser el caso, especificaciones comunes de sistemas de alto riesgo; y, también respecto de éstos, valorar posibles modificaciones del listado recogido en el anexo III.

59. En este sentido, se hace referencia a dos ámbitos. En primer lugar, las referidas a cuestiones en las que los EM deben cooperar intensamente para asegurar una implementación armonizada, como es el caso de los procedimientos de evaluación de conformidad o las medidas de apoyo a la innovación. En segundo término, las dirigidas a interpretar conceptos y conocimientos técnicos o jurídicos de manera uniforme por autoridades y operadores, debiendo el Comité facilitar directrices o índices de referencia.

60. Así, se le atribuye presentar propuestas de recomendaciones a la Comisión relacionadas los informes anuales de los EM acerca de la idoneidad, capacidad financiera y recursos humanos de las autoridades nacionales competentes (art. 70.6); o fomentar y facilitar, junto con la Comisión, la elaboración y adopción de códigos de conducta voluntarios dirigidos al cumplimiento de requisitos medioambientales (art. 112.7).

61. En este sentido, en colaboración con la OIA, le compete evaluar el cumplimiento de los objetivos previstos en los códigos de conducta a los que hace referencia el art. 112.7.

estableciendo la obligación de crear subgrupos estables, tal como ya se señaló, que pasaremos a estudiar seguidamente.

2. SUBGRUPOS PERMANENTES DE VIGILANCIA DEL MERCADO Y AUTORIDADES NOTIFICANTES

El mencionado art. 65.7 del RIA faculta al Comité para crear Grupos o subgrupos temporales o permanentes que aborden cuestiones específicas del Reglamento. No obstante, ordena la creación del Subgrupo Permanente de Vigilancia del Mercado (SPVM) y el Subgrupo Permanente de Autoridades Notificantes.

El SPVM, que actúa como instrumento de intercambio entre autoridades de vigilancia del mercado, se asocia al mecanismo estable de cooperación que se creó en aplicación del art. 29 del Reglamento (UE) 2019/1020⁶², esto es, la Red de la Unión sobre Conformidad de los Productos (EUCPN)⁶³. Como puede notarse, esto es otra manifestación de la configuración del RIA como una pieza más de la tupida red de legislación de armonización de productos de la Unión.

La incorporación del Subgrupo al EUCPN tiene por fin racionalizar las prácticas de vigilancia del mercado en la Unión, así como reforzar la implementación eficaz del Reglamento, desincentivando con ello la comisión de infracciones⁶⁴. De ahí que en el marco de la actividad de la red el SPVM se nutra de criterios armonizados de vigilancia del mercado aplicables a los sistemas de IA, se complete el marco para el desarrollo de investigaciones coordinadas y se brinde la estructura de apoyo técnico y administrativo al actuar como nodo de contacto y coordinación del Grupo de Cooperación Administrativa (AdCo)⁶⁵.

Por lo tanto, se trata de un grupo informal conformado por las autoridades de vigilancia del mercado y la Comisión⁶⁶. Su presidente es designado por y entre sus miembros, manteniendo reuniones de carácter regular bajo el apoyo administrativo y financiero de la Comisión⁶⁷. Su mandato final es promover que la supervisión de los sistemas de IA sujetos al RIA sea eficiente y responda a los principios de

62. Reglamento (UE) 2019/1020, relativo a la vigilancia del mercado y la conformidad de los productos.

63. EUCPN estructura el apoyo administrativo necesario para integrar y coordinar recursos, así como facilitar la cooperación e intercambio de información entre las autoridades de vigilancia del mercado de la Unión. Está conformada por estas autoridades de cada EM, expertos nacionales y la propia Comisión. Álvarez García, V., *Las normas técnicas armonizadas (una peculiar fuente del Derecho europeo)*, Iustel, Madrid, 2020, p. 53.

64. Cons. 55 del Reglamento (UE) 2019/1020, relativo a la vigilancia del mercado y la conformidad de los productos.

65. La integración del SPVM en EUCPN se implementa mediante su configuración como Grupo AdCo. Por tanto, el diseño final de este subgrupo y sus funciones están recogidas en el Reglamento de Vigilancia del mercado y conformidad de los productos, y no en el RIA que se configura como *lex specialis* de dicho marco normativo.

66. Art. 11.8 del Reglamento 2019/1020.

67. Esta financiación está recogida en el art. 32.e) del Reglamento (UE) 765/2008, de 9 de julio de 2008, por el que se establecen los requisitos de acreditación y vigilancia del mercado relativos a la comercialización de los productos.

vigilancia proactiva, proporcionalidad y cooperación⁶⁸. Las funciones específicas no difieren de las reconocidas para estos grupos por la normativa de vigilancia del mercado (art. 32 RVM), y tienen por fin ganar enteros en cuanto a la eficacia de la actividad de vigilancia de mercado proactiva⁶⁹ y reactiva. En relación con esto último, en caso de incidentes, accidentes o reclamaciones, esto es el ámbito propio de la actividad de vigilancia reactiva del mercado, el Subgrupo AdCo de IA se servirá del Sistema Europeo de Información y Comunicación para la vigilancia en el mercado (ICSMS), plataforma de apoyo para el intercambio de información y coordinación de actividades entre los AdCos de productos no alimentarios, que actuará como nodo para informar sobre incidentes en el uso de productos y sistemas de IA presuntamente no conformes o que presenten riesgos.

Por su parte, al Subgrupo Permanente de Autoridades Notificantes (SPAN), se le encarga la cooperación en cuestiones relacionadas con los organismos notificados, sin especificar con mayor detalle su composición y funciones. Este escueto mandato, no obstante, ha de completarse con la normativa europea relativa a productos, y más concretamente con lo previsto por la «Guía Azul»⁷⁰, que también recoge el mandato de establecer instrumentos de cooperación.

VIII. OTROS ENTES DE ASESORAMIENTO, APOYO Y COLABORACIÓN

Además de las entidades de gobernanza ya examinadas, el Reglamento introduce otras adicionales destinadas: el Foro Consultivo, el Grupo de Expertos Científicos Independientes (GECI), y el Centro Europeo para la Transparencia Algorítmica (ECAT), junto a las estructuras de apoyo a pruebas de IA. Están diseñadas o se vinculan para integrar diversas perspectivas, así como experticia y conocimiento técnico a fin de promover una implementación armónica y efectiva de la IA en el mercado único.

1. FORO CONSULTIVO

Para responder equilibradamente a los desafíos de la IA, no basta con la pericia de reguladores armados con un potente arsenal de poderes y competencias. Es necesario incluir y valorar la perspectiva del resto de actores del ecosistema europeo. En este sentido, la incorporación de éstos a través del Foro Consultivo es un acicate para que la implementación del RIA sea efectiva, y se acompañe a la rápida evolución

68. En consonancia con el Art. 30 del Reglamento 2019/1020.

69. Respecto a la vigilancia proactiva, el Subgrupo AdCo de IA se encargará de promover la aplicación uniforme del RIA y el Reglamento 2019/1020, fomentar la comunicación y confianza entre las autoridades de vigilancia del mercado de los EM, coordinar proyectos conjuntos y desarrollar metodologías comunes para asegurar una supervisión efectiva, especialmente en caso de actividades transfronterizas. Además, este mecanismo permite intercambiar información acerca de mejores prácticas y alinearlas con las generales ya recogidas en el ámbito de la supervisión del mercado, abordando cuestiones de interés común para proponer enfoques unificados y facilitar evaluaciones específicas del sector, incluyendo análisis de riesgo y avances científicos. En definitiva, permite optimizar y racionalizar las actividades de control previas.

70. Comunicación de la Comisión «Guía azul» sobre la aplicación de la normativa europea relativa a los productos, (2016/C 272/01), de 26 de julio de 2016.

tecnológica integrando experticia técnica con las distintas sensibilidades económicas y sociales. De esta forma se aquilata la coexistencia del enfoque basado en riesgos con el principio de innovación.

La aproximación del Reglamento a esta cuestión, aunque más amplia, no introduce novedades respecto al marco comunitario de seguridad de los productos, que ya exigía la representación y participación efectiva de los actores interesados⁷¹. El Reglamento adopta y amplía esta estrategia colaborativa, como se refleja en su considerando 150 y art. 67, con el objetivo de fomentar la legitimidad y aceptación del marco comunitario.

Al foro corresponde aportar conocimientos técnicos adicionales, y se le reconoce facultad para preparar opiniones y recomendaciones dirigidas al CEIA y la Comisión. Su composición debe reflejar el esfuerzo por incorporar el amplio espectro de intereses y perspectivas de los *stakeholders* del ecosistema de IA de la Unión, incorporando equilibradamente a la academia, industrias y empresas, así como startups, PYMES y sociedad civil. Esta representación busca promover un diálogo constructivo, atendiendo las implicaciones para distintos sectores y colectivos, así como la consideración de intereses comerciales, económicos y sociales.

La Comisión debe designar a los miembros del Foro por su probada experiencia en inteligencia artificial, garantizando la diversidad. Su mandato, inicialmente de dos años, puede extenderse a un máximo de cuatro. Debe contar con autonomía para aprobar su propio reglamento y elegir copresidentes entre sus miembros. Se reunirá, al menos, semestralmente pudiendo invitar expertos y otros actores del ecosistema de IA, si bien la Agencia de los Derechos Fundamentales de la Unión (FRA), ENISA, y los organismos europeos de normalización (CEN, CENELEC y ETSI) tendrán el estatus de miembros permanentes.

2. GRUPO DE EXPERTOS CIENTÍFICOS INDEPENDIENTES

El Acuerdo Provisional de las Negociaciones Interinstitucionales introdujo la figura del Grupo de Expertos Científicos Independientes (GECI), para apoyar la aplicación e implementación del Reglamento, especialmente respecto a los modelos de propósito general (art. 68). La Comisión debe crearlo y ponerlo en funcionamiento mediante un acto de ejecución, siguiendo el procedimiento de examen recogido por el art. 5 del Reglamento (UE) 182/2011⁷².

La Comisión determinará el número de expertos tras consultar al CEIA, asegurando una representación geográfica y de género equitativa. Serán seleccionados basándose, como mínimo, en su probada experiencia científica o técnica en IA; su independencia respecto a proveedores de sistemas o modelos de propósito general; y su capacidad para actuar de manera diligente y objetiva. Están sujetos a obligaciones de imparcialidad, objetividad y confidencialidad, operando sin recibir instrucciones de terceros para preservar su independencia. Para promover la transparencia, la

71. Véase, así, el art. 5 Reglamento (UE) 1025/2012.

72. Reglamento (UE) 182/2011 del parlamento europeo y del consejo, de 16 de febrero de 2011, por el que se establecen las normas y los principios generales relativos a las modalidades de control por parte de los Estados miembros del ejercicio de las competencias de ejecución por la Comisión.

OIA implementará procedimientos específicos para prevenir conflictos de interés, incluyendo la obligatoria publicación de declaraciones de intereses accesibles públicamente.

Aunque el espectro de responsabilidades del GECI es más amplio, su rol se concentra en brindar apoyo y asesoramiento técnico sobre modelos de propósito general (cons. 151 y art. 68.3). Este rol se diferencia del atribuido al Foro Consultivo, que no sólo incorpora conocimiento técnico de vanguardia, sino que también abarca una perspectiva más amplia prestando atención a las diversas sensibilidades sectoriales. De este modo, las funciones de ambos entes se complementan y enriquecen el marco de gobernanza con una visión sólida, integral y multidisciplinar.

Las funciones del GECI resultan instrumentales para asegurar la eficacia de la OIA, proporcionando un marco de referencia científico y técnico indispensable. Entre sus atribuciones destaca la emisión de alertas cualificadas a la OIA sobre modelos de propósito general que presenten riesgos sistémicos a nivel comunitario. Igualmente, la elaboración de metodologías que evalúen tanto las capacidades como los riesgos asociados a los sistemas de IA de uso general, fortaleciendo la capacidad de respuesta de la Oficina ante situaciones que requieren medidas de salvaguardia.

Si bien las funciones de asesoramiento se dirigen especialmente a la OIA, se extienden a las autoridades de vigilancia del mercado, correspondiendo a la Comisión garantizar el acceso al *pool* de expertos, aunque ello pueda conllevar el pago de tasas y honorarios (art. 69).

3. CENTRO EUROPEO PARA LA TRANSPARENCIA ALGORÍTMICA Y LAS ESTRUCTURAS DE APOYO A LAS PRUEBAS DE INTELIGENCIA ARTIFICIAL

El Centro Europeo para la Transparencia Algorítmica (ECAT), establecido en el marco del Reglamento de Servicios Digitales⁷³, está llamado a colaborar activamente con la OIA⁷⁴. Sin querer agotar su regulación y funciones, se configura como una estructura de apoyo de DG Connect integrada en el Centro Común de Investigación de la Comisión. Su labor principal es ofrecer asesoramiento científico y técnico en investigaciones sobre sistemas algorítmicos implementados por plataformas en línea, a fin de asegurar que se ajustan al Derecho comunitario⁷⁵.

Al ECAT se le atribuye, en general, la inspección y desarrollo de pruebas técnicas de sistemas algorítmicos para entender su funcionamiento, identificar y cuantificar riesgos sistémicos de plataformas en línea y motores de búsqueda de gran tamaño (VLOSEs), y, en definitiva, desarrollar metodologías para valorar la equidad de los modelos algorítmicos. Dada esta experticia, así como el hecho de que algunos modelos algorítmicos sujetos a la DSA pueden aplicar modelos de IA, la colaboración entre el ECAT y la OIA es indispensable. Las fórmulas específicas para implementar

73. Reglamento (UE) 2022/2065 del Parlamento Europeo y del consejo, de 19 de octubre de 2022, relativo a un mercado único de servicios digitales.

74. Así lo recoge el art. 5.2.a de la Decisión de Creación de la Oficina de Inteligencia Artificial.

75. La doctrina se ha ocupado de ellos. Al respecto, véase entre otros, Ilichman, D., «European Approach to Algorithmic Transparency», *Charles University in Prague Faculty of Law Research Paper*, n.º 2023/II/1, 2023, p. 11 y ss.

esta cooperación aún están por definirse, *per se* atribuye a la OIA la responsabilidad de desarrollar tales mecanismos (art. 3 de la decisión).

En cuanto a las estructuras de apoyo a las pruebas de IA, el art. 84 encomienda a la Comisión designar instalaciones de ensayo de la Unión para este fin. Este tipo de instalaciones, cuya regulación se encuentra recogida en el Reglamento 2019/1020 (art. 21), apoyan la labor de las Autoridades de Vigilancia del Mercado, la EUPCN, la Comisión y otras entidades públicas.

Una vez designadas por la Comisión, se les atribuye las mismas funciones generales recogidas por el art. 21.6 del RVM, pero en el ámbito de la IA. Esto es en nuestro caso, realizar pruebas y evaluaciones de sistemas y productos o con componentes de IA previa solicitud de las AVM, la OIA o la Comisión; brindar asesoramiento técnico o científico independiente en el marco de la cooperación que pueda desarrollarse a través de EUPCN y las AdCos; y, desarrollar nuevas técnicas y métodos de análisis para valorar los riesgos y modelos de IA. Además de esto, tanto el CEIA, como la Comisión o las AVM pueden solicitarles asesoramiento técnico o científico independiente, si lo estiman pertinente.

IX. PROCEDIMIENTOS DE SALVAGUARDIA, VIGILANCIA DEL MERCADO Y CONTROL DE LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL EN LA UNIÓN

Como ya se ha comentado en varios momentos, el RIA forma parte de la legislación de armonización de la Unión y, por tanto, debe ajustarse la estructura institucional, técnicas y procedimientos previstos por el Reglamento 765/2008, el RVM y la Decisión 768/2008.

Aunque es de sobre conocido, conviene recordar que el anclaje de este poder de armonización y aproximación de legislaciones de la Unión está recogido en el art. 114 TFUE⁷⁶, que además se configura como una de las bases del RIA⁷⁷. Esta previsión admite que la legislación de armonización recoja y autorice el uso de una cláusula de salvaguardia por los EM (art. 114.10), a partir de la cual desarrollaron procedimientos específicos de salvaguardia, que hoy están regulados con carácter general por el capítulo III del Reglamento 765/2008 y la Decisión 768/2008⁷⁸. El RIA, actúa como *lex specialis* por lo que completa y especifica ese marco general en los art. 79-83, que pasaremos a estudiar seguidamente.

No obstante, y aunque esto se aborda en otros capítulos de esta obra, interesa señalar que estos procedimientos de salvaguardia y medidas de vigilancia del mercado se activan una vez se identifican, detectan o notifican sistemas de IA desplegados o en funcionamiento en el mercado pero que presentan un riesgo grave que puede afectar negativamente la salud, seguridad o derechos fundamentales de

76. Sobre esto se ha ocupado la doctrina. Entre otros, Barnard, C., *The Substantive Law of the EU. The Four Freedoms*, séptima edición, Oxford University Press, Nueva York, 2022, pp. 557 y ss.

77. La otra como ya se comentó en otro capítulo de este trabajo es el art. 16 TFUE, que sirve de base para regular algunos usos de IA que conllevan tratamientos de datos personales.

78. Sobre estos procedimientos, con carácter general, véase Álvarez García, V., *Derecho de la regulación económica. VIII. Industria*, Iustel, Madrid, 2010, pp. 474-475.

las personas⁷⁹. La normativa comunitaria general regula los supuestos de riesgos graves que tienen incidencia nacional, diferenciándolos de aquellos que pueden alcanzar otros EM. También los supuestos de productos conformen que, no obstante, planteen un riesgo para la salud y la seguridad, y los incumplimientos formales. Todos estos son recogidos por el RIA que, no obstante, introduce un procedimiento específico para los sistemas calificados por los proveedores como no de alto riesgo en aplicación del anexo III.

1. PROCEDIMIENTO APLICABLE A LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL QUE PRESENTEN UN RIESGO A NIVEL NACIONAL

La normativa de vigilancia del mercado regula un primer procedimiento en caso de incumplimiento o incidente que se limite al territorio de un Estado miembro. El procedimiento, recogido por el art. 79, específica para los sistemas de IA los lineamientos recogidos por el Reglamento (CE) 765/2008, el art. 19 del RVM, así como el art. R31 del anexo I de la decisión 768/2008/CE. En este sentido, si una autoridad de vigilancia del mercado tiene indicios de la existencia de un riesgo, iniciará el procedimiento con miras a evaluar si los riesgos son de gravedad y se está incumpliendo con los requisitos y obligaciones recogidas por el RIA, especialmente en el caso de los sistemas recogidos por el art. 5 que afectan personas vulnerables. Para ello, cuentan con el cúmulo de poderes señalados, y que se refuerzan por el art. 14.3 RVM, y de manera particular potestades de acceso a la información señaladas previamente, estando los proveedores y demás operadores obligados a cooperar. A dichos operadores, así como a otros del ciclo de vida del sistema de IA afectados, se les notificará del procedimiento. Además, se informará en caso de riesgos para los derechos fundamentales a las autoridades nacionales con competencia para supervisar y garantizar las obligaciones en esta materia, debiendo cooperar con éstas.

Dado que el fundamento último del procedimiento es garantizar una rápida intervención para evitar la materialización de los riesgos o contener su expansión, una vez acreditado el riesgo se exigirá al operador correspondiente bien que adopte medidas correctoras para ajustarse a la normativa, o bien para retirarlo⁸⁰ o recuperarlo⁸¹ del mercado, en un plazo razonable y proporcional que establezca la autoridad de vigilancia en atención a la naturaleza del riesgo, si bien no podrá exceder de 15 días hábiles o del plazo que prevé la legislación de armonización pertinente.

Si el operador correspondiente no adopta las medidas, el reglamento faculta a la autoridad de vigilancia del mercado para adoptar medidas provisionales que habrán de ajustarse a las previsiones del Derecho interno. Estas medidas pueden ser todas aquellas que resulten adecuadas y coherentes a fin de retirar o recuperar el sistema o, en su caso, prohibir o restringir la comercialización en el mercado nacional.

79. La definición de sistema de IA que presente riesgo debe entenderse no sólo por lo previsto por el art. 79.1 RIA, sino también por el art. 3.19 del RVM. Por tanto, a efectos de definir el riesgo ha de considerarse si el riesgo es razonable y aceptable en atención a la finalidad prevista, usos normales o usos previsibles.

80. Por tal se entiende cualquier medida destinada a impedir la distribución, exposición u oferta de un sistema de IA.

81. Esto es, la adopción de medidas para recuperar un sistema ya puesto a disposición de los usuarios.

En el marco del procedimiento han de respetarse los derechos y garantías recogidas por el art. 18 del Reglamento 2019/1020, por lo que tanto la valoración acerca del riesgo grave como las medidas, así como el plazo para adoptarlas por parte del operador, han de estar motivadas adecuadamente, comunicando además las vías de recurso correspondiente en atención al Derecho interno. Además, ha de garantizarse el derecho de audiencia, bien sea antes de la adopción de la decisión y en un plazo que no supere los 10 días, o bien con posterioridad, en caso de que el retraso en su adopción suponga la materialización del riesgo, y por tanto aconseje una intervención inmediata.

Si el incumplimiento excede el territorio de un Estado miembro, tras evaluar y valorar el riesgo, la autoridad de vigilancia del mercado ha de informar a la Comisión y al resto de socios comunitarios, acerca del resultado de ésta, así como de las medidas correctoras que el operador correspondiente debe adoptar. Dichas medidas, por otra parte, han de ser adecuadas no sólo para dar respuesta a nivel interno, sino alcanzar a todos los sistemas comercializados en la Unión. Nuevamente, en caso de que el operador no las adopte, se le faculta para adoptar medidas provisionales con el alcance comentado, informando sobre ello —nuevamente— a la Comisión y demás EM.

Dado el alcance de los efectos del incumplimiento, en estos supuestos debe garantizarse que el intercambio de información sea más granulado. De ahí que la propuesta se encargue de establecer los extremos mínimos que han de comunicarse (art. 79.6). Estos son, los datos necesarios para identificar el sistema no conforme; la trazabilidad y su origen; la naturaleza de la no conformidad y del riesgo, así como la naturaleza y duración de las medidas nacionales adoptadas; y, los argumentos formulados por el operador correspondiente. Además, la autoridad de vigilancia del mercado debe especificar si la no conformidad se debe al no respeto de la prohibición de prácticas prohibidas del art. 5, al incumplimiento de las obligaciones horizontales de los sistemas de alto riesgo, a la insuficiencia de las normas armonizadas o especificaciones comunes, y/o el incumplimiento de las obligaciones de transparencia de los proveedores y usuarios de sistemas de IA de uso general, que generen contenido sintético de audio, imagen, vídeo o texto.

Interesa destacar que el inicio de este procedimiento faculta a las autoridades de vigilancia del mercado de otros EM, a adoptar las medidas correctoras que estimen para su territorio, cuestión que han de comunicar también al resto de EM y a la propia Comisión. En línea con esto, todos los EM también están obligados a comunicar cualquier información adicional acerca de la no conformidad del sistema de IA concernido.

Recibida la información, la Comisión y las autoridades de vigilancia del mercado de los EM cuentan, con carácter general, con tres meses para presentar objeciones acerca de las medidas notificadas en casos generales, y treinta días en supuestos referidos a las prácticas prohibidas del art. 5. En caso no oposición, se entienden justificadas, estando todas las autoridades de vigilancia del mercado obligadas a garantizar la eficacia de las medidas restrictivas que afectan al sistema de IA, incluyendo su retirada. En caso de objeción, se inicia el procedimiento comunitario al que hacemos referencia seguidamente.

2. PROCEDIMIENTO DE SALVAGUARDIA DE LA UNIÓN

El art. 81 del Reglamento regula procedimiento de salvaguardia comunitario que, en línea con lo comentado, mantiene los grandes trazos recogidos por la normativa general sobre esta materia.

Así pues, esta fase del procedimiento se inicia si en los plazos señalados desde la notificación de las medidas adoptadas en el marco de un procedimiento de salvaguardia nacional, alguna autoridad de vigilancia del mercado formula objeciones acerca de las medidas adoptadas, o si la propia Comisión considera que estas medidas pueden resultar contrarias al Derecho de la Unión. En tales casos, se abre un proceso de consultas bajo el liderazgo de la Comisión, con la participación de las autoridades de vigilancia de los EM y los operadores afectados, a fin de valorar la procedencia y adecuación de las medidas adoptadas por el Estado miembro que inició el procedimiento en fase nacional.

Concluida esta fase, la Comisión decidirá y notificará sobre la justificación de la medida o medidas cuestionadas, en un plazo que no excederá de seis meses desde la notificación recogida por el art. 79.5, o sesenta días en caso de tratarse de un supuesto de prácticas prohibidas. Si considera justificada la medida, todos los EM deben garantizar su eficacia adoptando las medidas restrictivas procedentes, incluyendo la retirada del sistema, mediando notificación a la Comisión; caso contrario, la autoridad de vigilancia del mercado del Estado miembro que inició el procedimiento ha de proceder a retirarla, debiendo también notificar de ello a la Comisión.

Además de los supuestos anteriores, es posible que la administración comunitaria considere no justificada la medida nacional, no por razones de no conformidad, sino debido a deficiencias o inadecuado desarrollo de las normas armonizadas o, de ser el caso, de las especificaciones comunes. En este supuesto, se activaría el procedimiento contemplado a tales efectos por el art. 11 del Reglamento 1025/2012 sobre normalización europea. A tales fines debe notificarse a los órganos europeos de normalización, así como informar y consultar al Comité Permanente de Representantes de los EM. El comité procederá a decidir, una vez realizadas las consultas pertinentes con los organismos de normalización.

