Comentarios al Reglamento Europeo de Inteligencia Artificial

Director

Moisés Barrio Andrés

LA LEY

LA LEY Soluciones Legales, 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

Edición: octubre 2024

Depósito Legal: M-21762-2024
ISBN versión impresa: 978-84-18662-88-1
ISBN versión electrónica: 978-84-18662-89-8

Diseño, Preimpresión e Impresión: LA LEY Soluciones Legales, S.A.U.
Printed in Spain

ÍNDICE SISTEMÁTICO

de decisiones públicas, entre otras funciones. Sin embargo, este necesario avance tecnológico para la mejora del servicio público puede entrar, pues, en tensión con la seguridad jurídica, derechos fundamentales como la igualdad, la intimidad o la protección de datos personales; o con principios u obligaciones de actuación administrativa como la obligación de motivación, la transparencia o la eficacia del derecho de defensa o recurso, pues no siempre será posible conocer por qué de los resultados proporcionados por algunas inteligencias artificiales» (Capdeferro, 2020: 1; Cerrillo, 2019; Cerrillo, 2023: 525)

Ahora bien, los resultados de la IA dependen de su uso y de los datos utilizados. Existe la posibilidad de sesgar, intencional o involuntariamente, tanto el diseño como los datos objeto de tratamiento. Como recuerda Menéndez (2021: 35), las discriminaciones y sesgos que muchas veces presiden las decisiones y acciones humanas se proyectan también en la red a través del *software* y de sistemas de inteligencia artificial. Aunque, de entrada, podría pensarse que el espacio digital conlleva la ventaja de ser un espacio neutro y aportar objetividad, de manera que promueva relaciones igualitarias y equitativas, alejado de la discriminación que, en ocasiones, preside juicios y elecciones humanas, sin embargo, no es así, habiendo dado lugar a amenazas que adoptan nuevas formas, como son los sesgos algorítmicos (Belloso, 2020: 45).

Otro de los peligros es utilizar la inteligencia artificial para tomar decisiones influenciadas por la etnia, el sexo o la edad incluidos en los datos al contratar o despedir, ofrecer préstamos o incluso en procesos penales. La IA también supone riesgos para la privacidad y la protección de datos al utilizarse, por ejemplo, en equipos de reconocimiento facial o para el seguimiento en línea y la creación de perfiles de personas (Bueno, 2020: 3). Asimismo, esta tecnología presenta riesgos para la democracia, al crear, por ejemplo, cámaras de eco por internet basadas en el comportamiento previo de alguien en la red, al mostrar solo un contenido específico. Los sistemas de IA también pueden usarse para crear vídeos, audios o imágenes falsas pero realistas, conocidos como *«deepfakes»*. Este contenido puede implicar riegos financieros, daños reputacionales y problemas en la toma de decisiones. De igual forma, la libertad de reunión y protesta está también amenazada por la IA, ya que ésta podría rastrear y controlar a las personas vinculadas a determinadas creencias o ideologías políticas.

> Basta recordar que los ejes sobre los que se erige la Carta ética europea, sobre el uso de la inteligencia artificial en los sistemas judiciales y su entorno, adoptada por la Comisión Europea para la Eficacia de la Justicia (CEPEJ), el 3 de diciembre de 2018, incluyen junto a la calidad, la seguridad, la transparencia y el control del usuario, la no discriminación, la imparcialidad, la equidad y el respeto a los derechos fundamentales.
>
> Sin ir más lejos, hace escasas fechas nos hacíamos eco de la noticia de que la Unión Europea ultima la implantación de sistemas de reconocimiento facial en todas sus fronteras, con el propósito de recopilar datos biométricos de los ciuda-