3. PROCEDIMIENTO RESPECTO DE SISTEMAS DE INTELIGENCIA ARTIFICIAL CONFORMES QUE PRESENTEN UN RIESGO

En sintonía con los procedimientos explicados, la normativa general de seguridad de los productos regula el procedimiento de sistemas conformes que presenten riesgos. En concreto por la señalada decisión 768/2008/CE. El reglamento, en su art. 82, no se separa sustancialmente de lo previsto por tal decisión, y en buena medida se replica el *iter* procedimental comentado respecto al procedimiento de salvaguardia comunitario. De ahí que únicamente mencionemos los aspectos diferenciales.

El procedimiento resulta de aplicación, igualmente, cuando una autoridad de vigilancia del mercado identifique un riesgo para la salud, la seguridad de las personas, o el incumplimiento de las obligaciones recogidas por el Derecho comunitario o nacional dirigidas a proteger los derechos fundamentales u otros bienes de imperioso interés público, informando de ello a la Comisión y los demás

Estados miembros⁸². No obstante, de la valoración realizada por la autoridad nacional se constata que el sistema es conforme con las normas técnicas armonizadas y las previsiones del Reglamento.

Ante este supuesto, debe consultar con la autoridad competente en materia de derechos fundamentales, en atención al art. 77.1. También ha de solicitar al operador que adopte las medidas correctoras de retirada o recuperación oportunas, fijadas por la autoridad nacional, y que han de alcanzar a todos los sistemas comercializados en la Unión. Posteriormente se notificará a la Comisión y a los EM, se realizará el proceso de consultas señalado, y será aquella la que decidirá sobre la justificación y adecuación de las medidas adoptadas, notificando a los EM y a los operadores económicos afectados.

4. PROCEDIMIENTO EN CASO DE INCUMPLIMIENTO FORMAL

Finalmente, el Reglamento recoge un en su art. 83 un procedimiento dirigido a dar respuesta a incumplimientos de algunas obligaciones formales: Marcado de conformidad que no se ajuste a las especificaciones recogidas por el art. 48; inexistencia del marcado de conformidad; no elaboración o elaboración incorrecta de la declaración UE de conformidad; no designación, en su caso, de representante autorizado; no disposición de documentación técnica; o no haber realizado el registro en la base de datos de la Unión, de conformidad con el art. 71.

En todos estos supuestos, la autoridad de vigilancia del mercado pedirá al proveedor que subsane el incumplimiento, y en caso de negativa o persistencia, deberá adoptar medidas adecuadas para restringir o prohibir la comercialización o, en su caso, su recuperación o retirada del mercado sin demora indebida.

5. PROCEDIMIENTO APLICABLE A LOS SISTEMAS CALIFICADOS POR LOS PROVEEDORES COMO NO DE ALTO RIESGO EN APLICACIÓN DEL ANEXO III

El Artículo 80 del RIA establece un novedoso procedimiento para evaluar la clasificación de sistemas de IA que los proveedores consideren no de alto riesgo. Como puede inferirse, el objetivo es asegurar que la clasificación que realicen los proveedores se ajuste adecuadamente a su nivel de riesgo, en atención a los criterios recogidos por el art. 6.3 y las correspondientes directrices desarrolladas por la Comisión.

En sintonía con los supuestos comentados, corresponde a la autoridad de vigilancia del mercado evaluar el sistema. En caso de constatar que se ha calificado erróneamente, exigirá al proveedor que cumpla los requisitos y obligaciones establecidas para los sistemas de alto riesgo, además que adopte las medidas correctoras pertinentes en un plazo razonable que fijará. También en esta misma línea, si el uso del sistema en cuestión excede el ámbito territorial de la AVM, debe

82. En dicha notificación deben proporcionar específicamente los datos requeridos para identificar el sistema de IA afectado, así como determinar su origen y cadena de suministro. También debe explicitarse la naturaleza del riesgo presentado y describir la naturaleza y duración de las medidas implementadas.

notificar a la Comisión Europea y a los Estados miembros, informando tanto de la evaluación como de las medidas adoptadas.

Una vez notificado, el proveedor debe adoptar medidas para cumplir con los requisitos y obligaciones procedentes en atención al sistema de alto riesgo correspondiente. También debe asegurarse de dar cumplimiento a las medidas correctoras que, de ser el caso, deben alcanzar a todos los sistemas comercializados en territorio de la Unión.

En caso de incumplir con tales requerimientos se abrirán procedimientos sancionadores que pueden concluir con la imposición de sanciones previstas por el art. 99. Misma medida disuasoria se recoge para el supuesto en que se determine que, la calificación del sistema como no de alto riesgo por parte del proveedor, pretendía eludir los requisitos establecidos por los art. 8 a 15 del Reglamento.

El régimen sancionador en el Reglamento de inteligencia artificial

F. JAVIER SEMPERE

Director de Supervisión y Protección de Datos del Consejo General del Poder Judicial. Doctorando por CEU Escuela Internacional de Doctorado (CEINDO)

I. INTRODUCCIÓN

El RIA únicamente dedica dos preceptos para regular su régimen sancionador, a los que se debe añadir los correspondientes considerandos que explican su contenido. Éstos son los artículos 71 y 72, denominados respectivamente «Sanciones» y «Multas administrativas» a las instituciones, agencias y organismos de la Unión. Se trata de dos preceptos cuya finalidad resulta diferente, ya que el primero es el que realmente regula el régimen sancionador, mientras el segundo se dedica a atribuir la potestad sancionadora en el marco de las instituciones comunitarias al Supervisor Europeo de Protección de Datos (SEPD).

Ambos, pero, sobre todo, el primero, en lo referente a la técnica normativa utilizada, guarda estrecha relación con los artículos que, a su vez, regulan el régimen sancionador del Reglamento Europeo de Protección de Datos (RGPD)¹, adoleciendo también de las mismas carencias, como la deficiente tipificación de las infracciones, por lo que en ocasiones nos referiremos al mismo.

Además, hay que tener en cuenta, la vinculación de los sistemas de inteligencia artificial con la protección de datos personales, lo que supondrá, como expondremos la necesaria comunicación entre la Autoridad de Control de IA y la Autoridad de

1. DOUE de 4 de mayo de 2016.

<https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex%3A32016R0679>

Con posterioridad, se publicaron sendas correcciones de errores, el 23 de mayo de 2018 y el 4 de marzo de 2021. En algunos casos, se trata más que una corrección de errores, una de carácter material, cambiando determinados aspectos del contenido de la norma. Téngase en cuenta, que la primera se realiza dos años después de la primera publicación de la norma, y por tanto, aunque todavía no era aplicable, había sido objeto de análisis detallado. Y la segunda, una vez que la norma se aplica.

Pueden consultarse en:

[https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32016R0679R\(02\)&from=ES](https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32016R0679R(02)&from=ES)

[https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32016R0679R\(03\)&from=ES](https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32016R0679R(03)&from=ES)

Control de Protección de Datos cuando ambas sean diferentes, puesto que algunos países han determinado que la Autoridad en materia de inteligencia artificial sea la de protección de datos personales, sin necesidad de crear una entidad nueva.

Junto a estos dos preceptos, también deben considerarse algunos otros que aparecen en el resto del texto del RIA, como pueden ser el artículo 58 relativo a las «Tareas del Consejo», el 59 sobre «Designación de autoridades nacionales competentes», y el 65 «Procedimiento para tratar los sistemas de IA que presentan un riesgo a nivel nacional». Nos referiremos también a ellos con más detalle por su relación con el contenido del citado artículo 71.

Respecto a los considerandos, si bien únicamente el 84 es el que se dedica a la parte sancionadora, para completarlo se puede acudir también al 79, referente a las funciones de las denominadas autoridades de vigilancia del mercado.

En consecuencia, observamos que el régimen sancionador resulta, a primera vista, bastante escueto con únicamente dos preceptos, uno de ellos dedicado al SEPD, y un considerando explicativo, pudiendo añadir un segundo. Probablemente, más allá de como ya hemos adelantado, algunas deficiencias que señalaremos, la razón venga dada porque cada uno de los Estados miembros, debe en parte, desarrollar este régimen sancionador.

Asimismo, y sin perjuicio tanto del contenido de estos preceptos, así como la norma que en un futuro de se adopte en aras de realizar ese desarrollo, a la hora de ejercitar la potestad sancionadora por la autoridad de vigilancia de mercado, que, en nuestro caso, se llevará a cabo por la Agencia Española de Supervisión de Inteligencia Artificial (AESIA), se debe dar cumplimiento a los principios regulados en la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público (LRJSP)², como son legalidad (art.25), irretroactividad (art.26), tipicidad (art.27), responsabilidad (art.28), proporcionalidad (art.29), prescripción (art.30), y concurrencia de sanciones (art.31). Se trata de principios que tienen su origen en el Derecho Penal pero que el Tribunal Constitucional en su sentencia 18/1981, de 8 de julio³, señaló que su finalidad principal consiste en dotar de garantías al procedimiento.

II. EL ARTÍCULO 71 DEL REGLAMENTO: NECESIDAD DE DESARROLLO LEGISLATIVO E INTERACCIÓN CON OTRAS NORMATIVAS Y FIGURAS ALTERNATIVAS A LA MULTA Y ESPECIFICACIONES PARA LAS ADMINISTRACIONES

Comenzamos con este precepto, al que dedicamos un análisis más pormenorizado y detallado pues de su contenido se desprende que regula el régimen sancionador aplicable en esta materia.

Partimos del texto inicial de la propuesta de RIA de la Comisión Europea, pudiendo diferenciar varios apartados: la remisión a los Estados miembros para que desarrollen el régimen sancionador, incluyendo también que determinen o no

2. BOE n.º 236 de 2 de noviembre de 2015. <https://www.boe.es/buscar/act.php?id=BOE-A-2015-10566>

3. Rebollo Puig, M., Izquierdo Carrasco, M., Alarcón Sotomayor, L., y Bueno Armijo, A. M^a., *Derecho Administrativo sancionador*, Editorial Lex Nova, Primera Edición, Valladolid, 2010.

la posibilidad de imponer multas al sector público; la tipificación de infracciones así como la cuantía de las multas; los criterios para determinar la cuantía de la multa a imponer cuando se comete una infracción; y la habilitación para imponerlas no sólo por órganos administrativos sino también jurisdiccionales, dependiendo del ordenamiento jurídico aplicable de cada país. Procedamos a analizar con detalle cada uno de estos apartados.

En cuanto al primero, supone que el régimen sancionador no se agota con lo previsto en el artículo 71, sino que los Estados miembros «*determinarán el régimen de sanciones, incluidas las multas administrativas, aplicable a las infracciones del presente Reglamento y adoptarán las medidas necesarias para garantizar su aplicación adecuada y efectiva*». Es decir, será necesario un desarrollo legislativo por cada uno de los Estados, que complete lo recogido en el citado precepto.

Así, se observa que respecto a cómo se debe responder ante la comisión de una infracción por parte de la autoridad de vigilancia, únicamente se ha contemplado la posibilidad de imponer multas, sin que aparezca otro tipo de alternativas como pueden ser el apercibimiento, la advertencia, o la orden de cumplimiento, figuras que en cambio, sí contempla el RGPD⁴. Recordemos que el apercibimiento, según la modificación de la Ley Orgánica 3/2018, de 3 de diciembre, de protección de datos de carácter personal y garantía de derechos digitales (LOPDGDD), carece de naturaleza sancionadora⁵; la advertencia se puede aplicar cuando es posible la comisión de una infracción pero sin que se tramite el correspondiente procedimiento sancionador; y la orden de cumplimiento, aunque la Agencia Española de Protección de Datos (AEPD) lo aplica en las resoluciones sancionadoras, como pueden ser las que versan sobre videovigilancia, en las que se ordena, por ejemplo, que la cámara no grabe la vía pública o que se dé cumplimiento al derecho de información con el correspondiente cartel, otras Autoridades utilizan esta opción sin necesidad de tramitar un procedimiento sancionador.

A este respecto, debemos partir que no necesariamente, ante la posible comisión de una infracción se tiene que imponer una multa, sino que, dependiendo del caso concreto, se puede resolver el perjuicio causado sin necesidad de acudir al binomio apertura de procedimiento sancionador para imponer una multa. De hecho, la propia propuesta de RIA, aunque como apuntamos sólo ha recogido en el artículo 71 la multa, de otro de sus preceptos, el artículo 65 se desprende esta posible forma de actuación.

Así, según el mismo, se establece un procedimiento específico cuando se trate de sistemas de IA que presenten un riesgo a nivel nacional respecto a la salud, seguridad o protección de los derechos fundamentales, de manera que la autoridad de vigilancia, cuando tenga conocimiento, no necesariamente tiene que dictar el

4. Véase al respecto el apartado 2 del artículo 58 del RGPD.

5. Véase al respecto Ley 11/2023, de 8 de mayo, de transposición de Directivas de la Unión Europea en materia de accesibilidad de determinados productos y servicios, migración de personas altamente cualificadas, tributaria y digitalización de actuaciones notariales y registrales; y por la que se modifica la Ley 12/2011, de 27 de mayo, sobre responsabilidad civil por daños nucleares o producidos por materiales radioactivos; modifica determinados artículos de la LOPDGDD. BOE n.º 110 de 9 de mayo de 2023.

acuerdo de inicio del procedimiento sancionador, sino que efectuará una evaluación del sistema para verificar que cumple todos los requisitos. Y si nos los cumple, exigirá medidas correctoras de inmediato, o bien retirarlo del mercado, pudiendo también adoptar medidas provisionales para prohibir o restringir la comercialización del sistema.

Por cuanto a la interacción con otras normativas y figuras alternativas a la multa, este artículo 65, desglosa también medidas que pueden calificarse igualmente de sancionadoras, como son la retirada, prohibición o restricción de un determinado producto de IA. Pensemos, además, que no necesariamente la multa puede ser la que cause un mayor perjuicio, sino las citadas. La retirada, dejaría durante un tiempo sin la posibilidad de ingresos al no estar el producto de IA disponible en el mercado; la restricción que los ingresos fuesen menores; y la mayor sanción sería prohibir el producto en cuestión.

En consecuencia, ese desarrollo podría contemplar todas estas medidas que puede adoptar la autoridad de vigilancia de mercado, a imagen y semejanza de lo que contempla el RGPD cuando regula los poderes de las autoridades de control. Recapitulando, podrían ser el apercibimiento, advertencia, medidas de cumplimiento, así como la retirada, prohibición y restricción de un producto de IA.

Respecto de las especificaciones para las Administraciones Pública, esta futura norma debe recoger si las Administraciones públicas pueden ser objeto de multa en caso de cometerse una o varias infracciones, puesto que el proyecto de RIA deja que cada uno de los estados miembros de la UE, decida al respecto. Esto supone, que probablemente, no exista una uniformidad ya que unos países podrían contemplar esta posibilidad, y otros no, como precisamente ha ocurrido con el RGPD, que contiene una previsión similar, y en la que la regla general ha sido que se pueden multar, salvo tres países, siendo los que no, Francia, Luxemburgo y España. Por ello, lo más probable, es que nuestro legislador adopte la misma previsión que la vigente en la actualidad en la LOPDGDD, recientemente modificada, para sustituir la posibilidad de apercibir a las Administraciones pública por el sistema de la antigua Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal, consistente en señalar que se ha cometido una o varias infracciones, la posibilidad de instar a cumplir medidas, y en su caso, aunque en la práctica es un «rara avis», proponer que se inicie un procedimiento disciplinario contra el presunto responsable⁶.

En este sentido, no compartimos que no se recoja esta posibilidad, puesto que, por una parte, supone un agravio comparativo respecto al sector privado, y por otra, como se ha demostrado en el ámbito de la protección de datos supone una «relajación» en el cumplimiento en el sector público. Habría que añadir que en nuestro ordenamiento jurídico existen otras normas, que, en cambio, sí prevén esa posibilidad, como pueden ser algunas normas en materia de defensa de la competencia⁷, o más recientemente

6. Véase al respecto el artículo 77 «Régimen aplicable a determinadas categorías de responsables y encargados de tratamiento» de la LOPDGDD.

7. Ortega Fernando, J. *Comentario a la STS de 18 de julio de 2016*, en el blog jurídico almacenderecho.org, 2017.

<https://almacenderecho.org/cuando-se-puede-sancionar-la-administracion-ambito-la-defensa-la-competencia>

en la Ley 2/2023, de 20 de febrero, reguladora de la protección de las personas que informen sobre infracciones normativas y lucha contra la corrupción⁸.

En todo caso, si la finalidad de un Reglamento comunitario consiste en que todos los afectados por su contenido cumplan las mismas reglas, con este tipo de habilitaciones en favor de los Estados miembros se pierde la uniformidad buscada, preguntándonos ¿Por qué un Ayuntamiento de Italia puede ser multado si incumple el RIA y un Ayuntamiento de cualquier Comunidad Autónoma, en la misma situación, a lo máximo que llegará es a ser apercibido?

III. PRESCRIPCIÓN Y TIPIFICACIÓN DE INFRACCIONES Y CUANTÍA DE LAS MULTAS EN EL ARTÍCULO 71

Por otra parte, la norma española sí puede llegar también a cubrir otras cuestiones, de tipo procedimental, pero, sobre todo, lo referente a regular la figura de la prescripción de las infracciones y de las sanciones. Recordemos que la prescripción es un elemento fundamental en nuestro ordenamiento jurídico y dota de seguridad jurídica, sobre todo al presunto infractor a los efectos de que, una vez cometido unos hechos, comienza a computarse el plazo de prescripción, de forma que, si transcurre el mismo, no podrá ser sancionado. La LRJSP en su artículo 30, regula la prescripción tanto de infracciones como de sanciones, partiendo de la regla general de que ambas prescriben según lo que dispongan las leyes que las establecen, y en su defecto, en el caso de las infracciones las muy graves a los tres años, la graves a los dos y las leves a los seis, y en el caso de las sanciones, el plazo es el mismo salvo las impuestas por faltas leves que será de un año.

En el caso que nos ocupa, obviamente la propuesta del RIA no contempla la prescripción ni de infracciones ni de sanciones, pero, sobre todo, y como expondremos más adelante tampoco califica las infracciones entre leves, graves ni muy graves (las sanciones tampoco) puesto que utiliza un sistema de tipificación en «términos amplios», de idéntica manera que ha hecho el RGPD, que podemos calificar también como exhaustivo⁹.

Por tanto, tendrá que ser la norma española la que cubra la no existencia de la figura de la prescripción, pero muy vinculada a la tipificación de infracciones, y de manera similar a lo hecho en la LOPDGDD¹⁰, lo que nos lleva a analizar el segundo apartado destacable de este artículo 71, consistente en la citada tipificación, así

8. BOE n.º44 de 21 de febrero de 2023.

<https://www.boe.es/buscar/act.php?id=BOE-A-2023-4513>

Esta norma no ha contemplado un régimen específico para las Administraciones públicas, puesto que según su art.62.1 «Estarán sujetos al régimen sancionador establecido en esta ley las personas físicas y jurídicas que realicen cualquiera de las actuaciones descritas como infracciones en el artículo 63», con la diferencia que, a nivel de cuantía, el art.65.1 en su apartado a) fija como máxima a las personas físicas 30.000 € y en el apartado b) para las personas jurídicas 1.000.000 €.

9. Hernández Corchete, J.A., *Exhaustividad y estándares del principio de legalidad sancionadora*, en Derecho Digital e Innovación, n.º 2, Editorial Wolters Kluwer, abril-junio de 2019.

10. Véase al respecto sus artículos 72, 73 y 74 que regula la prescripción para las infracciones muy graves, graves y leves.

como las cuantías de las multas a imponer. Concretamente, los puntos 3, 4 y 5 del mencionado precepto, que pasamos analizar.

En cuanto al punto 3, determina que cualquier vulneración tanto de la prohibición de las prácticas de inteligencia artificial del artículo 5 como de los requisitos del artículo 10 pueden ser multadas con hasta 30 millones de euros o si se tratase de una empresa hasta el 6% del volumen de negocio anual del ejercicio financiero anterior si esta cuantía fuese superior.

Sobre el artículo 5, como se desprende de su denominación «Prácticas de inteligencia artificial prohibidas» contiene todo un listado de prácticas que no pueden tener lugar debido a esta prohibición. Dicho listado aparece en el apartado 1 de este precepto, consistentes en la introducción al mercado, puesta en servicio o utilización de un sistema de IA *«que se sirva de técnicas subliminales que trasciendan la conciencia de una persona para alterar de manera sustancial su comportamiento de un modo que provoque o sea probable que provoque perjuicios físicos o psicológicos a esa persona o a otra»*, que tenga como finalidad *«evaluar o clasificar la fiabilidad de las personas físicas durante un período de tiempo atendiendo a su conducta social o a características personales o de su personalidad conocidas o predichas»* pudiendo provocar esta clasificación un trato perjudicial; o el uso de sistemas de identificación biométrica remota en tiempo real en accesos de espacios públicos cuando no sea aplicable alguna de las excepciones previstas.

Sin embargo, en alguno del resto de apartados del mencionado artículo 5, también se puede desprender la existencia de conductas que en caso de incumplimiento podrían ser susceptibles de sanción. Así, en el apartado 2 se indica que en el caso de un sistema de identificación biométrica en tiempo real para un espacio de acceso público con fines contemplados en la ley deben cumplir con salvaguardas y condiciones necesarias sobre su uso y particularmente respecto a limitaciones temporales, geográficas y personales; y en el apartado 3 obliga a que antes de su funcionamiento, resulta necesario obtener la previa autorización por la autoridad judicial o administrativa. En consecuencia, se podrían cometer infracciones relativas a vulnerar su límite temporal, geográfico o personal, o ponerlo en marcha sin haber obtenido la citada autorización.

Por tanto, aunque la norma se refiere únicamente al listado de prohibiciones, en los dos apartados citados anteriormente también aparecen otras conductas que podrían ser susceptibles de infracción en caso de incumplimiento.

Por lo que se refiere al artículo 10, denominado «Datos y gobernanza de datos», recoge los criterios de calidad del conjunto de datos que se utilicen para el entrenamiento, validez y prueba de los sistemas de IA de alto riesgo. Estos criterios de calidad se desglosan en los apartados 2 a 5 de este precepto, debiendo diferenciarse en función del contenido de cada uno de ellos.

Así, el apartado 2 se refiere a las prácticas adecuadas de gobernanza y datos, entre ellas, elegir un diseño adecuado, la recopilación de datos, o el examen atendiendo a los sesgos; el apartado 3, exige que los datos de entrenamiento, validación y prueba sean pertinentes, representativos y carezcan de errores, con propiedades estadísticas adecuadas, características que pueden ser para cada dato o combinación de éstos; el apartado 4, ordena que los datos de entrenamiento, validación y prueba, tengan en cuenta las características o elementos particulares del contexto geográfico, conductual o funcional específico; y finalmente, el apartado 5, habilita a los proveedores de estos

sistemas de alto riesgo, siempre que sea estrictamente necesario para garantizar la vigilancia, detección y corrección de los sesgos asociados, la posibilidad de tratar categorías especiales de datos del art.9.1. del RGPD así como del art.10.1 del Reglamento 2018/1725¹¹, siempre que se ofrezcan salvaguardas adecuadas para los derechos y libertades fundamentales de las personas físicas, incluyendo establecer limitaciones técnicas a la reutilización y utilización de medidas de seguridad y protección de la privacidad recientes, como seudoanonimización o cifrado, cuando la anonimización pueda afectar significativamente al objetivo perseguido.

Atendiendo al contenido de este artículo 10 debemos concluir que no necesariamente debe existir un incumplimiento de todos los requisitos para poder constatar una infracción, sino que la falta de uno de ellos ya sería sancionable. Por ello, podría darse, a modo de ejemplo, que respecto al apartado 2 se hubiese realizado un diseño adecuado y recopilado los datos, pero se hubiese obviado el examen atendiendo a posibles sesgos; sobre el apartado 3, los datos de entrenamiento, validación y prueba carecieran de errores pero no fuesen pertinentes; sobre el apartado 4, se hubiese tenido en cuenta el contexto geográfico y conductual pero no el específico; y sobre el apartado 5, se hubiesen adoptado unas limitaciones técnicas a la reutilización estableciendo medidas de seguridad pero no todas las que hubiesen sido necesarias. Como observamos de estos ejemplos, si se cumplen unos requisitos, pero otros no, su incumplimiento podría devenir en la comisión de la respectiva infracción aparejando la sanción, que sería la multa.

Asimismo, respecto al último apartado deberá realizarse una doble labor interpretativa para su aplicación, puesto que, por una parte, esa habilitación se limita a que sea «estrictamente necesario», por lo que podría darse a que no lo fuese y se hubiesen utilizado ese tipo de datos; y por otra, al tratarse dicha habilitación sobre protección de datos personales, la competente probablemente sea la Autoridad de Control de esa materia, y no la Autoridad de Control de la IA. Esta interacción con la normativa de protección de datos, no sólo en este precepto sino también en otros, debe quedar claramente definida en el texto de manera que no provoque conflictos futuros¹².

En suma, de estos dos preceptos se deducen múltiples incumplimientos, por lo que habrá que ir a cada uno de ellos para realizar una labor de disgregación a los efectos de poder determinar las conductas que fuesen sancionables. No obstante, los mismos, como hemos descrito, se refieren a un listado de actividades de alto riesgo (artículo 5) y a requisitos (artículo 10), por lo que dentro de lo que cabe, esta actuación se puede desarrollar sin inconvenientes, aunque hubiese sido más garantista que hubiese aparecido en la norma un listado de las conductas susceptibles de ser multadas, o en su caso, sancionadas con otra figura.

11. Reglamento (UE) 2018/1725 del Parlamento Europeo y del Consejo, de 23 de octubre de 2018, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por las instituciones, órganos y organismos de la Unión, y a la libre circulación de esos datos, y por el que se derogan el Reglamento (CE) n.º 45/2001 y la Decisión n.º 1247/2002/CE. DOUE de 21 de noviembre de 2018.
12. CEPD-SEPD: *Dictamen conjunto 5/2021 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia artificial)*. 2021. https://edpb.europa.eu/system/files/2021-10/edpb-edps_joint_opinion_ai_regulation_es.pdf

No obstante, dicho listado debería existir, sobre todo si tenemos en consideración la total ausencia de seguridad jurídica que provoca el siguiente apartado, el cuarto, del artículo 71, que preceptúa que cualquier «*incumplimiento por parte del sistema de IA de cualquiera de los requisitos u obligaciones establecidos en el presente Reglamento distinto de los artículos 5 y 10 están sujetos hasta multas administrativas de hasta 20 millones de euros, o si el infractor es una empresa, hasta el 4% del volumen de negocio total anual mundial del ejercicio financiero anterior, si esa cuantía fuese superior*». Es decir, esto supone tener que revisar cada uno de los artículos de la norma y tratar de dilucidar donde puede existir un incumplimiento y, por tanto, susceptible de imponer una multa, y donde no por ser el contenido del correspondiente artículo y sus apartados meramente declarativo.

Así, a modo de ejemplo, del artículo 11 referente a la «Documentación técnica», que contiene tres apartados, de los dos primeros se desprenden obligaciones, cuestión que no ocurre con el tercero. De esta forma, según el primero, se debe preparar una documentación técnica antes de la introducción al mercado del sistema de IA así como actualizarla; según el segundo, cuando se introduzca en el mercado o se ponga en servicio un sistema de IA de alto en el anexo II, sección A, se elaborará una única documentación técnica que contenga toda la información estipulada en el anexo IV, así como la información que exijan dichos actos legislativos; y en cambio, según el apartado tercero, no se contiene ninguna obligación, puesto que su contenido supone una reserva en favor de la Comisión para adoptar actos delegados.

Esta labor, como exponemos, debe realizarse de todos los preceptos de la norma, si bien, donde puede haber una mayor incidencia es en el Título III sobre los Sistemas de Alto Riesgo, que incluye su Capítulo 3 «Obligaciones de los proveedores y usuarios de sistemas de IA de alto riesgo y de otras partes».

En cuanto al apartado 5 de este artículo 71 contempla el tercer bloque de incumplimientos referente a la «presentación de información inexacta, incompleta o engañosa a organismos notificados y a las autoridades nacionales competentes en respuesta a una solicitud» podrá ser multada hasta 10 millones de euros o, en el caso de que el infractor fuese una empresa, el 2% del volumen de negocio anual mundial del ejercicio financiero anterior, si esta cuantía fuese superior.

De nuevo se debe realizar una labor «titánica» revisando aquellos artículos de la norma europea donde exista una obligación de facilitar información a los citados organismos notificados y autoridades nacionales competentes, que, si no se realizase correctamente, sería susceptible de ser multada. Así, por ejemplo, el ya citado artículo 11 sobre la documentación técnica, señala que se proporcionará a las autoridades nacionales competentes y los organismos notificados toda la información necesaria para poder evaluar si el sistema de IA cumple los correspondientes requisitos; o el artículo 22 titulado «Obligación de información», que como su nombre indica, configura una obligación sobre el proveedor de manera que si éste es consciente de que el sistema de IA de alto riesgo presenta un riesgo en el sentido del artículo 65 («Procedimiento aplicable a los sistemas de IA que presenten un riesgo a nivel nacional»), debe informar de inmediato sobre las el mismo y las medidas correctoras adoptadas a las autoridades nacionales.

Asimismo, y además de las obligaciones de información que recoge la norma, también habría que considerar los propios requerimientos de información en el marco de una investigación, que suelen tener lugar antes de dictar el acuerdo de

inicio, en la fase de actuaciones previas, y cuya finalidad consiste en conocer si existen indicios suficientes para considerar la posible comisión de una o varias infracciones, y, en consecuencia, dictar el acuerdo de inicio. Pues bien, esa falta de contestación a estos requerimientos, o una contestación que no llegue a todo lo requerido, podría igualmente ser sancionable.

Por otra parte, respecto a las cuantías de los tres apartados del artículo 71 que hemos descrito, si tenemos en cuenta que las mismas van de 30 millones o 6% (apartado 3 sobre los incumplimientos de los artículos 5 y 11); 20 millones o 4% (apartado 4 para el resto de incumplimientos); y 2% (apartado 5 para los derivados de remitir información inexacta, incompleta o engañosa), a los efectos de que la futura norma española regule la figura de la prescripción, atendiendo a estas cuantías, tendría hecha la clasificación entre infracciones muy graves, graves y leves. Además, de la misma forma que en su día contempló el RGPD, esa falta de tipificación de infracciones se podría «arreglar» por la vía de la prescripción, describiendo las conductas a tal efecto, pero que pueden ayudar para conocer cuáles son susceptibles de sanción.

IV. CRITERIOS PARA DETERMINAR LA CUANTÍA DE LA MULTA EN EL ARTÍCULO 71

Siguiendo con las cuantías, para determinar la multa a imponer, se deben utilizar unos factores o criterios al respecto que pueden actuar como agravantes o atenuantes y que aparecen en el apartado 6 de este artículo 71, y que suponen aplicar el principio de proporcionalidad, que será vulnerado si no se concretan las circunstancias que motivan la imposición de una multa¹³.

Tengamos presente que nuestro ordenamiento jurídico, en la LRJSP también lo recoge en su artículo 29, siendo éstos, el grado de culpabilidad o la existencia de intencionalidad; la continuidad o persistencia en la conducta infractora; la naturaleza de los perjuicios causados; y la reincidencia, por comisión en el término de un año de más de una infracción de la misma naturaleza cuando así haya sido declarado por resolución firme en vía administrativa.

Por su parte, el proyecto de RIA, únicamente contempla como criterios la naturaleza, gravedad y la duración de la infracción y de su consecuencia; si otras autoridades de vigilancia del mercado han impuesto ya multas administrativas al mismo operador por la misma infracción; y el tamaño y la cuota de mercado del operador que comete la infracción. Procedemos a explicar cada uno de ellos.

En cuanto a la naturaleza, gravedad y duración de la infracción y sus consecuencias, supone valorar qué tipo de infracción se trata en función de las tres categorías existentes en relación con los apartados 3, 4 y 5, puesto que debido a que las cuantías de las multas son diferentes, y a su vez, ligado a la naturaleza de cada una de ellas; de esta forma, si la infracción es del apartado 3, la cuantía tendrá a ser superior que si fuese del apartado 5; también supone tener en cuenta el perjuicio que se haya causado a los posibles afectados; también el número de afectados y la

13. Hernández Jiménez, H. M. *Aplicación práctica de los principios de la potestad sancionadora de la Administración en la nueva Ley 40/2015*, en *Actualidad Administrativa*, n.º 2, 2017.

duración, de forma que cuántos más haya y más se prolongue en el tiempo, la cuantía deberá ser superior.

Respecto a si otras autoridades de vigilancia del mercado han impuesto ya multas administrativas al mismo operador por la misma infracción, debemos considerar que se haya interpuesto por otra autoridad de otro país, lo que conlleva necesariamente a que exista una comunicación entre las diferentes autoridades. A este respecto, debemos realizar dos precisiones: por una parte, este criterio no supone que no vayan a ser sancionadas porque ya lo hayan sido, puesto que si fuese así, se habría contemplado de forma expresa; y por otra, podría interpretarse como un elemento agravante, y en cierta medida relacionado con la reincidencia, puesto que si ese operador, que actúa en varios países de la Unión, ya ha sido sancionado en uno de ellos por un producto de IA, debería haber corregido los hechos que han supuesto dicha sanción, de manera que no provoque más perjuicios.

Sobre el tamaño y cuota del mercado del operador, se desprende que se valorará si se trata de una pequeña o mediana empresa o de una mayor, y qué en cierta medida, aparece en las cuantías de las multas cuando se refiere a los tantos porcentuales del volumen de negocio de las empresas, en consonancia con el apartado 1 de este artículo 71 que se refiere a considerar en las sanciones a imponer *«los intereses de los proveedores a pequeña escala y las empresas emergentes, así como su viabilidad económica»*. Obviamente, cuanto mayor sea el tamaño y la cuota de mercado, más alta será la cuantía.

Cabe mencionar también que si bien estos son los únicos criterios contemplados en el proyecto de IA, a través de esa norma española que establezca otras medidas sancionadoras, así como la prescripción, podría incluir también otros a valorar, como podrían ser la culpabilidad (si existe dolo o negligencia), el posible beneficio obtenido, la reincidencia, las medidas adoptadas para paliar el perjuicio causado, la cooperación con la autoridad de vigilancia, o si en caso de causas daños o perjuicios a personas, si fuesen menores de edad.

Y todo ello sin perjuicio de la función atribuida al Comité Europeo de Inteligencia Artificial en el artículo 58 de adoptar documentos orientativos incluyendo directrices para fijar las multas administrativas.

Por último, por lo que se refiere al último apartado del artículo 71, el 8, dado que algunos países no imponen multas sus órganos administrativos, deja abierta la posibilidad a que se impongan por los juzgados y tribunales. Se trata de una previsión, al igual que otras, recogidas igualmente en el RGPD.

Con posterioridad, en el texto adoptado en el Consejo Europeo de 6 de diciembre de 2022, posición común sobre el proyecto de RIA, se propone introducir numerosas modificaciones, las cuales podemos englobar en dos grupos diferentes como serían una mayor consideración respecto a las pequeñas y medianas empresas, y dotar al régimen sancionador de una mayor seguridad jurídica, incluyendo en este caso algunas previsiones que anteriormente habíamos señalado como necesarias.

Respecto al primer grupo, en los apartados 3, 4 y 5 del artículo 71 se contempla que, en el caso de pequeñas y medianas empresas, el límite máximo de la cuantía a imponer en caso de cometer alguna infracción, sería inferior. Así, en el apartado 3 frente 6% del volumen total anual del ejercicio financiero inferior, si afectase a las pymes, especialmente empresas emergentes, el límite se fija en un 3%; en el apartado

2, de un 4% pasa a un 2%; y en el apartado 5, de un 2% a un 1%. No obstante, este límite en las bajadas, a efectos prácticos, podría no resultar necesario y haber seguido aplicando los límites anteriores con los criterios de graduación, el referente al tamaño de la empresa y cuota de mercado, así como lo dispuesto en el apartado 1, el cual contiene un mandato para igualmente considerar a las pymes.

Mayor relevancia contienen las modificaciones del segundo grupo que hemos calificado en base a la finalidad de dotar de una mayor seguridad jurídica, sobre todo las tendentes a mejorar el régimen de tipificación de las infracciones en el apartado 4 del artículo 71. Recordemos que en su redacción inicial recogía que cualquier infracción de las no previstas en el artículo 3, que a su vez establecía las infracciones sobre los artículos 5 y 10, sería multada hasta 20 millones de euros o 4% del volumen de negocios, lo cual suponía revisar de arriba abajo la norma buscando posibles incumplimientos en cualquiera de sus preceptos.

Pues bien, con la nueva redacción, en aras de dotar de una mayor seguridad jurídica, aparece todo un listado de preceptos susceptibles de posibles incumplimientos, como son vulnerar las obligaciones de los proveedores conforme a los artículos 4 ter («requisitos de los sistemas de IA de uso general y obligaciones de los proveedores de dichos sistemas») y 4 quater («excepciones al artículo 4 ter»); las obligaciones de los proveedores con arreglo al artículo 16 («obligaciones de los proveedores de sistemas de IA de alto riesgo»); las obligaciones de otras personas con arreglo al artículo 23 bis («obligaciones para que otras personas estén sujetas a las obligaciones de un proveedor»); las obligaciones de los representantes autorizados con arreglo al artículo 25 («representantes autorizados»); las obligaciones de los importados con arreglo al artículo 26 («obligaciones de los importadores»); las obligaciones de los distribuidores con arreglo al artículo 27 («obligaciones de los distribuidores»); las obligaciones de los usuarios con arreglo al artículo 29, apartados 1 a 6 bis («obligaciones de los usuarios de los sistemas de IA de alto riesgo»); requisitos y obligaciones de los organismos notificados con arreglo al artículo 33 «requisitos relativos a los organismos notificados», artículo 34 apartados 1, 3 y 4 («filiales de organismos notificados y subcontratación por parte de estos»), y artículo 34 bis («obligaciones operativas de los organismos notificados»); y obligaciones de transparencia para los proveedores y usuarios con arreglo al artículo 52 («obligaciones de transparencia de los proveedores y usuarios de determinados sistemas de IA»). De esta forma, se precisa, al menos, que artículos del texto son susceptibles de que en el caso que se vulnere su contenido, se pueda ser sancionado. No obstante, a pesar de esta precisión, habrá que analizar cada precepto para determinar cuando existe una obligación susceptible de incumplimiento y cuando no, al no imponerse nada y ser una redacción meramente enunciativa. Señalar que también que desaparecen las posibles infracciones del artículo 10 que se contemplaban en el apartado 3 del artículo 71, quedando ahora únicamente en las relativas al artículo 5.

Asimismo, y aunque sigue faltando un apartado en el que se indique claramente quienes pueden ser los posibles sujetos responsables de una infracción, del listado de artículos referidos se desprende que serían proveedores, representantes de proveedores establecidos fuera de la Unión Europea, usuarios y organismos notificados.

Siguiendo con este segundo grupo, se dota también de mayor seguridad jurídica al introducirse en el apartado 6 nuevos criterios para cuantificar la cuantía de las multas a imponer, que como expusimos anteriormente, en el proyecto de IA eran bastante escuetos. De hecho, aparecen algunos de los que hicimos hincapié en que precisamente, deberían estar también, como son la intencionalidad o negligencia en la infracción; y cualquier medida adoptada por el operador para poner remedio a la infracción y mitigar los posibles efectos adversos de la misma.

También se incluye una tercera recogiendo la posibilidad de valorar si ese operador ha sido multado por otras autoridades por infracciones legislación nacional o de la Unión, cuando esas infracciones se deriven de la misma actividad u omisión que constituya una infracción pertinente del RIA. Ciertamente, este criterio se redacta de forma farragosa, puesto que da a entender que un determinado hecho que constituya una infracción de la norma, lo cual supone analizar el contenido del texto del RIA buscando donde pueden existir infracciones, que, a su vez, pudiesen serlo de otra materia, regulada por normativa nacional de los Estados miembros o de la Unión. Desde nuestro punto de vista, esta confluencia podría darse en el ámbito de la protección de datos personales, cuando se vulnera alguno de los preceptos del RIA donde se recoge expresamente circunstancias u obligaciones sobre el tratamiento de datos personales.

Por otra parte, y además de los cambios ya señalados, se añaden otros dos más, consistentes en la inclusión en el apartado primero del artículo 71 la previsión del uso del sistema de la IA sobre el contexto de una actividad personal no profesional, y un nuevo apartado 9 garantizando que la autoridad de vigilancia de mercado actúe conforme a las garantías procesales del Derecho de la Unión y los Estados miembros, entre las que se encuentra la tutela judicial efectiva.

Sobre esa actividad personal no profesional, de hecho, el artículo 2.8 exceptúa la aplicación del RIA cuando se de esta situación salvo lo previsto en el artículo 52 que contiene obligaciones de transparencia que recaen sobre los usuarios tanto de un sistema de categorización biométrica como de un sistema de reconocimiento de emociones así como cuando se trate de un sistema que genere o manipule imagen, sonido o vídeo que se asemeje a personas, objetos, lugares u otras entidades o sucesos existentes y que puedan inducir erróneamente a pensar que son auténticos o verídicos.

Esta no aplicación, salvo la excepción citada, nos recuerda mucho a la denominada «excepción doméstica» del RGPD¹⁴, que supone su no aplicación cuando las

14. Véase al respecto:

TRIBUNAL DE JUSTICIA DE LA UNIÓN EUROPEA, Asunto C-101/01. Bodil Lindqvist y Göta hovrätt. 6 de noviembre. ECLI:EU:2003:596, 2003.

<https://curia.europa.eu/juris/document/document.jsf?docid=48382&doclang=ES>

TRIBUNAL DE JUSTICIA DE LA UNIÓN EUROPEA, Asunto C-212/13. Frantisek Rynes y Urad pro ochranu osobnich udaju. 11 de diciembre. ECLI: EU: C:2014:2428, 2014.

<https://curia.europa.eu/juris/document/document.jsf?docid=160561&doclang=ES>

TRIBUNAL DE JUSTICIA DE LA UNIÓN EUROPEA, Asunto C-25/17. Tietosuojavaltuutettu con la intervención de Jehovan todistajat — uskonnollinen yhdyksunta.

actividades llevadas a cabo por una persona física, como podrían ser la libreta de direcciones del móvil o correo electrónico, ambos particulares y no profesionales, puedan ser calificadas como tal. Las autoridades de control de protección de datos personales interpretan esta excepción de forma muy restrictiva, de manera que, por ejemplo, una publicación de una fotografía o vídeo de contenido sexual de un tercero en una red social supondría que quien ha realizado dicha publicación pueda ser sancionado por infringir el RGPD, y más concretamente, por haber efectuado dicha publicación en ese canal de forma, permitiendo su acceso indiscriminado, y existiendo falta de legitimación para ello.

En todo caso, la previsión del apartado 1 del artículo 71 debe ser entendida cuando estemos ante las circunstancias del artículo 52, ya que de lo contrario no tendría sentido, puesto que para el resto de los casos no se aplicará el RIA.

Respecto a las garantías procesales de actuación de la autoridad de supervisión de mercado se extienden tanto en lo referente al procedimiento administrativo que debe seguirse en la tramitación del expediente sancionador correspondiente, en el que el presunto responsable podrá alegar y presentar pruebas, así como la posibilidad de que una vez se haya resuelto en vía administrativa siendo sancionado pueda acudir a la vía jurisdiccional.

V. PROPUESTAS Y CAMBIOS DEL PARLAMENTO Y EN LAS VERSIONES FINALES

1. CAMBIOS PROPUESTOS EN EL PARLAMENTO

Tras este texto adoptado en el Consejo Europeo de 6 de diciembre de 2022, posición común sobre el proyecto de RIA, el 14 de junio de 2023, propone también introducir numerosas enmiendas en el seno del Parlamento Europeo, afectando considerablemente al contenido del artículo 71, siendo las más relevantes las siguientes.

En primer lugar, la facultad otorgada a los países de la Unión para determinar el régimen de sanciones se limitaría respecto a infracciones cometida por cualquier operador, teniendo en cuenta que esta figura engloba, según su definición al «proveedor, el fabricante de productos, el usuario, el representante autorizado, el importador o el distribuidor».

En segundo lugar, aumenta el contenido de la multa a imponer contempladas en el apartado 3 del artículo 71, sobre incumplimientos de la prohibición de prácticas de inteligencia artificial del artículo 5, pasando a 40 millones de euros o 7% del volumen de negocio total anual mundial del ejercicio financiero anterior, si la cuantía fuese superior; un nuevo apartado 3 bis establece que los incumplimientos, además del artículo 10, del artículo 13, podrán ser multados hasta 20 millones de euros o 4% del volumen de negocio; los incumplimientos del resto de preceptos que no sean los citados artículos 5, 10 y 13, serán de hasta 10 millones de euros o 2€ del volumen de negocio; y en el apartado 5 del citado precepto, respecto a incumplimientos sobre

10 de julio. ECLI:EU:C:2018:551, 2018.<https://curia.europa.eu/juris/document/document.jsf?docid=203822&doclang=ES>

información inexacta, incompleta o engaños a organismos notificados y autoridades nacionales, se reduce la multa a 5 millones de euros o 1% de ese volumen de negocio.

En tercer lugar, aparecen otros instrumentos derivados del ejercicio de la potestad sancionadora, a lo que hicimos alusión que podrían ser regulados por la norma interna de cada país, como son órdenes o advertencias, pudiendo acudir a ellos en vez de a la multa.

En cuarto lugar, se introducen más criterios para valorar la cuantía de la multa, la mayoría ya previstos en el RGPD, como son «las acciones emprendidas por el operador para mitigar los perjuicios o los daños sufridos por las personas afectadas»; «la intencionalidad o negligencia»; «el grado de cooperación con las autoridades nacionales competentes con el fin de poner remedio a la infracción y mitigar sus posibles efectos adversos»; «el grado de responsabilidad del operador, teniendo en cuenta las medidas técnicas y organizativas que aplique»; «la forma en que las autoridades nacionales competentes tuvieron conocimiento de la infracción, en particular si el operador notificó la infracción y, en tal caso, en qué medida»; «la adhesión a códigos de conducta o a mecanismos de certificación aprobados»; «cualquier otra infracción previa pertinente del operador»; y «cualquier otro factor agravante o atenuante aplicable a las circunstancias de cada caso».

Para conocer el contenido de cada uno de ellos, en el ámbito de la protección de datos personales, pero que podría dar una idea de su aplicación sobre las infracciones de proveedores de IA se puede acudir a los documentos aprobados al respecto por el CEPD¹⁵.

En quinto lugar, se establece la obligación de informar anualmente a la Oficina de IA sobre las multas que se hayan impuesto, que se constituye como un organismo independiente de la Unión Europea, pudiendo adoptar directrices, conjuntamente con la Comisión, sobre la norma, debiendo ser tenidas en cuenta a efectos sancionadores.

Y en sexto y último lugar, se añade un apartado 8 bis en el artículo 71 prohibiendo que las sanciones, costes de litigios y reclamaciones de indemnización no podrán ser objeto de cláusulas contractuales ni otra forma de reparto de cargas entre proveedores y distribuidores, importadores, implementadores o cualquier tercero. Sobre esta prohibición, y su aplicación al ámbito de protección de datos personales, se ha manifestado nuestro Tribunal Supremo en relación a que los incumplimientos de un responsable no pueden ser atribuidos a su encargado de tratamiento para que sea éste el que pague las multas o indemnizaciones¹⁶.

15. Véase al respecto: COMITÉ EUROPEO DE PROTECCIÓN DE DATOS, *Directrices 04/2022 sobre el cálculo de multas administrativas del RGPD, Versión 2.0*. 24 de mayo de 2023. https://edpb.europa.eu/system/files/2023-06/edpb_guidelines_042022_calculationofadministrativefines_en.pdf

GRUPO SOBRE PROTECCIÓN DE DATOS DEL ARTÍCULO 29, *Directrices sobre la aplicación y la fijación de multas administrativas a efectos del Reglamento 2016/679*, 3 de octubre de 2017.

<https://ec.europa.eu/newsroom/article29/items/611237/en>

16. Véase al respecto TRIBUNAL SUPREMO, SALA DE LO CIVIL, STS 1543/2023 — ECLI:ES:TS:2023:1543, 19 de abril de 2023.

2. LAS NOVEDADES EN LOS TEXTOS FINALES

Con el objeto de cerrar el texto, durante la presidencia española de Consejo, y durante tres días a principios de diciembre de 2023, se celebran los trilogos cerrando el texto, que, tras unos ajustes, en su versión, si bien sin publicarse todavía en el DOCE, se publica a principios de febrero de 2024¹⁷.

De esta versión final, sobre el art.71, comparando con las otras versiones expuestas anteriormente, destacar cambios sobre las cuantías de las multas a imponer por los incumplimientos que puedan darse, así como los criterios de graduación de las infracciones, desapareciendo alguno de los propuestos.

Así, sobre las cuantías a imponer en caso de incumplimiento, nos encontramos con la siguiente escala:

— La multa más alta, hasta 35 millones de € o 7% de facturación anual durante el ejercicio anterior si el responsable es una empresa, el que sea mayor, cuando el artículo 5 haya sido infringido.

— Hasta 15 millones de €, o 3%, el que sea mayor, cuando la infracción verse sobre obligaciones de los proveedores del art.16; obligaciones de los representantes conformes al art.25; obligaciones de los importadores del art.26; obligaciones de los distribuidores del art.27; obligaciones de los implementadores del art. 29 apartados 1 a 6 bis; y requisitos y obligaciones de los organismos notificados según el art.33 y art.34 apartados 1, 3 y 4, y art.34 bis; y obligaciones de transparencia para proveedores y usuarios según el art.52.

— Y en el último escalón, 7.500.000 €, o 1%, cuando se suministre información incorrecta, incompleta o engañosa, y al igual que los anteriores, la que sea mayor.

— Como excepción cuando se trate de una PYME, no será la cuantía mayor sino la menor.

Respecto a los criterios de graduación, son la naturaleza, gravedad y duración de la infracción; si se han aplicado multas por otros organismos de vigilancia de mercado o por otra autoridades; el tamaño, volumen de negocios y cuota de mercado del operador; cualquier otro agravante o atenuante, como los beneficios obtenidos o pérdidas evitadas; el grado de cooperación con las autoridades nacionales con el fin de remediar la infracción y mitigar los perjuicios; el grado de responsabilidad del operador atendiendo a los aspectos técnicos y medidas organizativas implementadas; la forma en que se tuvo conocimiento; la intencionalidad o negligencia; y cualquier medida adoptada por el operador para mitigar el daño.

Por otra parte, se consolida el hecho de que por los Estados miembros deban realizar dos actuaciones al respecto, como son establecer otras sanciones más allá de la multa, y si las Administraciones públicas son multadas o no.

17. CONSEJO EUROPEO DE LA UNIÓN EUROPEA, Nota de prensa «Ley de inteligencia artificial: el Consejo y el Parlamento llegan a un acuerdo sobre las primeras normas para la IA en el mundo», 9 de diciembre de 2023. Actualizado el 2 de febrero de 2024.

Ley de inteligencia artificial: el Consejo y el Parlamento llegan a un acuerdo sobre las primeras normas para la IA en el mundo — Consilium (europa.eu)

Finalmente, y tras su paso en marzo de 2024 por el Parlamento Europeo, el texto aprobado¹⁸ modifica la numeración de este artículo 71 que pasa a ser el artículo 99. También se precisa en el apartado 1 que las medidas que establezcan los Estados sobre el régimen de sanciones y otras medidas de ejecución, pueden ser advertencias o medidas de carácter no pecuniario.

Sobre las cuantías a imponer en concepto de multa, son las mismas que las descritas en el texto resultante de la Presidencia Española, pero modificando el número correspondiente de cada artículo en relación a las multas de hasta 15 millones de euros o 3% del volumen de negocios mundial del infractor cuando se vulnere: las obligaciones de los proveedores del artículo 16; las obligaciones de los representantes autorizados del artículo 22; las obligaciones de los importadores del artículo 23; las obligaciones de los distribuidores del artículo 24; las obligaciones de los responsables del despliegue del artículo 26; los requisitos y obligaciones de los organismos notificados del artículo 31, artículo 33 apartados 1, 3 y 4 y artículo 34; y las obligaciones de transparencia de los proveedores y usuarios con arreglo al artículo 50.

Asimismo, con alguna matización y añadido, los criterios para graduar las multas serán los siguientes: la naturaleza, gravedad y duración de la infracción y sus consecuencias, teniendo en cuenta la finalidad del sistema de IA, y si procede, el número de personas afectadas y el nivel de los daños sufridos; si otras autoridad de vigilancia de uno o varios Estados han impuesto multas al mismo operador por la misma infracción; si otras autoridades han impuesto multas al mismo operador por otro tipo de infracciones que se deriven de la misma actividad u omisión que constituya una infracción pertinente del RIA; el tamaño, volumen de negocios anual y la cuota de mercado del operador que comete la infracción; cualquier otro factor agravante o atenuante como beneficios financieros o pérdidas evitadas, directa o indirectamente a través de la infracción; grado de cooperación con las autoridades nacionales a fin de poner remedio a la infracción y mitigar sus efectos adversos; grado de responsabilidad del operador, atendiendo a las medidas técnicas y organizativas aplicadas; forma en que las autoridades nacionales han tenido conocimiento de la infracción, en particular si el operador notificó la infracción y, en tal caso, en qué medida; intencionalidad o negligencia; y las acciones del operador para mitigar los perjuicios sufridos por las personas afectadas.

En todo caso, se echa en falta dos aspectos. Primero, un apartado en que claramente se especifique quienes pueden ser considerados infractores, aunque de todo el precepto se desprende que serán los operadores y organismos notificados. Segundo, parece limitarse el régimen sancionador a la vulneración de los preceptos descritos, con lo que pueden existir obligaciones en el resto del texto, que, en caso de incumplimiento, quedarían sin sanción.

18. PARLAMENTO EUROPEO, Resolución legislativa del Parlamento Europeo, de 13 de marzo de 2024, sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión.

VI. EL ARTÍCULO 72 SOBRE LA POTESTAD SANCIONADORA DEL SUPERVISOR EUROPEO DE PROTECCIÓN DE DATOS

Este segundo precepto completa el régimen sancionador previsto en el RIA, si bien matizando que su objeto es diferente puesto que principalmente regula la potestad sancionadora del Supervisor Europeo de Protección de Datos (SEPD) respecto a posibles incumplimientos por parte de las instituciones, agencias y organismos de la Unión Europea a los que se le aplique la norma.

Comenzamos, al igual que hicimos con el artículo 71, con el análisis del contenido de la Propuesta de RIA de la Comisión Europea, cuyo contenido podemos dividir en tres apartados claramente diferenciado.

El primero, atribuye esa competencia sancionadora, incluyendo la posibilidad de imponer multas, al Supervisor Europeo de Protección de Datos, de manera que en el ámbito de la Administración de la Unión Europea no se va a crear un organismo específico para ejercer la supervisión en materia de IA. Esta opción podría haberse aplicado igualmente en nuestro país, atribuyendo esta función a la Agencia Española de Protección de Datos, si bien ha recaído en la Agencia Española de Supervisión de la Inteligencia Artificial (AESIA). De hecho, el artículo 59 «Designación de autoridades nacionales competentes», permite que hubiese sido la AEPD, sin necesidad de crear un organismo nuevo.

Asimismo, y a diferencia del artículo 71 que atribuye a cada país la competencia para que decida si su sector público puede ser multado, en el caso de las instituciones, organismos y agencias de la Unión, la norma sí contempla esta posibilidad, en consonancia con el Reglamento 2018/1725, que atribuye también esta posibilidad de imponer multas al SEPD respecto a los anteriormente citados, cuando los incumplimientos afecten a protección de datos personales.

En suma, se aprovecha la experiencia del Supervisor que pasará a ostentar una doble labor de supervisión y control, tanto en inteligencia artificial como en protección de datos personales. De hecho, este organismo es partidario de que la autoridad de supervisión de IA sea la autoridad de protección de datos, debido a su experiencia en la gestión de riesgos que afecten a los derechos fundamentales, así como lograr una aplicación coherente de la norma¹⁹.

El segundo, referido a las cuestiones más relevantes sobre el ámbito sancionador como son la tipificación de infracciones, multas previstas y criterios de graduación, cuya diferencia principal, si lo comparamos con el contenido del artículo 71, radica en la cuantía de las multas a imponer. Así, serán sensiblemente inferiores, siendo de hasta 500.00 € por el incumplimiento de los artículos 5 y 11, y hasta 250.000 € para el resto de los incumplimientos diferentes a los preceptos citados.

También se introducen como nuevo criterio de graduación de las cuantías la cooperación con el SEPD con el fin de poner remedio a la infracción y mitigar sus posibles efectos adversos incluida el cumplimiento de sus órdenes; así como la reincidencia, en base a toda infracción anterior similar cometida por la institución, agencia u organismo de la Unión. Extraña que estos criterios no estén entre los

19. SUPERVISOR EUROPEO DE PROTECCIÓN DE DATOS, *Dictamen 44/2023 sobre la propuesta de Ley de Inteligencia Artificial a la vista de la evolución legislativa*, de 2023. https://edps.europa.eu/system/files/2023-10/2023-0137_d3269_opinion_en.pdf

previstos en el artículo 71, puesto que también resulta aplicable cuando actúe una autoridad de vigilancia de mercado. Sí aparece en cierta medida el segundo, pero redactado de forma más compleja de manera que no podemos afirmar que su contenido se califique como «reincidencia». El otro criterio que contempla este artículo 72, son similares a los ya citados del artículo 71, como son la graduación, debiendo valorarse, además de la naturaleza, gravedad y duración de la infracción.

En cuanto al tercero, complementa el anterior al regular elementos de la propia tramitación del procedimiento sancionador como son el derecho a ser oído que podrá ejercer el presunto infractor antes de ser sancionado y el derecho de acceso al expediente, sin perjuicio de garantizará el interés legítimo de personas físicas y empresas para proteger sus datos personales o secretos comerciales.

Por su parte, en el texto de este artículo 72 de la Propuesta de RIA del Consejo de la Unión Europea de 6 de diciembre de 2022, no se proponen cambios de su contenido.

En cambio, sí los hay por parte del texto de las enmiendas del Parlamento Europeo, que podemos clasificarlas igualmente en dos grupos: uno sobre los criterios de cuantificación de las multas; y el otro sobre la tipificación y la cuantía de las multas, y que como expondremos es donde existen mayores cambios.

Sobre el primero, se introducen nuevos criterios, muchos de los cuales, se ha propuesto su inclusión en el artículo 71, cuestión lógica puesto que no tendría mucho sentido que en el ámbito de aplicación de este precepto se aplicasen unos, y en el del artículo 72 otros, sin perjuicio de que puede existir alguno que no fuese aplicable si tenemos en cuenta la naturaleza de los organismos e instituciones de la Unión Europea. Así, se propone incluir que junto a la naturaleza, gravedad y duración se considere también el propósito del sistema de la IA, el número de personas, daños sufridos y cualquier infracción anterior; cualquier medida para mitigar el perjuicio a las personas (aunque este criterio en cierta manera aparece ya en el texto del proyecto de IA de la Comisión); el grado de responsabilidad de la institución, agencia u organismo de la Unión, considerando las medidas técnicas y organizativas que aplique; la forma en que el SEPD ha tenido conocimiento de la infracción; y el presupuesto anual del organismo. A todo ellos, debe añadirse que las multas no afectarán al funcionamiento efectivo de la institución, órgano u organismo de la Unión sancionado, que puede utilizarse igualmente como un criterio a la hora de imponer la multa.

Sobre el segundo, el incumplimiento de las prohibiciones del artículo 5 será multado hasta 1'5 millones de euros; del artículo 10 hasta 1 millón euros; y el resto 750.000 euros, con lo que observamos que las cuantías de las multas suben considerablemente si lo comparamos con el texto del proyecto de RIA de la Comisión.

Respecto al texto surgido de la presidencia española y los trílogos citados ambos anteriormente, al igual que con el artículo 71, se aprecian modificaciones en cuanto las cuantías e infracciones, siendo dos grupos. La infracción del artículo 5 puede llegar hasta 1.500.000 de €, y cualquier otra infracción diferente que afecte a otros preceptos de la norma, de hasta 750.000 €.

En cuanto al texto final aprobado el RIA, este artículo 72 cambia de numeración siendo el artículo 100, las cuantías son idénticas a las anteriores, así como los criterios de graduación. La mayor novedad aparece en el artículo siguiente, el 101, al otorgar a la Comisión la capacidad de multar a los proveedores de modelos de IA de uso general

que no superen el 3% de volumen de negocios mundial o 15 millones de euros, si esta cifra es superior, cuando tenga lugar, ya sea mediante culpa o negligencia, alguna de estas conductas: incumplir el RIA; no atender la solicitud de información del art.91 o facilitarla de forma inexacta, incompleta o engañosa; incumplir una medida solicitada del artículo 93; o no dar acceso a la Comisión al modelo de IA de uso general o de uso con riesgo sistémico para llevar a cabo la evaluación del artículo 92.

Es decir, además de la potestad sancionadora ejercida por los Estados, también la propia Comisión podrá imponer multas siempre y cuando se den los requisitos descritos.

VII. CONCLUSIONES.

La principal premisa de la regulación del régimen sancionador, independientemente de la materia que se trate, es dotar de seguridad jurídica, de manera que los destinatarios de la norma puedan conocer en todo momento los hechos por los que pueden ser sancionados, las cuantías, y hasta cuándo. Del contenido analizado en sus diferentes versiones, se desprenden tres cuestiones que no logran regularse de forma plena, de manera que no se pueda alcanzar esa seguridad jurídica. Nos referimos a la ausencia de sujetos presuntamente infractores, la deficiente técnica de tipificación de infracciones, y la no regulación del régimen de la prescripción. Curiosamente, estas tres deficiencias aparecen también en el RGPD, por lo que se vuelvan a repetir los mismos errores, si bien, respecto a la parte de protección de datos personales, algunos han sido en cierta medida solventados.

Sobre los sujetos presuntamente responsables de cometer una infracción, si bien del propio art.71 se desprende que alcanzará a todos los operadores, hubiese sido más clarificador un apartado en el que se estableciese de una forma más concreta.

En cuanto a las infracciones, si bien es cierto que finalmente se han acotado una serie de preceptos sobre cuyo incumplimiento se aplicaría el régimen sancionador, como son los artículos 5, 16, 22, 23, 24, 26, 31, 33 apartados 1, 3 y 4, 34 y 50, lo que supone, como expusimos anteriormente, ir precepto a precepto para valorar donde puede cometerse una infracción y donde no. Deberían incluirse cuáles son las conductas que realmente son susceptibles de ser sancionadas. Por ello, consideramos que debería existir un Anexo describiendo las posibles infracciones, indicando las conductas punibles, que podría completarse determinando también el plazo de prescripción.

De lo contrario, y puesto que esa seguridad jurídica sigue siendo necesario, será nuestro legislador el trate de dotarla, como ha hecho con la LOPDGDD, si bien teniendo presente que no puede tipificar las infracciones, puesto que dicha tipificación, aunque de forma muy amplia, está en el RIA. Seguramente, se adopte por la misma solución que en la citada LOPDGDD: descripción de conductas sancionables a los efectos de fijar el plazo de prescripción.

Por otra parte, resulta bastante llamativo que la única medida sancionadora que se contempla sea la multa, cuando se pueden utilizar también otros instrumentos. Por ello, consideramos que tendría que incluirse el apercibimiento, la advertencia, medidas de cumplimiento, así como la retirada, prohibición y restricción de un producto de IA. Las tres últimas, en cambio, sí aparecen en el artículo 65.

También se necesita completar los criterios para imponer las multas, siendo el Parlamento Europeo el que ha propuesto más enmiendas para introducir nuevos criterios. Recordemos que el texto de la Comisión únicamente incluía tres.

En suma, debería mejorarse la tipificación de infracciones y criterios de graduación de las multas, así como introducir quienes pueden ser los presuntamente responsables, la figura de la prescripción, y más medidas a aplicar que no sea únicamente la multa.

Como complemento a todo ello, y atendiendo a la experiencia del RGPD, consideramos que la norma no debería dejar a cada país que decidiese sobre si aplicar las multas o no a su sector público, puesto que, de ser así, unos países, probablemente la mayoría lo contemplarán, pero algunos no, perdiendo uniformidad. Debería recogerse esa posibilidad de multar. Además, se advierte cierta contradicción que la norma lo deje abierto y en cambio sí le otorgue esa facultad al SEPD respecto a los organismos e instituciones de la Unión Europea.

Por último, y para finalizar, aplaudimos que al SEPD se le haya atribuido esta potestad sancionadora en relación con los anteriormente citados, sin necesidad de crear un organismo específico, lo que redundaría en una mayor efectividad, eficiencia, y reducción de tiempos. Si existiesen dos organismos, uno en protección de datos, y otro en IA, como ocurre en nuestro país, resulta indispensable una estrecha colaboración entre ambos, atendiendo sobre todo a las implicaciones entre ambas materias.

Derecho a presentar una reclamación y derecho a una explicación. Vías de recurso para los particulares en el reglamento de inteligencia artificial

AURELIO LOPEZ-TARRUELLA MARTÍNEZ

Profesor Titular Derecho internacional privado
Universidad de Alicante

I. INTRODUCCIÓN

El objetivo del presente trabajo es analizar la Sección 4 («Vías de recurso») del Capítulo IX («Vigilancia poscomercialización, intercambio de información, vigilancia del mercado») del RIA, que abarca los artículos 85 a 87. Como su título indica, en dicha sección se recogen las vías de recurso con las que cuentan los particulares contra el incumplimiento del Reglamento por parte de proveedores, responsables del despliegue o cualquier otro operador involucrado en la cadena de valor de la IA.

Se trata de una Sección introducida por el Parlamento Europeo que no aparecía en la Propuesta inicial de la Comisión Europea ni en la Posición Común del Consejo. Con ella se refuerza la posición de las personas que pueden resultar afectadas por las decisiones adoptadas a partir de la información proporcionada por los sistemas de IA. Se trataba de una omisión que en la inicial propuesta de la Comisión había sido criticada por la doctrina. De hecho, esta sección es la única del Reglamento en la que se recogen derechos para las personas afectadas por el funcionamiento de un sistema de IA: el derecho a presentar una reclamación; y el derecho a explicación de decisiones tomadas individualmente. En cualquier caso, como veremos, el texto finalmente adoptado no es totalmente satisfactorio por cuanto reduce la efectividad de estos derechos.

El trabajo se divide en tres partes. Las dos primeras referidas respectivamente al artículo 85, donde se regula el derecho a presentar una reclamación ante una autoridad de vigilancia del mercado; y al artículo 86, relativo al derecho a explicación de decisiones tomadas individualmente. En la tercera parte, se explican los artículos 87 y 110, los cuales tienen un carácter auxiliar y contienen, respectivamente, una remisión a la Directiva 2019/1937 sobre la protección de los denunciantes¹; y

1. Directiva (UE) 2019/1937 del Parlamento Europeo y del Consejo, de 23 de octubre de 2019, relativa a la protección de las personas que informen sobre in-

una modificación del Anexo de la Directiva 2020/1828² para garantizar que las asociaciones de representación de los intereses colectivos de los consumidores pueden iniciar acciones colectivas por incumplimiento del Reglamento.

II. EL DERECHO A PRESENTAR UNA RECLAMACIÓN ANTE UNA AUTORIDAD DE VIGILANCIA DEL MERCADO

La doctrina ha sostenido, acertadamente, que las leyes digitales europeas³, entre las que se encuentra el RIA, se inspiran en el Reglamento 2016/679 general de protección de datos (en adelante RGPD) tanto en su estructura como en el contenido de alguna de sus disposiciones⁴.

A primera vista, podría suponerse que el artículo 85 es una manifestación de esta inspiración por cuanto su título y el contenido del primer apartado es similar a la regulación del derecho a presentar una reclamación recogido en el artículo 77 RGPD, artículo 53 RSD, artículo 14.1 *in fine* y 24.1 *in fine* RGD, y artículo 38 RD. Esta interpretación resulta apoyada por la redacción del artículo 110, que abre la vía para que, al igual que en estos instrumentos, los organismos de representación colectiva de los intereses de los consumidores puedan presentar, en el marco de la Directiva 2020/1828, reclamaciones por el incumplimiento del Reglamento.

No obstante, esa suposición resulta desacreditada por la remisión que el apartado 2 del artículo 85 realiza al Reglamento 2019/1020 sobre vigilancia del mercado⁵ a efectos de la regulación de estas reclamaciones. Y ello porque, esta remisión conlleva la atribución a ese derecho de un contenido totalmente diferente al establecido en el RGPD y el resto de leyes digitales. A mi modo de ver, esta regulación especial reduce enormemente la utilidad y efectividad práctica de este derecho, circunstancia que puede constituir un obstáculo a la consecución de los objetivos del Reglamento.

fracciones del Derecho de la Unión, disponible en <http://data.europa.eu/eli/dir/2019/1937/oj>

2. Directiva (UE) 2020/1828 del Parlamento Europeo y del Consejo de 25 de noviembre de 2020 relativa a las acciones de representación para la protección de los intereses colectivos de los consumidores, y por la que se deroga la Directiva 2009/22/CE, disponible en <http://data.europa.eu/eli/dir/2020/1828/oj>
3. A los efectos del presente trabajo, además del RIA, se entiende por «leyes digitales europeas»: Reglamento 2022/868 de gobernanza de los datos (RGD), Reglamento 2022/1925 sobre mercados digitales (RMD); el Reglamento 2022/2065 de servicios digitales (RSD) y el Reglamento 2023/2854 sobre normas armonizadas para un acceso justo a los datos y su utilización (RD).
4. En el mismo sentido, GASCÓN MACÉN, A., «El Reglamento General de Protección de Datos como modelo de las recientes propuestas de legislación digital europea», *CDT*, Vol 13(2), 2021, pp. 209-232. <https://doi.org/10.20318/cdt.2021.6256>; PÁPAKONSTANTINOÚ, V. / DE HERT, P. «Post GDPR EU laws and their GDPR mimesis. DGA, DSA, DMA and the EU regulation of AI», *European Law Blog*, 1 abril 2021, <https://europeanlawblog.eu/2021/04/01/post-gdpr-eu-laws-and-their-gdpr-mimesis-dga-dsa-dma-and-the-eu-regulation-of-ai/>
5. Reglamento (UE) 2019/1020 del Parlamento Europeo y del Consejo, de 20 de junio de 2019, relativo a la vigilancia del mercado y la conformidad de los productos y por el que se modifican la Directiva 2004/42/CE y los Reglamentos (CE) 765/2008 y (UE) 305/2011, disponible en <http://data.europa.eu/eli/reg/2019/1020/oj>

La disposición, además, debe ponerse en concordancia con otras disposiciones del RIA, en particular con el artículo 74. De acuerdo con esta disposición, la competencia para conocer (o, mejor dicho, «tener en cuenta») estas reclamaciones, no corresponde a una única autoridad (en el caso de España, a la AESIA), sino a varias, dependiendo de la clasificación del sistema IA. Esta circunstancia puede dificultar el ejercicio de este derecho y menoscabar, más si cabe, su utilidad práctica. Es más, el margen de discreción que establece esta disposición puede generar situaciones indeseadas de *forum shopping* entre las autoridades de vigilancia de mercado de diferentes Estados miembros.

1. EVOLUCIÓN DEL TEXTO DE LA DISPOSICIÓN EN LOS TRABAJOS PREPARATORIOS

Como se ha indicado en la Introducción, la propuesta inicial de la Comisión europea no incluía este derecho, circunstancia criticada por la doctrina⁶ y en la Opinión Conjunta del Supervisor y el Comité europeo de datos personales⁷. A pesar de ello, el Consejo no creyó conveniente incluir este derecho en su Posición Común de noviembre de 2022⁸. Sí que lo hizo el Parlamento Europeo en las enmiendas 628 y 629 a la Propuesta de la Comisión aprobadas el 14 junio 2023⁹. La inclusión de este derecho venía acompañada de otras disposiciones complementarias inspiradas igualmente en el RGPD:

Artículo 68 bis. Derecho a presentar una reclamación ante una autoridad nacional de supervisión

1. Sin perjuicio de cualquier otro recurso administrativo o acción judicial, toda persona física o grupo de personas físicas tendrá derecho a presentar una reclamación ante una autoridad nacional de supervisión, en particular en el Estado miembro en el que tenga su residencia habitual, lugar de trabajo o lugar de la supuesta infracción, si considera que el sistema de IA que le concierne infringe el presente Reglamento.

6. SMUHA, N. et al, «How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act», 2021, disponible en <https://ssrn.com/abstract=3899991>; EBERS, M. et al., «The European Commission's Proposal for an Artificial Intelligence Act-A Critical Assessment by Members of the Robotics and AI Law Society (RIALS)», *J*, vol. 4, 2021, pp. 589-603. <https://doi.org/10.3390/j4040043>; VEALE, M. y ZUIDERVEEN BORGESIUS, F.J., «Demystifying the Draft EU Artificial Intelligence Act — Analysing the good, the bad, and the unclear elements of the proposed approach», *Computer Law Review International*, vol. 22, 2021, pp. 97-112, esp. 111; LÚCIA RAPOSO, V. «Ex machina: preliminary critical assessment of the European Draft Act on artificial intelligence», *International Journal of Law and Information Technology*, Vol. 30, Issue 1, 2022, p. 102.
7. EUROPEAN DATA PROTECTION BOARD / EUROPEAN DATA PROTECTION SUPERVISOR, «Joint Opinion 5/2021 on the Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)», 18 June 2021, p. 18, disponible en https://edpb.europa.eu/our-work-tools/our-documents/edpbedps-joint-opinion/edpbedps-joint-opinion-52021-proposal_en
8. Posición común del Consejo de 22 noviembre 2022 (disponible en <https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf>).
9. https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_ES.html

2. La autoridad nacional de supervisión ante la que se haya presentado la reclamación informará al reclamante sobre el curso y el resultado de la reclamación, inclusive sobre la posibilidad de acceder a la tutela judicial en virtud del artículo 78.

Artículo 68 ter. Derecho a la tutela judicial efectiva contra una autoridad nacional de supervisión

1. Sin perjuicio de cualquier otro recurso administrativo o extrajudicial, toda persona física o jurídica tendrá derecho a la tutela judicial efectiva contra una decisión jurídicamente vinculante de una autoridad nacional de supervisión que le concierna.

2. Sin perjuicio de cualquier otro recurso administrativo o extrajudicial, toda persona física o jurídica tendrá derecho a la tutela judicial efectiva en caso de que la autoridad nacional de supervisión competente con arreglo al artículo 59 no dé curso a una reclamación o no informe al interesado en el plazo de tres meses sobre el curso o el resultado de la reclamación presentada en virtud del artículo 68 bis.

3. Las acciones contra una autoridad nacional de supervisión deberán ejercerse ante los tribunales del Estado miembro en el que esté establecida la autoridad nacional de supervisión.

4. Cuando se ejerzan acciones contra una decisión de una autoridad nacional de supervisión que haya sido precedida de un dictamen o una decisión de la Comisión en el marco del procedimiento de salvaguardia de la Unión, la autoridad de supervisión remitirá al tribunal dicho dictamen o decisión.

El artículo 68.bis propuesto por el Parlamento es una copia prácticamente literal del artículo 77 RGPD. No deja lugar a dudas de que ello la equivocación que se comete al referirse en apartado 2 a «la posibilidad de acceder a la tutela judicial en virtud del artículo 78». Esta disposición, *pero del RGPD*, es la que regula el «derecho a la tutela judicial efectiva contra una autoridad de control» La remisión debería ser al artículo 68.ter, disposición que regula dicha posibilidad en el RIA. El error es demostrativo de la voluntad del Parlamento de introducir un derecho a presentar una reclamación con el mismo contenido que el artículo 77 RGPD y otras leyes digitales europeas.

La necesidad de llegar a un texto de compromiso en fase de triálogo conlleva dos cambios importantes. Por un lado, el artículo 68.ter acaba teniendo una redacción sustancialmente diferente de la propuesta por el Parlamento. Estas diferencias son particularmente importantes en lo que respecta al apartado 2 del actual artículo 85.

Artículo 85. Derecho a presentar una reclamación ante una autoridad de vigilancia del mercado

1. Sin perjuicio de otros recursos administrativos o judiciales, las reclamaciones ante la autoridad de vigilancia del mercado pertinente podrán ser presentadas por cualquier persona física o jurídica que tenga motivos para considerar que se ha producido una infracción de las disposiciones del presente Reglamento.

2. De conformidad con el Reglamento (UE) 2019/1020, las reclamaciones se tendrán en cuenta a efectos de la realización de las actividades de vigilancia del mercado y se tramitarán con arreglo a los procedimientos específicos establecidos al efecto por las autoridades de vigilancia del mercado.

Por otro lado, se elimina el artículo 68.ter de la propuesta del Parlamento. Ello se debe a que, como se explicará en el siguiente epígrafe, de acuerdo con el R. 2019/1020, la autoridad de vigilancia del mercado no tiene obligación de adoptar una decisión

relativa a la reclamación. Consiguientemente, el derecho a presentar un recurso contra esa decisión ante una autoridad judicial pierde sentido. Esto evidencia que la regulación de este derecho en el RIA es completamente diferente a la prevista en el RGPD y el resto de leyes digitales europeas.

La razón principal que puede explicar la redacción final del artículo 85 es que, al contrario que en el RGPD y otras leyes digitales, el RIA no atribuye la función de vigilancia del mercado de los sistemas IA de alto riesgo a una única autoridad sino a varias. Más allá de la AESIA, el RIA atribuye competencias a las autoridades designadas por los instrumentos de implementación de la normativa recogida en el Anexo I, Sección A (recuérdese, instrumentos que forman parte del nuevo marco regulatorio que fijan los requisitos de entrada en el mercado de máquinas, juguetes, dispositivos médicos, vehículos de aviación, etc...). Estas autoridades, cuyas actividades están reguladas por el R. 2019/1020, no tienen entre sus funciones conocer de reclamaciones presentadas por particulares, y no cuentan con personal que pueda hacerse cargo de ellas, más todavía si se tiene en consideración las complejidades que pueden revestir las reclamaciones relativas al cumplimiento de los requisitos y obligaciones del RIA por parte de un sistema de IA.

Sin duda, en línea con las enmiendas del Parlamento Europeo, se podía haber optado por atribuir competencias para conocer de estas reclamaciones exclusivamente a una autoridad (AESIA). De hecho, el artículo 74.3 permite a los Estados miembros atribuir, «en circunstancias adecuadas», competencia a una autoridad de vigilancia del mercado diferente a la designada por los actos jurídicos del nuevo marco regulatorio. Así, por ejemplo, la competencia para conocer de todas las reclamaciones de los particulares se podría concentrar en la AESIA. Ahora bien, esto podría generar dos tipos de problemas: de descoordinación y solapamiento de funciones entre la AESIA y las autoridades de vigilancia del mercado en sectores específicos; de *forum shopping* en relación con autoridades de otros Estados miembros, como se explicará más adelante.

Siendo esta la razón principal, queda la sospecha de si el vaciado de contenido del derecho de reclamación se enmarca dentro de la tendencia observada al final de las negociaciones de rebajar los requisitos y obligaciones de ciertas categorías de sistemas de IA para facilitar su desarrollo en la Unión Europea.

Asimismo, podría argumentarse que la ausencia de un verdadero derecho a presentar una reclamación quedaría compensado por la obligatoriedad para los responsables del despliegue de sistemas de IA de alto riesgo, introducida en la última versión del Reglamento, de llevar a cabo evaluaciones de impacto relativa a los derechos fundamentales (artículo 27).

En cualquier caso, aunque la regulación final pueda justificarse en las razones expuestas ello no es obstáculo para afirmar, como explicamos a continuación, que la consecución de los objetivos del RIA (artículo 1) puede verse dificultada.

2. DIFERENCIAS ENTRE LA REGULACIÓN DEL DERECHO DE RECLAMACIÓN EN EL REGLAMENTO Y LA ESTABLECIDA EN EL RGPD Y OTRAS LEYES DIGITALES EUROPEAS

La doctrina considera el derecho a presentar una reclamación recogido en el RGPD una herramienta de gran utilidad por dos motivos. En primer lugar, porque

como ha puesto de manifiesto la doctrina¹⁰ y el TJUE¹¹, favorece la defensa del derecho fundamental a la protección de datos personales, al ofrecer a los interesados una vía más sencilla y gratuita para hacer valer sus derechos. En segundo lugar, porque incentiva a los responsables y encargados a cumplir con el Reglamento, puesto que, si la supervisión recae únicamente en las autoridades públicas, esta puede resultar afectada por motivos relativos a la falta de recursos de estas autoridades, de los escasos conocimientos de su personal, o de motivos políticos o geoestratégicos (piénsese en el caso de Irlanda o Luxemburgo). La introducción de este derecho permite a los particulares participar en la vigilancia del mercado y reclamar sus derechos, circunstancia que incentiva a las empresas a cumplir con sus obligaciones reglamentarias¹². A nadie escapa el gran impulso que ha supuesto para el cumplimiento efectivo del RGPD las actuaciones llevadas a cabo por Maximilian Schrems, NOYB o *La Quadrature du Net*.

La utilidad de este derecho es trasladable a todos aquellos reglamentos en los que se ha recogido el derecho con una regulación similar a la prevista en el RGPD. No es este el caso del RIA y ello porque, como decimos, la regulación es diferente.

En el RGPD, este derecho tiene un contenido mínimo irrenunciable establecido en el propio reglamento y en la jurisprudencia del TJUE. Así, el artículo 77.2 RGPD indica que «[l]a autoridad de control ante la que se haya presentado la reclamación informará al reclamante sobre el curso y el resultado de la reclamación, inclusive sobre la posibilidad de acceder a la tutela judicial». Además, el TJUE en su sentencia de 7 diciembre 2023, C-26/22, *SCHUEFA*, ha establecido que la disposición obliga a la autoridad de control a adoptar una decisión administrativa que sea susceptible de un control jurisdiccional pleno, relativa a los argumentos de fondo y no exclusivamente a cuestiones procedimentales¹³. El procedimiento de reclamación no se asemeja al de una petición, sino que se concibe como un mecanismo capaz de proteger de manera eficaz los derechos y los intereses de los interesados¹⁴.

En relación con aquellas cuestiones sobre la reclamación no reguladas expresamente en el Reglamento, la STJUE de 12 enero 2023, C-132/21, *Budapesti Elektromos Művek*, recuerda que corresponde a los Estados miembros, en aplicación del principio de autonomía procesal determinar los procedimientos a partir de los cuales se deben articular este derecho¹⁵. Ahora bien, la regulación de esas vías

10. DE MIGUEL ASENSIO, P., *Derecho privado de Internet*, 6ª Ed, Civitas, Madrid, 2022, p. 537; AGENCIA DE LOS DERECHOS FUNDAMENTALES DE LA UNIÓN EUROPEA, CONSEJO DE EUROPA, SUPERVISOR EUROPEO DE PROTECCIÓN DE DATOS, TRIBUNAL EUROPEO DE DERECHOS HUMANOS, *Manual de legislación europea en materia de la protección de datos: edición de 2018*, Oficina de Publicaciones de la Unión Europea, 2019, <https://data.europa.eu/doi/10.2811/60145>

11. SSTJUE de 31 diciembre 2016, C-203/15, «Tele2 Sverige», ap. 123; de 6 octubre 2015, «Schrems», ap. 41; de 8 abril 2014, «Digital Ireland», ap. 68.

12. Claro está que el éxito del sistema depende de que los Estados miembros doten a estas autoridades de los recursos necesarios para que lleven a cabo su labor de manera eficaz.

13. Ap. 70.

14. Ap. 58.

15. Ap. 45: «A falta de normativa de la Unión en la materia, cada Estado miembro debe configurar, en virtud del principio de autonomía procesal de los Estados miembros, la regulación de los procedimientos administrativos y judiciales destinados a ase-

procedimentales no debe poner en entredicho el efecto útil y la protección efectiva de derecho a presentar una reclamación¹⁶. Del mismo modo, «esa regulación no debe ser menos favorable que la referente a los recursos semejantes establecidos para la protección de los derechos reconocidos por el ordenamiento jurídico interno (principio de equivalencia) ni hacer imposible en la práctica o excesivamente difícil el ejercicio de los derechos conferidos por el ordenamiento jurídico de la Unión (principio de efectividad)»¹⁷.

Por último, de la jurisprudencia del TJUE se extrae que la interpretación de estas disposiciones debe tomar en consideración los considerandos 10 y 11 del Reglamento. De acuerdo con el primero, el objetivo del Reglamento es garantizar un nivel elevado de protección de las personas físicas por lo que se refiere al tratamiento de datos personales en la Unión. El considerando 11 del mismo Reglamento señala, además, que la protección efectiva de estos datos exige que se refuercen los derechos de los interesados¹⁸.

Esta jurisprudencia es difícilmente trasladable a la interpretación del artículo 85 RIA. En este caso, de acuerdo con el apartado 2, la regulación de la tramitación y efectos de la reclamación se remite al Reglamento 2019/1020. Dicho reglamento no regula un derecho de reclamación, en el sentido del RGPD. El artículo 11.3 de ese Reglamento, único en el que se contiene una referencia a los consumidores y operadores económicos, indica que éstos podrán presentar una reclamación que la autoridad de vigilancia tendrá en cuenta, entre otros factores, para decidir que comprobaciones realizar a los efectos de determinar si se cumple el Reglamento:

«A la hora de decidir qué comprobaciones realizar, de qué tipos de productos y a qué escala, las autoridades de vigilancia del mercado seguirán un enfoque basado en el riesgo, teniendo en cuenta los siguientes factores:

e) las reclamaciones de los consumidores y otra información recibida de otras autoridades, operadores económicos, medios de comunicación y otras fuentes que puedan indicar incumplimiento».

No existe en esta disposición, ni en ninguna otra del Reglamento, una obligación de la autoridad de vigilancia del mercado de informar al reclamante sobre el curso y el resultado de la reclamación. Y tampoco existe una obligación como la establecida por el TJUE en relación con el RGPD de otorgar una respuesta de fondo a la reclamación. La redacción del artículo 85 se explica por la necesidad de hacerla coincidir con la regulación del Reglamento 2019/1020: «*las reclamaciones se tendrán en cuenta a efectos de la realización de las actividades de vigilancia del mercado y se tramitarán con arreglo a*

gurar un nivel elevado de salvaguarda de los derechos que el Derecho de la Unión confiere a los justiciables».

16. Ap. 47: «[...]la regulación de la aplicación de las referidas vías de recurso concurrentes e independientes no debe poner en entredicho el efecto útil y la protección efectiva de los derechos garantizados por ese Reglamento».

17. La existencia de esta jurisprudencia no ha podido evitar la aparición de importantes diferencias en las regulaciones internas del derecho de reclamación previsto en el art. 77. Al respecto puede consultarse EDPS, *Study on the National Administrative Rules Impacting the Cooperation Duties for the National Supervisory Authorities — Final Report*, EDPS 2019/02-07, 2020, disponible en https://edpb.europa.eu/system/files/2023-04/call_7_final_report_07012021.pdf

18. Ap. 61, STJUE de 7 diciembre 2023, C-26/22, SCHUFA.

los procedimientos específicos». De esta disposición se puede inferir que los Estados miembros tienen libertad para decidir los procedimientos específicos para tramitar esas reclamaciones, y el valor que se le pueden atribuir. Lo único que establece el artículo 85.2 RIA y reitera el artículo 11.3 R. 2019/1020 es que dichas reclamaciones «se tendrán en cuenta». Pero no existe una obligación para las autoridades de vigilancia de tratar esas reclamaciones de manera individual, o de explicar las razones por las cuales esas reclamaciones, en su caso, no son finalmente tomadas en consideración a la hora de realizar comprobaciones o iniciar investigaciones. En definitiva, no existe una obligación de la autoridad de vigilancia del mercado de adoptar una decisión relativa a la reclamación.

La particular naturaleza que el artículo 85 atribuye al derecho a presentar una reclamación explica que, al contrario que en el RGPD (y otras leyes digitales), los particulares no puedan recurrir la decisión o la omisión de una decisión por parte de la autoridad de vigilancia de mercado ante la jurisdicción contencioso-administrativa. De ello se deriva la supresión en el texto final, del artículo propuesto por el Parlamento (artículo 68.ter) que recogía este derecho.

La escasa regulación de este derecho en el artículo 85 RIA deja, además, algunos interrogantes. Para empezar, está por ver si nuestro legislador adoptará una regulación especial en relación con las reclamaciones relativas a sistemas de IA cuya vigilancia corresponde a AESIA. En principio, la remisión del artículo 74 al R. 2019/1020 es aplicable a todos «los sistemas de IA cubiertos por el [...] Reglamento» lo que induce a pensar que las reclamaciones ante AESIA podrían recibir en mismo tratamiento que en ese tratamiento. No obstante, nada impide que se pueda adoptar una regulación similar al establecido en el RGPD en la que, al menos, la AESIA queda obligada a responder al reclamante, y que la decisión sea susceptible de recurso ante los juzgados del orden contencioso-administrativo. Sin duda ello favorecería la protección de los objetivos del Reglamento, pero plantea dos inconvenientes. Primero, que se ocasionaría un agravio comparativo: mientras los usuarios de sistemas IA sujetos a la vigilancia de AESIA tendrían un verdadero derecho de reclamación; el resto no. Segundo, que si el resto de Estados miembros no adoptaran una regulación similar a la que se propone, se podrían plantear casos de *forum shopping*: los usuarios (y, en particular, las asociaciones que los representan) optarían por presentar las reclamaciones ante las autoridades de Estados miembro que establece una regulación más beneficiosa para sus intereses del derecho a presentar una reclamación.

La remisión del artículo 85.2 al Reglamento 2019/1020 plantea dudas en relación con los sistemas de IA de alto riesgo en el sector financiero, y los que utilicen con fines policiales, gestión de fronteras, administración de justicia y procesos democráticos. Como veremos más adelante, de acuerdo con los apartados 6 y 8 del artículo 74, la autoridad competente en estos casos es la Comisión Nacional del Mercado de Valores y la AEPD respectivamente.

Parece lógico pensar que las reclamaciones ante la primera de estas autoridades se tramitarán en atención a la normativa específica del sector financiero, por lo que la remisión del artículo 85.2 al Reglamento 2019/1020 carece de sentido. La precisión es relevante por cuanto la CNMV cuenta con un servicio de reclamaciones propio¹⁹.

19. <https://www.cnmv.es/portal/inversor/reclamaciones.aspx>

Ahora bien, debe recordarse que el artículo 74.7 permite a los Estados miembros, en las circunstancias apropiadas y siempre que se garantice la coordinación, atribuir la competencia para supervisión la aplicación del RIA a una autoridad diferente (que podría ser AESIA).

Más difícil resulta llegar a una conclusión en el caso de la AEPD por cuanto, en puridad, la reclamación presentada con base en el artículo 85 RIA no será por una infracción del RGPD. Pero tampoco resulta lógico que la AEPD acabe aplicando un Reglamento (el 2019/1020) que no está incluido dentro de sus competencias.

Por último, en vista de que el artículo 85.1 RIA no ofrece un mecanismo efectivo para que los particulares puedan reclamar un incumplimiento del Reglamento, cabe plantearse si existe alguna otra vía para hacerlo. En este sentido, se puede plantear la presentación de una demanda ante tribunales civiles frente al operador presuntamente incumplidor en la que se solicite el cese de una actividad que no es conforme con el Reglamento, o una indemnización por los perjuicios que esa actividad haya podido ocasionar. No creo que sea un impedimento para ello el hecho de que, al contrario que ocurre en el artículo 79 RGPD, el texto del Reglamento no recoja expresamente esta posibilidad. Es más, cabe recordar, que el artículo 85.1 se abre indicando que el derecho de presentación de reclamaciones se disfruta *«[s]in perjuicio de otros recursos administrativos o judiciales»*.

Esta vía tiene el problema de que implica costes elevados y el procedimiento puede demorarse un período largo de tiempo. Se trata, a mi modo de ver, una opción viable únicamente para organismos de representación colectiva. Ciertamente, está por ver qué ocurre en la práctica, pero la doctrina ha planteado este escenario en relación con la otra ley digital europea en la que tampoco se regula el derecho a presentar una reclamación: el Reglamento de Mercados Digitales²⁰.

3. REGULACIÓN DEL DERECHO A PRESENTAR UNA RECLAMACIÓN EN EL REGLAMENTO

Habiendo explicado lo que hubiera sido deseable que fuera el derecho a presentar una reclamación en el RIA, pero no es; procede a continuación explicar lo que finalmente es. A tales efectos, cabe entender el derecho como una mera facultad de cualquier persona física o jurídica de presentar una reclamación ante autoridades de vigilancia de mercado en la que informa de una presunta infracción del RIA por parte de un operador contemplado en el Reglamento. Dicha reclamación también puede ir referida al incumplimiento por parte del responsable del despliegue de la obligación establecida en el artículo 86 de ofrecer una explicación sobre la decisión adoptada a partir de los resultados ofrecidos por determinados sistemas de IA de alto riesgo.

Como se ha explicado, de acuerdo con el artículo 85.2 RIA y artículo 13.1 R. 2019/1020, la única obligación de las autoridades es tener en cuenta dichas reclamaciones a la hora de decidir si inician comprobaciones que, en su caso, dará lugar a investigaciones en el sentido de los artículos 79 y 80 RIA. Cabe exceptuar de esta regla general a la CNMV que, a nuestro modo de ver, deberá tramitar las

20. G. MONTI «Procedures and institutions» en AAVV, *Effective and Proportionate Implementation of the DMA*, CERRE, 2023 pp. 164 ss, esp. 181, disponible en https://cerre.eu/wp-content/uploads/2023/01/DMA_Book-1.pdf

reclamaciones relativas a sistemas de IA utilizados en el sector financiero de acuerdo con sus reglas particulares.

Habiendo explicado estas cuestiones en el apartado anterior, corresponde en las líneas que siguen precisar alguna otra cuestión relativa a este derecho a presentar una reclamación.

En primer lugar, el artículo 85 indica que la legitimación para presentar la reclamación recae en «cualquier persona física o jurídica que tenga motivos para considerar que se ha producido una infracción de las disposiciones del presente Reglamento». Como veremos en el apartado IV, de acuerdo con el artículo 87, estas personas se pueden beneficiar de la protección que ofrece la Directiva 2019/1937 de protección de los denunciantes.

Al contrario que el artículo 80 RGPD (y otras leyes digitales europeas), el RIA no regula expresamente la posibilidad de que las personas afectadas por un incumplimiento otorguen mandato a una entidad, organización o asociación sin ánimo de lucro para que los represente ante la autoridad competente. Sin duda, esta ausencia de regulación se debe al sentido completamente diferente del derecho de reclamación en el RIA. No obstante, dicha representación se nos antoja posible por dos razones. Primero, porque de no ser así, el mandato del artículo 110 de incluir el RIA en el Anexo de la Directiva 2020/1828 sobre la protección de los intereses colectivos de los consumidores carecería de sentido. Segundo, porque el artículo 9 del Reglamento 2019/1020 establece la facultad de las autoridades de vigilancia del mercado de acordar con «organizaciones que representen a operadores económicos o a usuarios finales», la realización de actividades conjuntas con objeto de incentivar el cumplimiento o detectar casos de incumplimiento.

En segundo lugar, se plantea la cuestión de determinar la autoridad ante la que debe presentarse la reclamación desde una doble dimensión: la material (determinación de la autoridad competente por razón de la materia), y la territorial (determinación del Estado miembro ante cuya autoridad debe plantearse la reclamación).

En relación con la primera dimensión, el artículo 85 debe leerse conjuntamente con varias disposiciones que determinan las autoridades competentes para llevar a cabo la vigilancia del mercado de sistemas de IA. Al respecto, debe hacerse la siguiente clasificación.

a) Sistemas de IA basados en un modelo del IA de propósito general. La competencia corresponde a la Oficina de IA, la cual «tendrá todas las competencias de una autoridad de vigilancia del mercado en el sentido del Reglamento 2019/1020» (artículo 75).

b) Sistemas de IA de alto riesgo destinados a utilizarse como componentes de un producto o que constituyen en sí mismo un producto cubierto por la legislación armonizada enumerada en la sección A del Anexo I (artículo 74.3). Como se recordará, esa legislación, que forma parte del «nuevo marco normativo», está referida, entre otros, a juguetes, maquinas, embarcaciones, ascensores, equipos radioeléctricos, productos sanitarios o vehículos de motor. En estos casos, la reclamación deberá presentarse ante la autoridad designada en cada instrumento legislativo. Así, en

España, dependiendo del producto, la competencia puede corresponder a agencias o unidades administrativas pertenecientes a múltiples ministerios²¹.

c) Sistemas de IA de alto riesgo comercializados, puestos en servicio o utilizados por entidades financieras reguladas por la legislación de la Unión en la materia (artículo 74.6). En estos casos, la autoridad competente para conocer de la reclamación es la Comisión Nacional del Mercado de Valores. Ahora bien, como se ha adelantado, esta atribución de competencia podría cambiar, por cuanto el apartado 7 establece que, «en circunstancias justificadas y siempre que se garantice la coordinación, el Estado miembro podrá designar a otra autoridad pertinente como autoridad de vigilancia del mercado a efectos del presente Reglamento».

d) Sistemas se utilicen con fines policiales y para los fines enumerados en los puntos 6, 7 y 8 del anexo III, es decir, asuntos relacionados con la aplicación de la ley; gestión de la migración, el asilo y el control fronterizo; y la administración de justicia y los procesos democráticos (artículo 74.8). En este caso, la competencia para conocer de la reclamación corresponde a la AEPD.

e) Prácticas de IA prohibidas, sistemas IA de alto riesgo no enumerados en el anexo III (sistemas IA independientes) y sistemas IA que no son de alto riesgo. La competencia para conocer de reclamaciones en estos supuestos corresponde a AESIA.

Según se ha indicado, la distribución de la competencia para conocer de las reclamaciones en una pluralidad de autoridades puede dificultar al ejercicio de este derecho. Para reducir este problema, resultan de gran relevancia las obligaciones previstas en el artículo 70.2 para los Estados miembros: designación de un punto de contacto único; y puesta a disposición del público, por medios de comunicación electrónica, de información sobre la forma de contactar con las autoridades competentes y los puntos de contacto únicos. Además, la Comisión Europea está obligada a publicar en línea un listado de dichos puntos de contacto a nivel europeo.

En relación con la dimensión territorial, cabe plantearse ¿ante las autoridades de qué Estado miembro debe presentarse la reclamación? En el marco del RGPD esta cuestión plantea una relevancia especial por la naturaleza de la reclamación y por la necesidad de facilitar su presentación por parte de los particulares. Esto lleva a legislador europeo a establecer reglas de jurisdicción especiales que atribuyen la competencia a las autoridades de la residencia del interesado, y que suponen una excepción en relación con la regla general del Reglamento según la cual la competencia para supervisar a los responsables y encargados corresponde a las autoridades de los Estados miembros de establecimiento (o, en su caso, del establecimiento principal)²².

En el caso del derecho a presentar una reclamación del artículo 85 RIA, esta atribución de competencia reviste una menor importancia por dos razones. Primero, el carácter particular de la reclamación que, como se viene sosteniendo, se tramitará de acuerdo con el Reglamento 2019/1012 y las autoridades pueden o no tener en cuenta. Segundo, porque el RIA ha optado por un sistema descentralizado de competencia: las autoridades de vigilancia del mercado tienen jurisdicción para

21. El listado de autoridades de vigilancia del mercado pueden consultarse en https://single-market-economy.ec.europa.eu/single-market/goods/building-blocks/market-surveillance/organisation_en

22. Art. 56 RGPD.

conocer de las reclamaciones relativas a infracciones ocurridas en el territorio de su Estado miembro²³. En aquellos supuestos (presumiblemente habituales en la práctica) en los que la infracción del sistema de IA ocurra en más de un Estado miembro, este sistema de competencia descentralizado permite que la reclamación se presente ante la autoridad de cualquiera de esos Estados. Esta circunstancia supone una puerta abierta al *forum shopping*: es de esperar que los particulares y, en particular, las asociaciones de representación colectivas opten por presentar la reclamación en Estados miembros que establezca vías procedimentales para su presentación más amigables, o que resulten ser más proclive a iniciar investigaciones o que, sencillamente, cuente con mayores recursos, o mayores conocimientos y preparación técnica. De ahí que, como decíamos anteriormente, la concentración de competencias para conocer de reclamaciones de los particulares en la AESIA, aunque beneficioso para los particulares, puede tener un peligroso «efecto llamada».

III. DERECHO A UNA EXPLICACIÓN DE LA TOMA DE DECISIONES INDIVIDUALES

El segundo derecho otorgado a los particulares en el RIA es el referido a obtener una explicación de la toma de decisiones individuales. Con ello se refuerza la exigencia de explicabilidad, de acuerdo con apuesta europea por una IA fiable, deben tener todos los sistemas de IA desarrollados y comercializados en la Unión. Se trata de un derecho inspirado en el artículo 22 RGPD, por lo que consideramos conveniente recurrir a las *Directrices sobre decisiones individuales automatizadas* del antiguo Grupo de Trabajo del artículo 29²⁴. Precisamente, la relación con esa disposición del RGPD es la principal incógnita que plantea el artículo 86, razón por la cual se le va a dedicar un apartado específico. Asimismo, resulta relevante analizar la relación de este derecho con la protección que los derechos de propiedad intelectual y el secreto comercial puede otorgar a muchos de los elementos presentes en un sistema de IA.

1. EVOLUCIÓN DEL TEXTO DE LA DISPOSICIÓN EN LOS TRABAJOS PREPARATORIOS

El derecho a una explicación ha seguido un camino similar al del derecho a presentar una reclamación del artículo 85 RIA. Ni la Propuesta inicial de la Comisión, ni la Posición Común del Consejo de diciembre 2022 contenían referencia alguna a él. La primera referencia al mismo se encuentra en la Enmienda 630 del Informe del Parlamento Europeo de junio 2023, donde se propone la introducción de un nuevo artículo 68 quater con la siguiente redacción:

-
23. Un análisis comparativo de las ventajas e inconvenientes de los diferentes sistemas de supervisión de los instrumentos regulatorios europeos puede encontrarse en MONTI, G., DE STREEL, A., «Improving EU Institutional Design to Better Supervise Digital Platforms», *CERRE Report*, 2022, disponible en <https://cerre.eu/publications/improving-eu-institutional-design/>
24. GRUPO DE TRABAJO DEL ART. 29, «Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679, de 3 de octubre 2017», disponible en <https://ec.europa.eu/newsroom/article29/items/612053>

1. *Toda persona afectada sujeta a una decisión adoptada por el implementador sobre la base de la información de salida de un sistema de IA de alto riesgo que produzca efectos jurídicos o que le afectan significativamente de una manera que considera que perjudica a su salud, seguridad, derechos fundamentales, bienestar socioeconómico o cualquier otro de sus derechos derivados de las obligaciones establecidas en el presente Reglamento, tendrá derecho a solicitar al implementador una explicación clara y significativa, de conformidad con el artículo 13, apartado 1, sobre el papel del sistema de IA en el procedimiento de toma de decisiones, los principales parámetros de la decisión adoptada y los datos de entrada correspondientes.*

2. *El apartado 1 no se aplicará a la utilización de sistemas de IA para los que el Derecho nacional o de la Unión prevea excepciones o restricciones a la obligación prevista en el apartado 1 en la medida en que dichas excepciones o restricciones respeten la esencia de los derechos y libertades fundamentales y sean una medida necesaria y proporcionada en una sociedad democrática.*

3. *El presente artículo se aplicará sin perjuicio de lo previsto en los artículos 13, 14, 15 y 22 del Reglamento (UE) 2016/679.*

El texto finalmente adoptado no varía mucho si bien incluye importantes precisiones:

Artículo 86 Derecho a una explicación de la toma de decisiones individuales

1. *Toda persona afectada que sea objeto de una decisión adoptada por el responsable del despliegue sobre la base de los resultados de un sistema de IA de alto riesgo enumerado en el anexo III, con excepción de los sistemas enumerados en el punto 2, y que produzca efectos jurídicos o le afecte significativamente de manera similar de forma que considere que repercute negativamente en su salud, seguridad y derechos fundamentales, tendrá derecho a solicitar al responsable del despliegue explicaciones claras y significativas sobre el papel del sistema de IA en el procedimiento de toma de decisiones y los principales elementos de la decisión adoptada.*

2. *El apartado 1 no se aplicará al uso de sistemas de IA para los que se deriven excepciones o restricciones a la obligación establecida en el apartado 1 del Derecho de la Unión o nacional en cumplimiento del Derecho de la Unión.*

3. *El presente artículo sólo se aplicará en la medida en que el derecho contemplado en el apartado 1 no esté ya previsto en la legislación de la Unión.*

Los cambios entre ambos textos que cabe destacar son los siguientes:

a) El texto final limita el ámbito de aplicación del derecho a los sistemas de IA de alto riesgo del Anexo III, con la excepción de los previstos en el punto 2.

b) Se simplifica la redacción del apartado primero, al reducir los bienes jurídicos que pueden resultar afectados por la decisión, a «salud, seguridad y derechos fundamentales», lo cual resulta adecuado por cuanto coincide con los objetivos perseguidos por el RIA (véase el artículo 1) y se evita la ambigüedad que, a mi modo de ver, podía plantear el resto de bienes jurídicos mencionados en la versión inicial («bienestar socioeconómico o cualquier otro de sus derechos derivados de las obligaciones establecidas en el presente Reglamento»).

c) Se suaviza la obligación del responsable del despliegue refiriéndose de manera más genérica al «procedimiento de toma de decisiones y los principales elementos de la decisión adoptada». La referencia explícita en la versión del Informe del Parlamento a «los principales parámetros» y «los datos de entrada correspondiente»

podía dificultar el cumplimiento de la obligación y, además, generar dudas sobre si resultaban obligados a desvelar información susceptible de protección por derechos de propiedad intelectual o secreto comercial.

d) De la exclusión del apartado 2, se elimina la condición de que las excepciones o restricciones «respeten la esencia de los derechos y libertades fundamentales y sean una medida necesaria y proporcionada en una sociedad democrática». La exclusión es acertada por cuanto parece una condición que podría generar problemas interpretativos, y que desconoce que, por definición, las normas de la UE o de los ordenamientos nacionales que puedan establecer dichas excepciones deberían, por definición, cumplir esas exigencias por tratarse de ordenamientos donde se respeta el Estado de Derecho y los valores europeos.

e) La exclusión del apartado 3 sustituye la referencia explícita a los artículos del RGPD (*artículos 13, 14, 15 y 22*) por la remisión genérica a la «legislación de la Unión». En cualquier caso, como se verá a lo largo del análisis, el RGPD es el principal afectado por la disposición.

2. EL PRINCIPIO DE EXPLICABILIDAD DE LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL

El uso de complejos algoritmos para automatizar la toma de decisiones en nuestro día a día se ha convertido en algo habitual. Está presente en los procesos de selección de personal en las empresas públicas o privadas, la concesión de créditos, el precio de una póliza de seguros, o la personalización de la publicidad que recibe el usuario de una red social. Con carácter general, el funcionamiento de estos algoritmos resulta desconocido e ininteligible no sólo para las personas afectadas por las decisiones automatizadas, sino también por la empresa que utiliza el sistema de IA. Esto ocurre, en particular, si el algoritmo está basado en técnicas de aprendizaje automático y, más concretamente, de aprendizaje profundo. Esto ha llevado a la doctrina a acuñar el término de la sociedad de la «caja negra», en relación con el hecho de que las decisiones que mueven nuestra vida diaria, las cuales cada vez se adquieren mayor relevancia y mayor alcance, se adoptan por sistemas que no entendemos y que, por tanto, no podemos saber en qué se basan²⁵.

La opacidad de los sistemas de IA no sólo constituye un problema por el obstáculo que supone al derecho que toda persona tiene de conocer las razones por las que se adoptan decisiones que le afectan; sino también porque dicho desconocimiento impide investigar las razones por las que un sistema algorítmico que comete errores. Estos errores pueden tener consecuencias materiales (como, por ejemplo, el funcionamiento defectuoso de un dispositivo conectado al Internet de las cosas); y también personales, que pueden ir desde una lesión física (derivada, por ejemplo, de un accidente provocado por un vehículo autónomo²⁶) hasta perjuicios económicos o morales derivados de ser rechazados en una selección de personal,

25. PASQUALE, F., *The Black Box Society: The Secret Algorithms That Control Money and Information*, Cambridge-London, Harvard University Press, 2015.

26. LÍ, M., «Another Self-Driving Car Accident, Another AI Development Lesson», *Towards Data Science*, 20 noviembre 2019, disponible en <https://towardsdatascience.com/another-self-driving-car-accident-another-ai-development-lesson-b2ce3d-bb4444>

ver denegado un crédito por una entidad bancaria, no ser aceptada para participar en unas oposiciones públicas, o ser calificado como una persona con un alto riesgo de fuga²⁷. Estos errores pueden llegar a generar perjuicios en colectivos enteros que pueden incluso a contravenir valores fundamentales. Se habla en tales casos de sesgos discriminatorios²⁸. Tal es el caso cuando los sistemas de IA llevan a adoptar decisiones o realizar predicciones que encierran discriminaciones por motivos de raza u origen étnico²⁹, opiniones políticas, religión o creencias, afiliación sindical, condición genética o estado de salud u orientación sexual³⁰.

Estos errores pueden estar relacionados con el modelo (el algoritmo elegido, los parámetros utilizados, los pesos atribuidos a cada variable, etc...); o con los conjuntos de datos con los que se ha entrenado: la falta de volumen, variedad o calidad de los datos con los que se ha entrenado el sistema, o la existencia de defectos en el preprocesado de datos (duplicaciones, generalizaciones, etc...)³¹.

En las *Directrices éticas para una IA fiable*³², el Grupo de expertos de alto nivel sobre inteligencia artificial de la Comisión Europea indica, acertadamente, que los seres humanos y las comunidades solamente podrán confiar en el desarrollo tecnológico y en sus aplicaciones si contamos con un marco claro y detallado para garantizar su

-
27. Puede suceder que el uso de determinados algoritmos de la IA para predecir la reincidencia delictiva dé lugar a prejuicios raciales o de género, y prevea una probabilidad de reincidencia distinta para hombres y mujeres o para nacionales y extranjeros. Vease TOLAN S., MIRÓN M., GÓMEZ E. y CASTILLO C., «Why Machine Learning May Lead to Unfairness: Evidence from Risk Assessment for Juvenile Justice in Catalonia», *ICAIL 19: Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*, 2019, pp. 83-92, disponible en <https://doi.org/10.1145/3322640.3326705>
 28. COMISIÓN EUROPEA, «Libro Blanco sobre inteligencia artificial — un enfoque europeo orientado a la excelencia y confianza», Doc. COM(2020) 65 final, pp. 13-15; FRA — AGENCIA DE LA UNIÓN EUROPEA PARA LOS DERECHOS FUNDAMENTALES, *Data quality and artificial intelligence— mitigating bias and error to protect fundamental rights*, Luxembourg, Publications Office, 2019, <https://fra.europa.eu/en/publication/2019/artificial-intelligence-data-quality>
 29. CHIVERS, T., «Facial recognition... coming to a supermarket near you», 4 agosto, 2019, *The Guardian*, disponible en <https://www.theguardian.com/technology/2019/aug/04/facial-recognition-supermarket-facewatch-ai-artificial-intelligence-civil-liberties>
 30. En general, vease O'NEILL, C., *Weapons of Math Destruction*, Nueva York, Random House, 2016.
 31. En este sentido, merece la pena recordar la polémica generada por un tweet de Yan Lecun, investigador jefe de Facebook, relativo a un modelo de IA que había sido utilizado para transformar a Barack Obama en un hombre blanco y que, a la postre, llevó al investigador a abandonar la red social: «ML systems are biased when data is biased. This face upsampling system makes everyone look white because the network was pretrained on FlickrFaceHQ, which mainly contains white people pics. Train the *exact* same system on a dataset from Senegal, and everyone will look African». Vease «Yann LeCun Quits Twitter Amid Acrimonious Exchanges on AI Bias», *Synced*, 30 junio 2020, disponible en <https://syncedreview.com/2020/06/30/yann-lecun-quits-twitter-amid-acrimonious-exchanges-on-ai-bias/>
 32. GRUPO INDEPENDIENTE DE EXPERTOS DE ALTO NIVEL SOBRE INTELIGENCIA ARTIFICIAL (2018), *Directrices para una IA fiable*, disponible en <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

fiabilidad. Una inteligencia artificial fiable se basa en tres componentes: que sea ética; que sea lícita; y que sea robusta. En relación con el primero de estos componentes, las *Directrices* identifican cuatro principios éticos que deben cumplirse para garantizar que los sistemas de IA se desarrollen, desplieguen y utilicen de manera fiable: respeto de la autonomía humana, prevención del daño, equidad y explicabilidad³³.

En opinión del Grupo de expertos:

«La explicabilidad es crucial para conseguir que los usuarios confíen en los sistemas de IA y para mantener dicha confianza. Esto significa que los procesos han de ser transparentes, que es preciso comunicar abiertamente las capacidades y la finalidad de los sistemas de IA y que las decisiones deben poder explicarse —en la medida de lo posible— a las partes que se vean afectadas por ellas de manera directa o indirecta. Sin esta información, no es posible impugnar adecuadamente una decisión. No siempre resulta posible explicar por qué un modelo ha generado un resultado o una decisión particular (ni qué combinación de factores contribuyeron a ello). Esos casos, que se denominan algoritmos de “caja negra”, requieren especial atención. En tales circunstancias, puede ser necesario adoptar otras medidas relacionadas con la explicabilidad (por ejemplo, la trazabilidad, la auditabilidad y la comunicación transparente sobre las prestaciones del sistema), siempre y cuando el sistema en su conjunto respete los derechos fundamentales. El grado de necesidad de explicabilidad depende en gran medida del contexto y la gravedad de las consecuencias derivadas de un resultado erróneo o inadecuado».

El principio de explicabilidad recogido en las *Directrices* inspira una serie de requisitos introducidos en el RIA para los sistemas de IA de alto riesgo a los que es preciso referirse brevemente para poner en contexto en derecho a una explicación. Así, cabe recordar que el artículo 10 impone la obligación de implementar prácticas adecuadas de gobernanza y gestión de los datos; el artículo 11 requiere la existencia de documentación técnica; el artículo 12 la conservación de registros a lo largo de todo el ciclo de vida del sistema de IA para garantizar un nivel adecuado de trazabilidad de su funcionamiento; y el artículo 13, que exige que los sistemas de IA se diseñen de un modo que garanticen que funcionan con un nivel de transparencia suficiente para que los responsables del despliegue interpreten y usen correctamente su información de salida; en fin, el artículo 14, que indica que el sistema de IA debe diseñarse de tal manera que la persona física a la que se encomiende su vigilancia pueda llevar a cabo adecuadamente su labor.

El derecho a una explicación sirve de corolario a estos requisitos por cuanto si un responsable del despliegue no es capaz de ofrecer las explicaciones solicitadas por una persona afectada, ello puede ser debido a que el sistema de IA no cumple con algunos requisitos y, por lo tanto, estaremos ante un incumplimiento del Reglamento que no sólo afecta al particular que ejerció el derecho, sino a la sociedad en general. Ello debería ser un indicio suficiente para que la autoridad de vigilancia inicie actuaciones de investigación contra el responsable del despliegue y, en su caso, contra el proveedor.

3. CONDICIONES PARA EL EJERCICIO DEL DERECHO A UNA EXPLICACIÓN

El ejercicio del derecho a una explicación está sujeto a ciertas condiciones que, como se observará, reducen su impacto positivo.

33. *Idem*, p. 14.

En primer lugar, el derecho a una explicación sólo se disfruta en relación con las decisiones adoptadas por los responsables de despliegue de sistemas de IA de alto riesgo que figuren en el Anexo III, con excepción de los enumerados en el punto 2. Para entender el alcance de esta precisión es más sencillo referirse a los sistemas de IA excluidos.

Así, los sistemas de IA del punto 2 son los referidos a infraestructuras críticas. La relevancia que estas infraestructuras tienen para los poderes públicos provoca que su gestión quede enteramente en manos del Estado.

Se excluyen también los sistemas de IA que constituyen componentes de seguridad de los productos regulados por los actos legislativos enumerados en el Anexo I. Y ello, aunque el artículo 86 establece como una de las razones para solicitar explicaciones que la decisión adoptada tenga un efecto perjudicial para «su salud» o «su seguridad». Cabe pensar que, al igual que ocurre con el derecho a presentar una reclamación del artículo 85, la exclusión de estos sistemas de IA está basado en la idea de alterar en la menor medida posible el funcionamiento de las autoridades de vigilancia de los mercados de los productos a los que van referidos esos actos legislativos. En definitiva, esto significa que la facultad de exigir a una empresa que demuestre que el sistema de IA incorporado a uno de sus productos es «explicable» corresponde, exclusivamente, a la autoridad de vigilancia del mercado. En todo caso, la persona afectada podrá presentar una reclamación ante esa autoridad de acuerdo con el artículo 85 si bien dicha autoridad sólo está obligada a tenerla en cuenta.

Tampoco se disfruta este derecho en relación con sistemas de IA que no son considerados de alto riesgo, circunstancia que se justifica en las mismas razones que informan la regulación de estos sistemas: su escasa afectación a los derechos fundamentales. Y también cabe considerar excluidos los modelos IA de uso general.

En fin, la ausencia de una exclusión expresa, permite afirmar que el derecho sí se disfruta en relación con los modelos de IA de uso general. En estos casos, en ausencia de explicación por parte del responsable del despliegue, la reclamación deberá presentarse ante la Oficina de IA.

En segundo lugar, el derecho a una explicación previsto en el artículo 86 RIA tampoco está disponible en aquellos casos en los que:

a) existan excepciones o restricciones a la obligación de ofrecer explicaciones prevista en el artículo 86.1 derivadas del Derecho de la Unión o nacional de conformidad con el Derecho de la Unión (apartado 2 de la disposición).

b) el derecho a una explicación esté previsto de otro modo en el Derecho de la Unión (apartado 3). Como se explicará en el siguiente epígrafe, esta exclusión se refiere principalmente al derecho a impugnar una decisión automatizada prevista en el RGPD.

En tercer lugar, cabe precisar quiénes son los beneficiarios de este derecho. La disposición se refiere a «toda *persona* que se vea afectada por una decisión que el responsable del despliegue adopte». En la medida en que la redacción no establece diferencias, debe entenderse que pueden ejercer el derecho tanto personas físicas como personas jurídicas.

Ahora bien, en relación con las primeras, cabe recordar que el apartado 3 del artículo 86 indica que la disposición únicamente es aplicable «en la medida en que

el derecho a que se refiere el apartado 1 no esté previsto de otro modo en el Derecho de la Unión». Parece lógico afirmar que una decisión tomada individualmente por un sistema de IA que afecta una persona física estará basada en datos referidos a esa persona de carácter personal. Siendo así, el particular debería ejercer el derecho a impugnar una decisión individual automatizada recogida en el artículo 22 RGPD, y no este derecho. Como veremos en el siguiente epígrafe, esta circunstancia puede ser relevante por cuanto la doctrina ha afirmado que, como mínimo, resulta dudoso que el RGPD establezca un derecho a una explicación.

Por lo que respecta a la posibilidad de que la persona en cuestión esté representada por un organismo de protección colectiva, no parece que puedan albergarse dudas al respecto. Más aún con la modificación que el artículo 110 RIA introduce en la Directiva 2020/1828 que habilita a las asociaciones de protección colectiva de los intereses de los consumidores a presentar demandas colectivas por el incumplimiento del presente Reglamento.

Del mismo modo, como veremos en el apartado IV, de acuerdo con el artículo 87, estas personas se pueden beneficiar de la protección que ofrece la Directiva 2019/1937 de protección de los denunciantes. Esta remisión puede tener gran relevancia en aquellos casos en los que la persona que ejerce el derecho es un trabajador o un colaborador del responsable del despliegue que tiene conocimiento de primera mano de los errores que el sistema de IA puede cometer. No obstante, para poder ejercer el derecho, el denunciante deberá reunir las condiciones que se explican a continuación.

En cuarto lugar, la persona afectada sólo disfruta de un derecho a una explicación en relación con una decisión que «produzca efectos jurídicos o le afecte considerablemente del mismo modo, de manera que considere que tiene un efecto perjudicial para su salud, su seguridad o sus derechos fundamentales».

Para empezar, debe señalarse que el TJUE³⁴ sostiene que el término «decisión» debe interpretarse de manera amplia³⁵, circunstancia que le lleva a incluir los meros actos preparatorios que sirven para tomar la decisión³⁶, los cuales pueden haber sido llevados a cabo por personas diferentes.

A los efectos de interpretar esta condición, resulta adecuado tomar en consideración las *Directrices sobre decisiones individuales automatizadas*. En ellas se indica que una decisión produce «efectos jurídicos» si afecta a los derechos que el ordenamiento reconoce a una persona, y se pone como ejemplo la libertad de asociarse con otras personas, votar en unas elecciones o entablar acciones legales. También genera un efecto jurídico una decisión que afecte al estatuto jurídico de una persona (denegación de una prestación concedida por la ley, denegación de admisión en un país o la denegación de ciudadanía) o a sus derechos en virtud de un contrato (la cancelación de un contrato).

Por su parte, debe entenderse que una decisión que afecta «considerablemente del mismo modo a una persona», es lo mismo que una decisión que «afecta significativamente de modo similar» en el sentido del artículo 22 RGPD. Según las *Directrices*, se trata de decisiones que, si bien no producen ningún cambio en las

34. STJUE de 7 diciembre 2023, C-634/21, *SCHUFA*.

35. Aps. 44 y 45.

36. Aps. 61 y 62.

obligaciones o derechos jurídicos de la persona, le pueden afectar suficientemente como para exigir protección. Ahora bien, en estos casos, la decisión debe afectar «significativamente» o «considerablemente» a la persona. De acuerdo con las *Directrices*, esta circunstancia debe determinarse en cada supuesto, pero, en cualquier caso, los efectos de la decisión deben ser lo suficientemente importantes como para ser dignos de protección. Tal será el caso, por ejemplo, de las decisiones que afecten al acceso de una persona a estudios universitarios; o le denieguen una oportunidad laboral o la coloquen en desventaja; o, cuando el sistema de IA se utiliza con fines de mercadotecnia, si el perfilado de una persona provoca que se le ofrezcan productos o servicios a precios prohibitivamente elevados, circunstancia que impide que, en la práctica, pueda acceder a ellos³⁷.

No debe olvidarse que la decisión debe tener, en la persona afectada, un «efecto perjudicial para su salud, su seguridad o sus derechos fundamentales». La referencia debe ponerse en colación con los objetivos generales del RIA, previstos en el artículo 1.1, pero no parece que esta condición pueda suponer un obstáculo al ejercicio del derecho por cuanto es fácil imaginar que cualquier decisión adoptada a partir de la información proporcionada por un sistema IA afectará a uno de estos tres objetivos.

En quinto lugar, en claro contraste con el artículo 22 RGPD, el derecho a una explicación del artículo 86 RIA puede ejercerse en relación con una «decisión que el responsable del despliegue adopte basándose en los resultados de un sistema de IA». Su ámbito de aplicación es diferente del cubierto por la disposición del RGPD, la cual se refiere exclusivamente a «decisiones basadas únicamente en el tratamiento automatizado».

Este segundo supuesto cubre, por ejemplo, una decisión de denegación de una prestación social adoptada automáticamente por un sistema IA. En cambio, el artículo 86 RIA se refiere a la decisión de un funcionario que, en atención a la información o la sugerencia aportada por un sistema IA decide denegar la ayuda.

Resulta difícil pensar que la intención de las instituciones fuera excluir del ámbito del artículo 86 las decisiones individuales automatizadas. Por ello, a pesar de la defectuosa redacción, cabe interpretar que el derecho a una explicación puede ejercerse en relación con ambos tipos de decisiones.

En sexto lugar, resulta necesario precisar los requisitos que debe reunir la explicación a la que resulta obligado el responsable del despliegue por esta disposición. Estos requisitos están referidos a la forma («explicaciones claras y significativas») y al contenido («acerca del papel que el sistema de IA ha tenido en el proceso de toma de decisiones y los principales elementos de la decisión adoptada»).

Si bien la redacción establecida en el artículo 15.1. h) RGPD en relación con las decisiones individuales automatizadas no es idéntica («información *significativa* sobre la lógica aplicada, así como la importancia y las consecuencias previstas de dicho tratamiento para el interesado»), consideramos apropiado tomar como guía para interpretar estos requisitos las *Directrices sobre decisiones individuales automatizadas*.

En relación con la forma, este documento indica que se debe informar a la persona afectada de forma sencilla y de manera suficientemente exhaustiva para que entienda los motivos de la decisión.

37. GRUPO DE TRABAJO ART. 29, *Directrices ...*, op. cit., p. 24.

En relación con el contenido, se debe ofrecer información significativa sobre la lógica aplicada, no necesariamente una compleja explicación de los algoritmos utilizados o la revelación de todo el algoritmo³⁸. Al respecto debe recordarse que se ha rebajados los requisitos inicialmente propuestos por el Parlamento. Además, debe tenerse en cuenta el considerando 171 del RIA que añade que la explicación debe poder «servir de base para que las personas afectadas *puedan ejercer sus derechos*».

Como veremos en el epígrafe 4, la determinación del contenido de las «explicaciones claras y significativas» también tiene implicaciones desde el punto de vista de la propiedad intelectual y la protección de la información confidencial del proveedor y del responsable del despliegue de sistemas de IA.

Por último, cabe hacer referencia a las lagunas y dudas que plantea la regulación de este derecho. Primero, la disposición no establece un plazo al responsable del despliegue para ofrecer las explicaciones solicitadas. Esto es importante, a los efectos del ejercicio posterior de derechos al que se refiere el mencionado considerando 171. ¿Cuánto tiempo debe dejar pasar la persona afectada sin recibir una explicación antes de iniciar acciones legales? Desafortunadamente, el legislador europeo no ha tomado como ejemplo el artículo 12 RGPD que obliga al responsable a proporcionar la información requerida sin demora y en el plazo máximo de un mes, prorrogable por causas debidamente justificadas.

Segundo, la disposición no establece qué acciones legales se pueden tomar. Según se ha explicado, el artículo 85 no recoge un verdadero derecho a presentar una reclamación ante la autoridad de vigilancia del mercado. De acudir a esta vía, la reclamación presentada por omisión de explicaciones u explicaciones insatisfactorias será un elemento más que la autoridad tendrá en cuenta para determinar si inicia o no una investigación contra el responsable del despliegue. Alternativamente, la persona afectada puede iniciar una acción ante la jurisdicción civil (en el caso en que el responsable sea una entidad privada) o contencioso-administrativa (si se tratara de un organismo público). Como se ha indicado al analizar el artículo 85, el elevado coste del procedimiento judicial y su dilatación en el tiempo provoca que, en la práctica, únicamente sea una opción viable para entidades de representación de interés colectivos de los consumidores.

4. LA RELACIÓN ENTRE EL DERECHO A UNA EXPLICACIÓN DEL ARTÍCULO 86 Y EL RGPD

Como se ha explicado en el epígrafe anterior, de acuerdo con el artículo 86.3, el derecho a una explicación resulta aplicable «únicamente en la medida en que el derecho [...] no esté previsto de otro modo en el Derecho de la Unión».

Resulta necesario analizar si esta exclusión resulta aplicable al RGPD por cuanto no resulta claro si en él se recoge un «derecho a una explicación» o no. En caso, afirmativo, el RGPD resultaría aplicable con carácter preferente cuando la persona o personas afectadas que reclaman una explicación fueran personas físicas. En tal caso, el derecho a una explicación del artículo 86 RIA perdería parte de su utilidad por cuanto únicamente se beneficiarían de él las personas jurídicas. En cambio, si se llegara a la conclusión de que el RGPD no establece un derecho a una explicación, la

38. GRUPO DE TRABAJO ART. 29, *Directrices ...*, *op. cit.*, p. 28.

utilidad del artículo 86 RIA sería mucho mayor no sólo porque también beneficiaría a las personas físicas sino porque reforzaría los derechos que estas personas tienen en el artículo 22 RGPD en relación con las decisiones automatizadas.

La cuestión sobre la existencia de un derecho a una explicación en el RGPD ha sido ampliamente discutida por la doctrina³⁹. Se afirma, acertadamente, que se trata de un concepto multiforme. Por un lado, está referido a las explicaciones que el interesado tiene derecho a recibir sobre el funcionamiento del sistema (es decir, la lógica, el significado y las consecuencias que se derivan del mismo), o sobre la justificación, las razones o las circunstancias individuales que llevaron a adoptar una decisión determinada. Por otro, el derecho puede ejercerse antes de que se haya tomado la decisión automatizada (*ex ante*); o después (*ex post*)⁴⁰.

El derecho a una explicación *ex ante* está adecuadamente regulado en el RGPD. De acuerdo con su artículo 5.1, el responsable del tratamiento (titular de los datos) tiene la obligación de tratar los datos de manera lícita, leal y transparente. Cuando dichos datos son utilizados para la adopción de decisiones automatizadas (incluida la elaboración de perfiles), ello implica la obligación del responsable de ofrecer «información significativa sobre la lógica aplicada, así como la importancia y las consecuencias previstas de dicho tratamiento para el interesado» (artículo 15.1 h).

Ahora bien, ¿qué ocurre cuando, a pesar de cumplir con estas obligaciones, un sistema informático adopta una decisión automatizada presuntamente errónea? ¿Tiene el interesado un derecho a una explicación *ex post*? Para algunos, las obligaciones de información previamente citadas y el artículo 22.3 introducen este derecho⁴¹. En particular, esta última disposición otorga al interesado un derecho «a expresar su punto de vista y a impugnar la decisión». No obstante, ninguna de estas disposiciones recoge expresamente el derecho a una explicación, cuya referencia únicamente puede encontrarse en el considerando 71, en el cual se habla del «derecho a obtener intervención humana, a expresar su punto de vista, a recibir una explicación de la decisión tomada después de tal evaluación y a impugnar la decisión».

La falta de refrendo legal y el carácter no vinculante de los considerandos, ha llevado a algunos autores a interpretar que no existe un derecho a una explicación en el RGPD⁴². Por consiguiente, volviendo al RIA, podría entenderse que la exclusión del artículo 86.3 no resultará aplicable al RGPD, por lo que las personas físicas podrían beneficiarse del derecho a una explicación previsto en el apartado 1, en los términos

39. VILASAU I SOLANA, M. (2020), «La realización de perfiles y la salvaguardia de los derechos y libertades del afectado», en A. Cerrillo i Martínez y M. Peguera Poch, *Retos jurídicos de la inteligencia artificial*, Madrid, Aranzadi, 2020, pp. 181 ss.

40. WACHTER, S., MITTELSTADT, B., y FLORIDI, L., «Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation», *International Data Privacy Law*, Vol. 7, N.º 2, 2017, disponible en <https://ssrn.com/abstract=2903469>

41. Entre otros, GOODMAN, B. y FLAXMAN, S., «EU Regulations on Algorithmic Decision-Making and a «Right to Explanation», *AI Magazine*, vol 38, num. 3, 2017, disponible en 10.1609/aimag.v38i3.2741; MALGIERI, G. / COMANDÉ, G., «Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation», *International Data Privacy Law*, vol. 7, Issue 3, 2017, disponible en <https://ssrn.com/abstract=3088976>

42. WACHTER, S., MITTELSTADT, B., y FLORIDI, L., *op. cit.*, p. 6.

analizados anteriormente. De sostenerse esta interpretación, la protección de las personas físicas frente a las decisiones individuales automatizadas se vería reforzada por cuanto el RIA les otorga un derecho que no parece estar contemplado en el RGPD.

Ahora bien, la literalidad del artículo 86.3 permite llegar a otra conclusión. Y es que la disposición no excluye la aplicación del apartado primero cuando el derecho esté previsto en otro instrumento de Derecho de la Unión, sino cuando el derecho *esté previsto «de otro modo»* en el Derecho de la Unión. A mi modo de ver, se pueden interpretar que el derecho a una explicación está previsto «de otro modo» en el RGPD cuando la decisión automatizada se refiere a datos personales. De ser esta la interpretación finalmente adoptada, la utilidad práctica del artículo 86.1 se limita extraordinariamente por cuanto sólo podrían ser beneficiarios las personas jurídicas. Los trabajos preparatorios también apuntan a que la intención del legislador era la de excluir este derecho en aquellos supuestos en los que resulta aplicable el RGPD. Ello explica que, la versión del artículo introducida por el Parlamento no se refiere al Derecho de la Unión, sino a «*los artículos 13, 14, 15 y 22 del Reglamento (UE) 2016/679*».

5. LÍMITES AL DERECHO A UNA EXPLICACIÓN: LOS DERECHOS DE PROPIEDAD INTELECTUAL Y LA INFORMACIÓN CONFIDENCIAL

Como es conocido, los sistemas de IA son susceptibles de protección por distintas categorías de propiedad intelectual. En particular, los modelos constituyen programas de ordenador susceptibles de protección por derechos de autor o patentes⁴³; los pesos y parámetros utilizados para entrenar al sistema son susceptibles de protección como bases de datos o secreto empresarial⁴⁴; y los diferentes conjuntos de datos que se pueden utilizar para entrenar el modelo pueden ser protegidos, en sí mismos, por derechos de autor o derechos conexos, o en su conjunto como bases de datos o, en el peor de los casos y siempre que se garantice su confidencialidad, como secretos comerciales⁴⁵. Lo mismo ocurre con los resultados, los cuales podrán ser susceptibles de protección por derechos de autor o conexos o como secreto comercial. Los beneficiarios de esta protección pueden ser el proveedor del sistema IA, el responsable del despliegue o terceras personas.

Debe recordarse que la exclusividad que estas entidades obtienen sobre ese software, los pesos y parámetros de entrenamiento, y sobre los datos les otorga una ventaja competitiva en el mercado que es merecedora de protección por el ordenamiento jurídico.

Ante esta circunstancia, cuando se ejercita el derecho a una explicación, se plantea un conflicto de intereses: los de los responsables del despliegue por ofrecer la menor cantidad de información posible para así preservar los derechos de propiedad intelectual y la confidencialidad de los datos del sistema de IA; y los de la persona

43. MUÑOZ FERRANDIS, C. / DUQUE LIZARRALDE, M., «Open Sourcing AI: Intellectual Property at the Service of Platform Leadership» (January 26, 2022). Available at SSRN: <https://ssrn.com/abstract=4018413>

44. SOUSA E SILVA, N, «Are AI models weights protected databases?», 18 enero 2024, *Kluwer Copyright Blog*, disponible en <https://copyrightblog.kluweriplaw.com/2024/01/18/are-ai-models-weights-protected-databases/>

45. Extensamente, en LOPEZ-TARRUELLA MARTINEZ, A, *Propiedad intelectual e innovación basada en los datos*, Madrid, Dykinson, 2021.

afectada por obtener una explicación lo más detallada posible sobre la decisión que el sistema de IA ha adoptado sobre ella. A mi modo de ver, la obtención de una explicación lo más detallada posible no sólo interesa a la persona afectada sino a la sociedad en general por cuanto, gracias a ello, se favorece la detección de errores en los sistemas de IA que, al fin y al cabo, conllevan un beneficio general. Ante este conflicto, cabe preguntarse: ¿está el responsable del despliegue obligado a desvelar información susceptible de protección por derechos de propiedad intelectual o secreto empresarial cuando ello sea necesario para ofrecer una explicación clara y significativa acerca del papel que el sistema de IA ha tenido en el proceso de toma de decisiones automatizadas?

Ni el artículo 86 RIA ni los considerandos relacionados incluyen alguna precisión de cómo responder a esta pregunta. Ahora bien, el análisis de otras disposiciones del Reglamento que se refieren a la cuestión del tratamiento de los derechos de propiedad intelectual y la información confidencial permite llegar a una respuesta negativa.

Así, en todas las disposiciones en las que se obliga al proveedor u otro participante en la cadena de valor a ofrecer información sobre el sistema de IA, se indica que dicha obligación, se establece: «sin perjuicio de la necesidad de observar y proteger los derechos de propiedad intelectual e industrial, la información empresarial confidencial y los secretos comerciales, de conformidad con el Derecho nacional y de la Unión» (artículo 25.5, relativo a las obligaciones de los operadores económicos que participan en la cadena de valor de la IA, incluyendo a los responsables del despliegue; artículo 53.1 b), relativo a las obligaciones de los proveedores de modelos de IA de uso general).

La misma obligación recae sobre la Comisión a la hora de publicar la lista de modelos de IA de uso general con riesgo sistémico a la que se refiere el artículo 52.6.

En fin, con carácter general, el artículo 78 obliga a «las autoridades de vigilancia del mercado, los organismos notificados y cualquier otra persona física o jurídica que participe en la aplicación del presente Reglamento, de conformidad con el Derecho de la Unión y nacional, respetarán la confidencialidad de la información y los datos obtenidos en el ejercicio de sus funciones y actividades de modo que se protejan, en particular: a) los derechos de propiedad intelectual e industrial y la información empresarial confidencial o los secretos comerciales de una persona física o jurídica». Asimismo, en relación con las autoridades, el apartado 2 indica que «solo solicitarán los datos que sean estrictamente necesarios para la evaluación del riesgo que presentan los sistemas de IA y para el ejercicio de sus competencias en cumplimiento del presente Reglamento y del Reglamento 2019/1020».

Por lo tanto, si bien el artículo 86 no lo establece expresamente, del conjunto de disposiciones analizadas parece extraerse que el responsable del despliegue no tiene obligación de proporcionar información que puedan considerarse un secreto comercial o que esté protegida por derechos de propiedad intelectual e industrial, a la hora de ofrecer explicaciones sobre la decisión adoptada a partir de los resultados proporcionada por el sistema de IA. Y, en el caso de que el responsable del despliegue considere necesario proporcionarla, el artículo 78.1 obliga a la persona afectada a respetar la confidencialidad de la información y los datos obtenidos.

Ahora bien, hubiera sido adecuado la introducción de una precisión similar a la que contiene el considerando 63 RGPD: la necesidad de preservar sus derechos

de propiedad intelectual y el secreto de la información confidencial no puede tener como resultado el rechazo de la solicitud de explicaciones. Es decir, la solicitud debe ser atendida si bien proporcionando únicamente información que no perjudique los intereses del responsable del despliegue o a terceros.

IV. EL PAPEL DE DENUNCIANTES Y ASOCIACIONES DE INTERESES COLECTIVOS

Según se ha indicado en la introducción, la regulación del derecho a presentar una reclamación ante una autoridad de vigilancia del mercado del artículo 85, y del derecho a explicación de decisiones tomadas individualmente del artículo 86, se complementa con los artículos 87 y 110. El primero contiene una remisión a la Directiva 2019/1937 sobre la protección de los denunciantes. El segundo modifica el Anexo de la Directiva 2020/1828 para garantizar que las asociaciones de representación de los intereses colectivos de los consumidores pueden iniciar acciones por incumplimiento del Reglamento. Ambas disposiciones están, por tanto, referidas a la legitimación para ejercer estos derechos. Por motivos expositivos, se ha preferido tratar ambas disposiciones manera autónoma en este apartado.

En vista de los beneficios que se derivan para el interés público, la Directiva 2019/1937 otorga una protección a las personas que denuncien infracciones de la normativa europea. Estas denuncias permiten detectar, investigar y enjuiciar de manera efectiva esas infracciones, mejorando así la transparencia y la rendición de cuentas⁴⁶.

Si bien el RIA puede considerarse uno de los instrumentos que entra en el ámbito de aplicación de la Directiva por la referencia que su artículo 1.1 c) hace a las «infracciones relativas al mercado interior», el legislador europeo ha preferido despejar cualquier duda al respecto con una referencia expresa en el artículo 87 RIA:

La Directiva (UE) 2019/1937 se aplicará a la denuncia de infracciones del presente Reglamento y a la protección de las personas que denuncien tales infracciones.

En particular, la Directiva obliga a los Estados miembros a adoptar las medidas necesarias para garantizar que no se desvele la identidad del denunciante sin su consentimiento expreso (artículo 16), y para prohibir todas las formas de represalias contra estas personas (artículo 19 y 21). Asimismo, se deben habilitar medidas de apoyo para el denunciante, como ofrecimiento de información y asesoramiento completo e independiente, asistencia efectiva de las autoridades frente a represalias y asistencia jurídica (artículo 20). También se deben establecer medidas de apoyo para otras personas afectadas por su relación con el denunciante (artículo 22).

Esta protección del denunciante puede resultar de gran ayuda para favorecer el cumplimiento del RIA. En este sentido, cabe recordar que, en el pasado, el público y las autoridades tuvieron conocimiento de flagrantes infracciones de la normativa aplicable por parte de empresas tecnológicas gracias a las revelaciones llevadas a cabo por trabajadores o colaboradores de esas entidades. Tal fue el caso, por ejemplo, del escándalo de *Cambridge Analytica*.

46. Considerando 1.

Del mismo modo, el considerando 8 de la Directiva nos recuerda, en relación con la seguridad de los productos comercializados en el mercado interior (como puede ser el caso de productos que incorporan sistemas de IA), que «las empresas que operan en las cadenas de fabricación y distribución son la principal fuente de pruebas, de modo que la información de los denunciantes en esas empresas tiene un alto valor añadido ya que están mucho más cerca de la información sobre posibles prácticas abusivas e ilícitas de fabricación, importación o distribución relativas a productos inseguros. En consecuencia, existe una necesidad de que se introduzca la protección de los denunciantes en relación con los requisitos de seguridad aplicables a los productos regulados por la legislación de armonización de la Unión, tal como se establece en los anexos I y II del Reglamento (UE) 2019/1020».

Este considerando también indica que serán las autoridades de vigilancia del mercado o las autoridades judiciales las obligadas a garantizar al denunciante de una infracción del RIA la protección que les ofrece la normativa nacional de desarrollo de esta Directiva.

Por su parte, la inclusión del RIA en el Anexo de la Directiva 2020/1828 habilita a los organismos de protección de los intereses colectivos de los consumidores (denominadas «entidades habilitadas») a ejercer acciones de representación frente a actos de empresarios que infrinjan las disposiciones del RIA. Con ello se refuerza la confianza de los consumidores y los capacitaría para ejercitar sus derechos, se contribuye a una competencia más leal y a crear unas condiciones de competencia equitativas para los empresarios que ejercen su actividad en el mercado interior. En el entorno del RIA, la existencia de estos mecanismos, deben incentivar a proveedores y otros operadores de la cadena de valor de la IA a cumplir adecuadamente con los requisitos y obligaciones del Reglamento.

De acuerdo con la Directiva, los Estados miembros deben garantizar que los consumidores disponen, a escala de la Unión y nacional, de al menos un mecanismo procesal efectivo y eficiente de acciones de representación para obtener medidas de cesación y medidas resarcitorias. La referencia a «consumidores» implica que los Estados miembros no están obligados a garantizar estos mecanismos cuando las infracciones de la normativa europea perjudican a personas físicas o jurídicas consideradas empresarios.

En el caso del RIA, al contrario que en el resto de leyes digitales europeas, ese mecanismo procesal efectivo pasa necesariamente por el ejercicio de acciones judiciales pues, como se ha explicado en el primer epígrafe, no existe un verdadero derecho a presentar una reclamación ante las autoridades de vigilancia del mercado. Bien es cierto que, en estos casos, el artículo 9 Reglamento 2019/1020 establece la facultad de las autoridades de vigilancia del mercado de acordar con «organizaciones que representen a operadores económicos o a usuarios finales», la realización de actividades conjuntas con objeto de incentivar el cumplimiento o detectar casos de incumplimiento. En el mismo sentido, en relación con las acciones judiciales, la Directiva garantiza que las entidades habilitadas en un Estado miembro puedan ejercitar acciones de representación en otro Estado miembro, y deben poder unir fuerzas para presentar una única acción ante un único foro.

Debe señalarse la efectividad que la aplicación de la Directiva puede tener para favorecer el cumplimiento efectivo del RIA. Basta con recordar la efectividad que

ha tenido las acciones iniciadas por las entidades *NOYB* o *La Quadrature du Net* para favorecer el cumplimiento del RGPD; y las acciones iniciadas recientemente en los Países Bajos, al amparo de la *WAMCA*⁴⁷, contra gigantes tecnológicos como Apple, Google o Tik Tok en la que se solicitan la adopción de medidas de cesación por incumplimiento del Reglamento de servicios digitales, y de la normativa *antitrust* europea⁴⁸.

V. CONCLUSIONES

La Sección 4 («Vías de recurso») del Capítulo IX («Vigilancia poscomercialización, intercambio de información, vigilancia del mercado») del RIA se introdujo, en fase de tramitación legislativa por el Parlamento europeo, con el fin de garantizar a los particulares ciertos de derechos cuando se vieran afectados por la utilización de sistemas de IA. El derecho a presentar una reclamación ante una autoridad de vigilancia del mercado y el derecho a explicación de decisiones tomadas individualmente deben, por ello, ser bienvenidos. No obstante, su redacción definitiva ha vaciado sustancialmente el contenido que inicialmente le atribuía el Parlamento.

Así, por un lado, el primero no regula un derecho a presentar una reclamación en el mismo sentido que el RGPD u otras leyes digitales europeas. Más bien, lo que regula es un derecho a presentar una petición ante la autoridad de vigilancia del mercado, la cual lo tendrá en cuenta a la hora de determinar si iniciar operaciones de investigación contra el proveedor o responsable de despliegue el sistema de IA. Además, la presentación de estas peticiones resulta obstaculizado por la compleja distribución de competencias entre las autoridades de vigilancia del mercado que se deriva del artículo 74. Si bien, esta disposición permite que, hasta cierto punto, la competencia para conocer de estas reclamaciones se concentre en la AESIA, esta solución no está exenta de problemas.

Por otro lado, en relación con el segundo, cabe señalar que el derecho solo se disfruta en relación con ciertas categorías de sistemas IA de alto riesgo: aquellos enumerados en el Anexo III (con excepción del punto 2). Además, si nuestra interpretación de la relación del artículo 86.3 RIA con el RGPD es correcta, este derecho solo lo pueden ejercer las personas jurídicas. En el caso de las personas físicas, en la medida en que la toma de decisión automatizada que les afecta debería haberse llevado a cabo necesariamente a partir de datos que permiten identificarlas, la obtención de explicaciones deberá ejercerse a partir de las vías habilitadas por el RGPD.

En definitiva, nos encontramos ante bienintencionada regulación de las vías de recurso en los artículos 85 a 87 que, debido a las negociaciones de última hora en el trílogo, ha perdido gran parte de su efecto útil en perjuicio de los intereses de los particulares, y de los objetivos enunciados en el artículo 1 del Reglamento.

47. Ley de resolución de acciones colectivas (*Wet collectieve afwikkeling massaschade*).

48. X. KRAMER, «International tech litigation reaches the next level: collective actions against TikTok and Google», *Conflict of laws*, 12 marzo 2024, disponible en <https://conflictoflaws.net/2024/international-tech-litigation-reaches-the-next-level-collective-actions-against-tiktok-and-google/>

Acceso a documentación y confidencialidad en el Reglamento de inteligencia artificial

GABRIELE VESTRI

Doctor en Derecho, Fundador y Presidente del Observatorio Sector Público e Inteligencia Artificial

I. INTRODUCCIÓN

Como sugiere el título de esta contribución, nuestro propósito es analizar dos de los pilares esenciales del RIA. Esta normativa representa el último avance realizado por los poderes de la Unión Europea, bajo la dirección de la representación española el pasado mes de diciembre de 2023. Un reglamento que finalmente responde o intenta responder a lo que Salazar García ha denominado «shock tecnológico» que de alguna manera une los avances tecnológicos con el miedo que estos producen¹. Precisamente para dar respuesta a este cambio, el RIA crea un andamio de normas, a veces complejo que, desde una perspectiva horizontal, no se limita a sectores concretos sino que pretende paliar los efectos dañinos de la inteligencia artificial².

Así, en esta aportación nos referimos específicamente a las disposiciones del RIA relacionadas con el acceso a la documentación y la confidencialidad. En este contexto, nuestro objetivo es desentrañar las complejidades legales, administrativas y prácticas que rodean determinadas cuestiones críticas en el desarrollo normativo de la inteligencia artificial en la UE. Nos centraremos especialmente en los artículos 77 y 78 del RIA, aunque debemos advertir que estas normas plantean ciertas ramificaciones que intentaremos abordar y analizar.

En este escenario, es necesario introducir algunos conceptos que nos ayudarán a comprender y analizar el alcance de las cuestiones tratadas, las cuales, como es de imaginar, no carecen de complicaciones. Además, cabe destacar que el análisis que proponemos asume las definiciones presentes en el RIA, remitiéndose a ellas sin repetir las en este contexto.

1. Salazar García, I. «Privacidad e inteligencia artificial: ¿es posible su convivencia?» en Arellano Toledo, W. (Directora), en *«Derecho, Ética e Inteligencia Artificial»*, Tirant lo Blanch, (2023), p. 181.
2. En este sentido Véase: Barrio Andrés, M. «Inteligencia artificial, Internet de las cosas y blockchain» en Montero Pascual, J.J. (Coordinador), *«Digitalización y derecho. Curso de Derecho digital»*, Tirant lo Blanch, (2024), p. 266.

Resulta útil, esto sí, realizar una aproximación universal a ciertas nociones que, como mínimo, nos permitan trazar la senda a seguir en nuestra contribución.

Al abordar el acceso a la documentación y la confidencialidad, estamos tratando los criterios propios de la transparencia. El RIA se fundamenta en el principio imprescindible de transparencia. Para garantizar que los usuarios, las autoridades y también los ciudadanos comprendan plenamente el impacto y el funcionamiento de los sistemas de inteligencia artificial, se impone la obligación de proporcionar documentación detallada —transparencia que, con un enfoque más general, el RIA trata en su artículo 13 del que recomendamos su lectura—. Bien, esta documentación no solo debe ser clara y comprensible, sino también accesible, reflejando un compromiso con la participación informada de todas las partes interesadas. Precisamente en el ámbito del Reglamento UE y como bien señala Cotino Hueso, debemos tener en cuenta la tipología de información para que «el usuario o consumidor del sistema (y los técnicos que lo implementan), puedan manejar el sistema de IA correctamente, cumplir sus obligaciones y supervisarlos. En este punto se ha de tener en cuenta la naturaleza diferente de usuarios, importadores y distribuidores»³. El tema mencionado no es baladí, y es nuestra obligación señalar que, aunque existen ciertas diferencias en la aproximación, algunos Estados miembros ya han esgrimido argumentos en los cuales el acceso a la información ha desempeñado un papel central. Basta con recordar el caso español denominado Bono social-Fundación Civio, que se centró precisamente en la denegación de acceso a la información del sistema Bosco⁴. En el mismo sentido, diversas sentencias de tribunales italianos han destacado la necesidad del acceso a la información y de la comprensibilidad inherente⁵.

Este enfoque conduce a lo que en las legislaciones nacionales conocemos como el derecho de acceso. Un aspecto destacado es el reconocimiento del derecho de los usuarios a acceder a la documentación pertinente. Esto no solo fortalece la posición de los usuarios en un mundo cada vez más impulsado por la inteligencia artificial, sino que también fomenta la rendición de cuentas de los proveedores de estos sistemas. La transparencia —a través del acceso—, se convierte en un medio para empoderar y garantizar la autonomía informada⁶.

Naturalmente, dicho acceso deberá contemplar ciertas limitaciones. En aras de un enfoque equilibrado, el RIA también establece limitaciones a la divulgación completa de información, reconociendo que ciertos detalles pueden comprometer la seguridad pública, la privacidad o los derechos de propiedad intelectual. Las excepciones y

3. Cotino Hueso, L. «Transparencia y explicabilidad de la inteligencia artificial y “compañía” (comunicación, interpretabilidad, inteligibilidad, auditabilidad, testabilidad, comprobabilidad, simulabilidad...). Para qué, para quién y cuánta» en Cotino Hueso, L. Claramunt Castellanos, J. (Coordinadores). «*Transparencia y explicabilidad de la inteligencia artificial*». Tirant lo Blanch, (2022), p. 46.
4. Véase Vestri, G. «El acceso a la información algorítmica a partir del caso bono social vs. Fundación ciudadana Civio» en *Revista General de Derecho Administrativo*, n.º 61, (2022), pp. 1-24.
5. Para tener una visión sobre la orientación de la jurisprudencia italiana Véase: Vestri, G. «Sistemi algoritmici e principio di buona amministrazione algoritmica» en *Rivista Diritto di internet*, n.º 2, (2023), pp. 373-382.
6. Sobre transparencia algorítmica Véase: Vestri, G. «La inteligencia artificial ante al desafío de la transparencia algorítmica. Una aproximación desde la perspectiva jurídico-administrativa». *Revista Aragonesa de Administración pública*, n.º 56, (2021), pp. 368-398.

limitaciones buscan salvaguardar otros valores fundamentales sin despojar por completo a la regulación de su carácter transparente.

Asimismo, el RIA se ocupa de la confidencialidad que, en nuestra opinión, debe además considerarse en relación con la protección de datos. En efecto, en el ámbito de la confidencialidad, el RIA tiene en cuenta la delicada cuestión de la protección de los datos. Los desarrolladores de sistemas de inteligencia artificial manejan información estratégica, comercial y de investigación, lo que hace imperativo garantizar su confidencialidad para no comprometer la competitividad y la innovación. En este mismo contexto, es de suma importancia proceder con la correspondiente evaluación de la conformidad de las distintas formas de confidencialidad. Las evaluaciones de conformidad, piedra angular del marco regulatorio, también se someten al prisma de la confidencialidad. Este enfoque busca preservar la propiedad intelectual y los secretos comerciales asociados con los procesos de evaluación, a la vez que se asegura la integridad del sistema de regulación.

De crucial importancia es también la aproximación a la seguridad nacional y las limitaciones estratégicas. En este sentido, y reconociendo la necesidad de proteger la seguridad nacional, el RIA incorpora disposiciones que permiten limitar la divulgación de información que podría poner en peligro la seguridad del Estado miembro y de la Unión. Este matiz resalta la conciencia de la Unión Europea sobre la necesidad de equilibrar la innovación tecnológica con la seguridad nacional.

Para concluir esta parte introductoria, y como probablemente ya se ha entendido, el RIA de la Unión Europea se presenta como un marco normativo ambicioso, sustentado en principios sólidos de acceso (y por ende transparencia) y confidencialidad. Este enfoque cauteloso y equilibrado refleja el compromiso de la UE con la ética y la responsabilidad en el desarrollo de la inteligencia artificial. A medida que avanzamos en el estudio de los efectos de la inteligencia artificial, es imperativo seguir escudriñando estas disposiciones, evaluando su implementación y adaptándolas a un entorno tecnológico en constante evolución. La convergencia de la tecnología y el derecho exige una vigilancia continua y una reflexión crítica, y en este sentido, nos encontramos en el epicentro de un fascinante y desafiante terreno jurídico. Este análisis se realizará siempre desde un enfoque crítico, como intentaremos desglosar en esta contribución y a sabiendas de que la norma europea, al margen de su importancia estratégica, hubiera podido ser aún más ambiciosa de los que realmente es. La tendencia del RIA es desarrollar un ecosistema de excelencia y asimismo crear un ecosistema de confianza⁷. Quizá, solo el tiempo nos permitirá evaluar el impacto del RIA.

II. ANÁLISIS DEL CONTENIDO DEL ARTÍCULO 77 DEL REGLAMENTO

Es importante señalar que el artículo 77 del RIA se enmarca en lo que la misma norma establece como: «Poderes de las autoridades que protegen los Derechos Fundamentales». Ahora bien, aunque el título de la norma es sin duda elocuente, es quizá correcto señalar que la disposición legal en cuestión se configura como una norma que otorga facultades específicas a la autoridad nacional de supervisión en relación con conjuntos de datos empleados en actividades vinculadas a la inteligencia artificial o sistemas automatizados.

7. En sentido, Véase: Muñoz García, C. «Regulación de la inteligencia artificial en Europa. Incidencia en los regímenes jurídicos de protección de datos y de responsabilidad por productos» Tirant lo Blanch, (2023), p. 36.

Dicho esto, sin duda, el acceso a la documentación, el acceso al algoritmo de inteligencia artificial, guardan una relación muy estrecha con los Derechos fundamentales por lo menos en la perspectiva de que un sistema de inteligencia artificial plantea riesgos individuales y sociales que precisamente pueden poner en peligro los Derechos fundamentales⁸.

De conformidad con el texto, la autoridad nacional de supervisión, en el ejercicio de sus atribuciones y mediante la presentación de una solicitud debidamente fundamentada, ostenta el derecho de obtener acceso integral a los conjuntos de datos utilizados en las etapas de entrenamiento, validación y prueba por parte del proveedor o, en su caso, del implementador. Este otorgamiento de acceso se encuentra circunscrito a aquellos conjuntos de datos que resulten pertinentes y estrictamente necesarios para los propósitos que motivaron la solicitud de acceso. La implementación de dicho acceso debe llevarse a cabo empleando medios técnicos y herramientas apropiadas que finalmente permitan un acceso estructurado y proactivo.

Es imperativo resaltar que esta disposición normativa se orienta a asegurar la transparencia y la supervisión efectiva de las actividades asociadas con la inteligencia artificial, reconociendo la importancia de los conjuntos de datos en la evaluación y control de los sistemas automatizados. La presentación fundamentada de la solicitud y la limitación a la pertinencia y necesidad estricta de los datos buscan equilibrar la autoridad de supervisión con la protección de la confidencialidad y demás derechos legítimos de los proveedores o implementadores. Todo esto y parafraseando a Cotino Hueso, se trata de información valiosa que está estrechamente vinculada al principio de proporcionalidad, sirviendo como una alternativa tanto en general para el poder público como, en particular, cuando existen restricciones o impactos en derechos fundamentales⁹.

Asimismo, y en una perspectiva más amplia, también la Agencia Española de Protección de Datos se ha pronunciado, naturalmente en el ámbito de su materia, sobre el impacto del RIA sobre la cuestión aquí tratada. En este sentido señala que: «cuando los sistemas de IA se incluyen en, o son medios de, un tratamiento de datos personales los responsables del tratamiento deben obtener información sobre ellos suficiente para cumplir sus diferentes obligaciones de cumplimiento RGPD. Estas incluyen la transparencia para permitir el ejercicio de los derechos, cumplir el principio de responsabilidad activa, cumplir los requisitos de las Autoridades de Supervisión del RGPD en relación con sus poderes de investigación, y lo mismo para los organismos de certificación y supervisión del código de conducta»¹⁰.

Posteriormente, la norma establece un marco legal detallado para la autoridad nacional de supervisión en el contexto de sistemas de inteligencia artificial de alto riesgo. Se destaca en la norma que, en situaciones necesarias y previa presentación de una solicitud debidamente fundamentada, la autoridad nacional de supervisión tiene el

8. En este sentido Véase: Presno Linera, M.Á. «Derechos fundamentales e inteligencia artificial». Marcial Pons, (2022), pp. 23-24.
9. Véase Cotino Hueso, L. «Qué concreta transparencia e información de algoritmos e inteligencia artificial es la debida» en *Revista Española de la Transparencia*, n.º 16 primer semestre enero-junio (2023), p. 30.
10. Agencia Española de Protección de Datos, «Inteligencia artificial: transparencia», en <https://www.aepd.es/prensa-y-comunicacion/blog/inteligencia-artificial-transparencia> [Consultado el 28 de diciembre de 2023]

derecho de acceder al modelo entrenado y al modelo de entrenamiento de un sistema de inteligencia artificial, así como a los parámetros relevantes de estos modelos. Este acceso se concede después de agotar y demostrar la insuficiencia de todas las demás vías razonables para verificar la conformidad del sistema de inteligencia artificial de alto riesgo, incluyendo las contempladas en el apartado anterior. La evaluación de conformidad tiene como finalidad asegurar que el sistema de inteligencia artificial cumple con los requisitos pre-establecidos.

Es de suma importancia señalar que toda la información obtenida durante este procedimiento y conforme al artículo 78, se considera como información confidencial. Dicha información está sujeta a la normativa de la Unión Europea en materia de protección de la propiedad intelectual y de los secretos comerciales. Asimismo, se especifica que esta información será eliminada una vez concluida la investigación para la cual fue solicitada.

La introducción del apartado 2 bis enfatiza que los derechos procedimentales del operador, de acuerdo con el artículo 18 del Reglamento (UE) 2019/1020, no se ven afectados por las disposiciones anteriores. En otras palabras, se garantiza que el operador del sistema de inteligencia artificial de alto riesgo conserve sus derechos procesales durante el proceso de evaluación de conformidad llevado a cabo por la autoridad nacional de supervisión.

El texto dispone, adicionalmente, que las autoridades u organismos públicos nacionales investidos con la responsabilidad de supervisar la observancia de las obligaciones emanadas del Derecho de la Unión, especialmente aquellas vinculadas a derechos fundamentales y la no discriminación en el uso de sistemas de inteligencia artificial de alto riesgo, ostentan la facultad de requerir y obtener acceso a cualquier documentación generada o conservada en virtud del reglamento mismo. Este otorgamiento de acceso se materializa siempre que resulte indispensable para la ejecución de las competencias que les corresponden dentro de los límites de su competencia territorial.

Es imperioso subrayar que el acceso a la documentación debe efectuarse en una lengua y formato accesibles y, por consiguiente, comprensibles. Además, al formular una solicitud de esta índole, la autoridad u organismo público pertinente está obligado a comunicar a la autoridad de supervisión del mercado del Estado miembro correspondiente sobre dicha solicitud. Ahora bien, por extensión y como señala Belloso Martín, debemos quizá aspirar a que el algoritmo no solo sea explicable sino también justo y éste es el verdadero desafío¹¹.

En síntesis, el texto analizado persigue conferir facultades a las autoridades nacionales encargadas de la salvaguarda de derechos fundamentales para obtener la documentación pertinente en el ámbito de sistemas de inteligencia artificial de alto riesgo, garantizando, de esta manera, un control y supervisión eficaces en el cumplimiento de las obligaciones derivadas del Derecho de la Unión en esta materia.

Prosigue el artículo analizado estableciendo que en un plazo máximo de tres meses a partir de la entrada en vigor del Reglamento, cada Estado miembro deberá identificar a las autoridades u organismos públicos a los que se hace referencia en el apartado

11. Belloso Martín, N. «Sobre fairness y machine learning: el algoritmo ¿puede (y debe) ser justo?» en *Anales de la Cátedra Francisco Suárez* n.º 57, (2023), p.3. DOI: <https://doi.org/10.30827/acfs.v57i.25250>

3 de la normativa correspondiente¹². La identificación de estas entidades debe ser divulgada mediante la publicación de una lista en el sitio web de la autoridad nacional de supervisión del respectivo Estado miembro. Asimismo, los Estados miembros tienen la obligación de notificar esta lista tanto a la Comisión como a todos los demás Estados miembros y deben mantenerla actualizada.

En términos sencillos, este apartado establece un plazo específico para que cada Estado miembro identifique y publique las autoridades u organismos públicos mencionados en la normativa. La divulgación de esta información en el sitio web de la autoridad nacional de supervisión, junto con la notificación a la Comisión y otros Estados miembros, busca garantizar la transparencia y la comunicación efectiva entre los Estados miembros y la Comisión en el contexto de la implementación y aplicación del Reglamento.

El texto establece un procedimiento legal cuando la documentación disponible, según lo establecido en el apartado 3 de la normativa correspondiente, resulta ser insuficiente para determinar si ha ocurrido un incumplimiento de las obligaciones derivadas del Derecho de la Unión destinadas a proteger los derechos fundamentales en el contexto de sistemas de inteligencia artificial de alto riesgo.

Ahora bien, en situaciones en las cuales la documentación especificada en el apartado 3 no proporciona información suficiente para verificar si ha habido un incumplimiento de las obligaciones derivadas del Derecho de la Unión destinadas a proteger los derechos fundamentales en el ámbito de sistemas de IA de alto riesgo, la autoridad u organismo público mencionado en el mismo apartado tiene la facultad de presentar una solicitud fundamentada a la autoridad de vigilancia del mercado. La solicitud tiene como objetivo la organización de una verificación del sistema de IA de alto riesgo a través de medios técnicos.

Finalmente, la autoridad de vigilancia del mercado, a su vez, llevará a cabo las pruebas necesarias con la estrecha participación de la autoridad u organismo público solicitante. Este proceso debe realizarse en un plazo razonable después de recibir la solicitud. En esencia, este mecanismo permite a la autoridad u organismo público, cuando la documentación existente es insuficiente, solicitar a la autoridad de vigilancia del mercado la realización de pruebas técnicas para evaluar la conformidad del sistema de IA de alto riesgo con las obligaciones derivadas del Derecho de la Unión relacionadas con los derechos fundamentales.

Comprendiblemente, el acceso a la documentación parece convertirse en lo que los ordenamientos jurídico-nacionales suelen definir como «información pública» de manera que este criterio habrá de considerarse también en el seno del artículo 77¹³.

III. ANÁLISIS DEL CONTENIDO DEL ARTÍCULO 78 DEL REGLAMENTO

El precepto aludido, es decir, el artículo 78, tiene como objeto la delineación de los parámetros relativos a la confidencialidad, los cuales se estiman meticulosamente

12. Es sabido que España ya cuenta con la Agencia Española de Supervisión de la Inteligencia Artificial (AESIA).

13. Sobre el tema tratado se recomienda: Gutiérrez David, M.E. «Administraciones inteligentes y acceso al código fuente y los algoritmos públicos. Conjurando riesgos de cajas negras decisionales» en *Derecom*, n.º 30. Nueva Época. marzo-septiembre, (2021) pp.159-160.

precisos en su redacción. Por precisión y a modo de aproximación general, cabe señalar que la confidencialidad se refiere a la protección y preservación de la información sensible o confidencial manejada, en este caso, en entornos digitales. En este contexto, la confidencialidad se erige como un pilar fundamental para resguardar datos empresariales, secretos comerciales, información personal y otros activos digitales cruciales para las partes involucradas. De hecho, los sistemas destinados a garantizar la confidencialidad se centran en establecer y hacer cumplir medidas legales y técnicas, tales como acuerdos de confidencialidad, cifrado de datos y políticas de acceso restringido, con el propósito de asegurar que la información confidencial no sea divulgada ni utilizada de manera indebida. La confidencialidad, en este sentido, no solo protege los intereses de las partes involucradas, sino que también contribuye, o al menos intenta hacerlo, a la construcción de la confianza en el entorno digital, promoviendo la innovación y el desarrollo tecnológico de manera segura.

En tal sentido, el primer apartado instaura disposiciones concernientes a la confidencialidad de la información y los datos en el contexto de la ejecución del RIA. El texto prescribe que la Comisión, las autoridades de supervisión del mercado y los organismos notificados, así como cualquier ente de naturaleza física o jurídica involucrado en la implementación del citado Reglamento, están obligados a observar la confidencialidad respecto de la información y los datos que obtengan en el desempeño de sus atribuciones. Este resguardo de la confidencialidad debe ajustarse a las normativas del Derecho de la Unión o nacional.

Especial énfasis se concede a la salvaguardia de los derechos de propiedad intelectual, la información comercial confidencial, los secretos comerciales, incluyendo el código fuente, excepto en aquellos supuestos contemplados en la Directiva 2016/943 sobre la salvaguarda de conocimientos técnicos y la información empresarial no divulgados.

Asimismo, se enumeran diversas finalidades para las cuales se impera la protección de la confidencialidad:

- a) La efectiva implementación del presente Reglamento, especialmente en lo relativo a inspecciones, investigaciones o auditorías.
- b) La consideración de intereses públicos y de seguridad nacional.
- c) La integridad de la información clasificada de acuerdo con el Derecho de la Unión o nacional.

En otras palabras, se instaura una obligación para las autoridades vinculadas con la ejecución del Reglamento de requerir únicamente los datos estrictamente necesarios para evaluar el riesgo suscitado por el sistema de IA y para ejercer sus competencias en consonancia con los Reglamentos pertinentes. Además, se subraya la necesidad de implementar medidas idóneas y eficaces de ciberseguridad con miras a proteger la seguridad y confidencialidad de la información y datos obtenidos. Se impone la obligación de eliminar los datos recabados una vez que cesen de ser necesarios para la finalidad para la cual fueron solicitados, conforme a la legislación nacional o europea aplicable, haciendo expresa referencia al Reglamento 2019/1020.

La norma en cuestión, en este caso el segundo apartado, impone limitaciones a la divulgación de información que ha sido intercambiada de manera confidencial entre las autoridades nacionales competentes y entre estas autoridades y la Comisión, en el

marco de la utilización de sistemas de inteligencia artificial de alto riesgo, específicamente aquellos indicados en los puntos 1, 6 y 7 del anexo III.

Así, sin menoscabo de lo dispuesto en los apartados 1 y 1 bis se implanta una declaración según la cual las disposiciones aludidas no afectarán lo establecido en los apartados 1 y 1 bis del instrumento normativo correspondiente.

La información confidencialmente intercambiada entre las autoridades nacionales competentes y la Comisión: se refiere a datos confidenciales compartidos entre las autoridades nacionales competentes de los Estados miembros y la Comisión Europea. No será divulgada sin consulta previa: Establece la prohibición de divulgar tal información sin realizar una consulta previa a la autoridad nacional competente de origen y al usuario.

Cuando los sistemas de IA de alto riesgo mencionados en los puntos 1, 6 y 7 del anexo III sean empleados por las autoridades policiales, de control de fronteras, de inmigración o de asilo: La restricción se aplica específicamente cuando estos sistemas de inteligencia artificial de alto riesgo son utilizados en contextos vinculados con autoridades policiales, control de fronteras, inmigración o asilo.

Asimismo, cuando tal divulgación pudiera poner en peligro los intereses de la seguridad pública y nacional: se establece el criterio de que la divulgación solo puede evitarse si se considera que esta acción podría poner en peligro los intereses de la seguridad pública y nacional. Este intercambio de información no comprenderá los datos operativos sensibles en relación con las actividades de las autoridades policiales, de control de fronteras, de inmigración o de asilo: delimita la información intercambiada, excluyendo los datos operativos sensibles relacionados con las actividades de las autoridades mencionadas.

El tercer apartado del proyecto de acuerdo instaura una salvaguarda normativa, indicándose que los apartados 1, 1 bis y 2 no incidirán en determinados derechos y obligaciones específicos de la Comisión, los Estados miembros, sus autoridades competentes y los organismos notificados. La no incidencia se refiere particularmente al intercambio de información, la difusión de alertas, la cooperación transfronteriza y las obligaciones de las partes interesadas en el contexto del cumplimiento del Derecho penal de los Estados miembros.

Los apartados 1, [1 bis] y 2: Alude a las secciones 1, 1 bis y 2 de la normativa en cuestión, que contienen disposiciones específicas.

No incidirán en los derechos y obligaciones: establece que estas disposiciones no modificarán ni afectarán los derechos y obligaciones de los sujetos mencionados.

De la Comisión, de los Estados miembros y de sus autoridades competentes, así como de los organismos notificados: detalla los sujetos derechos y obligaciones no se verán afectados, incluyendo la Comisión Europea, los Estados miembros y sus respectivas autoridades competentes, así como los organismos notificados.

Por lo que respecta al intercambio de información y a la difusión de alertas, incluso en el contexto de la cooperación transfronteriza, el apartado en cuestión delimita la no incidencia a situaciones relacionadas con el intercambio de información y la difusión de alertas, especialmente en el contexto de la cooperación entre diferentes jurisdicciones.

Asimismo, se aclara que tampoco se afectarán las obligaciones de los actores involucrados (partes interesadas) en proporcionar información de acuerdo con el Derecho penal de los Estados miembros.

En otras palabras, opera como una cláusula de no incidencia, asegurando que ciertos derechos y obligaciones específicos relacionados con el intercambio de información, difusión de alertas, cooperación transfronteriza, así como las obligaciones bajo el Derecho penal, no se ven alterados por las disposiciones contenidas en los apartados mencionados.

Finalmente, el último apartado del artículo 78 instituye la posibilidad de intercambio de información confidencial entre la Comisión y los Estados miembros de la Unión Europea, así como las autoridades reguladoras de terceros países. La realización de dicho intercambio está condicionada a la necesidad y debe llevarse a cabo de acuerdo con las disposiciones específicas de los acuerdos internacionales y comerciales. Además, se destaca que este intercambio solamente puede efectuarse con aquellas autoridades reguladoras de terceros países con las cuales se hayan celebrado acuerdos bilaterales o multilaterales de confidencialidad que garanticen un nivel adecuado de protección de la información confidencial.

En términos estrictamente técnicos-jurídicos, se plantea un escenario determinado.

La Comisión y los Estados miembros podrán intercambiar: establece la facultad de la Comisión y los Estados miembros para llevar a cabo el intercambio de información.

En caso necesario y de conformidad con las disposiciones pertinentes de los acuerdos internacionales y comerciales: condiciona la realización de dicho intercambio a la necesidad y prescribe que debe ajustarse a las disposiciones específicas de los acuerdos internacionales y comerciales.

Además, se subraya que dicho intercambio solo puede realizarse con aquellas autoridades reguladoras de terceros países con las que se hayan celebrado acuerdos bilaterales o multilaterales de confidencialidad que aseguren un nivel adecuado de protección de la información confidencial: enfatiza que la comunicación de información solo puede llevarse a cabo con autoridades reguladoras de terceros países que cuenten con acuerdos bilaterales o multilaterales de confidencialidad, garantizando así un nivel suficiente de protección para la información confidencial.

IV. CONCLUSIONES

Lo hasta ahora analizado describe dos de las disposiciones legales concernientes a la regulación de la inteligencia artificial en el contexto de la Unión Europea. Las normas objeto de estudio se centran en la autoridad nacional de supervisión y sus facultades específicas para acceder a conjuntos de datos utilizados en actividades vinculadas a la inteligencia artificial, con el propósito de garantizar la transparencia y supervisión efectiva de las actividades asociadas con la inteligencia artificial, reconociendo la importancia de los conjuntos de datos en la evaluación y control de sistemas automatizados.

En primer lugar, se establece el derecho de la autoridad de supervisión a acceder a conjuntos de datos relevantes y estrictamente necesarios para los propósitos de entrenamiento, validación y prueba de sistemas de inteligencia artificial. Esta medida se presenta como un equilibrio entre la autoridad de supervisión y la protección de la confidencialidad y otros derechos legítimos de proveedores e implementadores.

La norma posteriormente se enfoca en sistemas de inteligencia artificial de alto riesgo, otorgando a la autoridad de supervisión el derecho de acceder al modelo entrenado y al modelo de entrenamiento, así como a los parámetros relevantes. La evaluación de conformidad tiene como objetivo asegurar que estos sistemas cumplan con los requisitos

establecidos en el marco legal. La información obtenida durante este proceso se considera confidencial y está sujeta a normativas de propiedad intelectual y secretos comerciales, con la obligación de eliminarla una vez concluida la investigación.

Se destaca la garantía de los derechos procesales del operador durante la evaluación de conformidad. Además, se confiere a las autoridades nacionales la facultad de requerir acceso a la documentación relacionada con el cumplimiento de las obligaciones derivadas del Derecho de la Unión en el ámbito de sistemas de inteligencia artificial de alto riesgo, asegurando un control efectivo.

El texto establece un plazo para que los Estados miembros identifiquen y divulguen las autoridades u organismos públicos responsables de supervisar las obligaciones del Derecho de la Unión en este ámbito. Esta divulgación busca garantizar la transparencia y comunicación efectiva entre los Estados miembros y la Comisión.

El artículo 77 establece parámetros detallados sobre la confidencialidad, destacando la protección de la propiedad intelectual, la información comercial confidencial y los secretos comerciales en el contexto de la ejecución del Reglamento. Se enfatiza la necesidad de requerir solo datos estrictamente necesarios y la implementación de medidas de ciberseguridad. También se establece la limitación a la divulgación de información confidencial intercambiada entre autoridades nacionales y con la Comisión, específicamente en el contexto de sistemas de IA de alto riesgo.

El artículo aborda situaciones en las que la documentación disponible es insuficiente, permitiendo a la autoridad de vigilancia del mercado llevar a cabo pruebas técnicas en colaboración con la autoridad u organismo público solicitante. Este mecanismo asegura una evaluación adecuada cuando la documentación existente no es suficiente.

En resumen, el texto busca equilibrar la supervisión efectiva de la inteligencia artificial con la protección de la confidencialidad y los derechos de los proveedores. Establece un marco legal detallado para sistemas de inteligencia artificial de alto riesgo, garantiza la transparencia y comunicación entre las autoridades y define parámetros claros para la confidencialidad y el intercambio de información.

Las disposiciones examinadas, así como la totalidad del texto del RIA, representan únicamente un primer paso en dirección a la regulación de la inteligencia artificial. Esta circunstancia implica que, de manera comprensible, se requiere aguardar las acciones de los diversos Estados miembros para luego proceder a un análisis, a modo de prueba de impacto, con el objetivo de determinar si la normativa europea, conjuntamente con la nacional, ha logrado instaurar una estructura y un entorno normativo y proactivo en el ámbito de la inteligencia artificial.

Asimismo, en el fondo, y no tan en el fondo, el RIA se asemeja mucho a una especie de tratado comercial. Este dato no necesariamente debe entenderse de manera negativa, sino que debemos comprender la dificultad que puede existir para que sus principios se reflejen directamente en las personas. El RIA establece más bien normas de convivencia comercial entre actores profesionales. Por eso, insistimos en que será crucial la aplicación nacional que los Estados miembros realicen del RIA. Será en este momento cuando las personas, los ciudadanos, podrán ver y sentir cómo las normas regulan su relación con la inteligencia artificial.

ESTUDIOS

El Reglamento Europeo de Inteligencia Artificial de la Unión Europea es una norma de una importancia enorme no sólo para la Unión, sino de referencia para todo el mundo. Va a marcar un antes y un después para el desarrollo y la innovación, que depende en muy buena medida de la inteligencia artificial en este siglo XXI. El Reglamento es una norma muy extensa y extraordinariamente compleja que precisa ser abordada sistemáticamente. Es por ello que, lejos de un comentario al articulado o de un acopio de estudios variados, se ha optado por un Tratado sistemático. Para ello, Lorenzo Cotino y Pere Simón han organizado exhaustivamente los temas relevantes del Reglamento y los han encargado a los mejores especialistas en cada tema, treinta y cinco expertos nacionales e internacionales y académicos de primer nivel. Este tratado es la obra imprescindible y de obligada referencia sobre el Reglamento, una guía tanto teórica como práctica para académicos, legisladores y profesionales del Derecho de la IA, así como proveedores e implementadores.

El precio de esta obra incluye la publicación en formato DÚO sin coste adicional (papel + libro electrónico)

ACCEDE A LA VERSIÓN ELECTRÓNICA SIGUIENDO LAS INDICACIONES DEL INTERIOR DEL LIBRO

ISBN: 978-84-1162-931-7



9 788411 629317