danos de terceros países, es decir, de aquellos ciudadanos procedentes de fuera del espacio Schengen. Hasta el momento, únicamente se les tomaban las huellas dactilares y se contrastaban dichos datos en tiempo real con las listas de personas señaladas por las autoridades y fuerzas de seguridad. En cambio, ahora la información quedará almacenada en una gran base de datos que aspira a recopilar y gestionar las huellas dactilares y las imágenes faciales de más de 400 millones de personas, información que gestionará la Comisión Europea y a la que podrán acceder los Estados miembros. El desarrollo del Shared Biometric Matching Service (sBMS) corrió a cargo del consorcio IDEMIA y Sopra Steria, dos multinacionales ubicadas en Paris, y su presupuesto oscila los 302.500.000 de euros. La aplicación de esta tecnología en las fronteras europeas desató un huracán de opiniones contrarias de activistas, políticos y tecnólogos, conscientes de que este tipo de herramientas pueden potenciar la discriminación, incurrir en sesgos y criminalizar a la ciudadanía (Guilabert, 2022: 89).

Cuando hacemos alusión al término «cámara de eco» nos estamos refiriendo a aquel estado de aislamiento intelectual que resulta de unas determinadas prácticas de acceso a la información. En tiempos pasados, cuando el acceso a la información se producía fundamentalmente a través de unos pocos canales y medios de comunicación, las cámaras de eco se producían cuando una persona, voluntariamente, decidía informarse de manera limitada en tan solo algunos de ellos, un efecto que, además, se veía reforzado cuando muchos de sus contactos tomaban la misma opción. Con la adopción masiva de internet y la llegada de las redes sociales hemos experimentado un incremento brutal del número de fuentes de información disponibles, el cual unido a la perversa influencia que genera el *gobierno* de los algoritmos que pueblan estos nuevos medios, está ocasionando dos importantes efectos estrechamente vinculados. En primer término, esta *novedosa* combinación está produciendo un importante sesgo de confirmación de los individuos, tal y como evidenció el asalto al Capitolio de Estados Unidos. Y, en segundo lugar, el análisis detallado de nuestro tráfico como usuarios de los recursos digitales está generando un tesoro de información de dimensiones desconocidas sobre el comportamiento humano.

El término *deepfakes* apareció en 2017 para hacer alusión a vídeos manipulados con el propósito de hacer creer a los usuarios que ven a una determinada persona, tanto si es anónima como si es personaje público, realizando declaraciones o acciones que nunca ocurrieron. Para la creación de dichos vídeos, se utilizan herramientas o programas dotados de tecnología de inteligencia artificial que permiten el intercambio de rostros en imágenes y la modificación de la voz. El objetivo de tales videos no es otro que el de poder realizar copias digitalizadas de cualquier personaje público o privado para poder hacer que esta copia haga o diga lo que el autor o autores de esta creación pretende. Esta técnica es relativamente fácil de implementar mediante el uso de un *software* de aprendizaje profundo que permita el tratamiento de videos que estén presentes en la web en código abierto. Dicho *software* funciona gracias a un mecanismo que permite esquivar los sistemas de detección, a través del uso de dos algoritmos que actúan de la siguiente forma: el primero, copia una vídeo multitud de veces de forma idéntica importando una cara externa y, el segundo, detecta la calidad de los

> vídeos creados por el primer algoritmo con el fin de excluir aquellos marcos o *frames* menos creíbles (Lavanda, 2022: 86).

En efecto, como ha reconocido la propia Unión en diversos pronunciamientos, la IA genera numerosas dudas entre los usuarios, investigadores, especialistas, autoridades y la propia industria encargada de su desarrollo. En singular, estas preocupaciones se centran en lo que concierne a los aspectos relativos al cumplimiento normativo, el respeto de los derechos y libertades fundamentales de los interesados (privacidad, igualdad y no discriminación, dignidad, etc.) o la seguridad jurídica de todos los intervinientes en aquellos procesos en los que la innovación digital se erige como componente primordial; cuestiones esenciales que, ante la inacción de los poderes públicos, terminan por constituir un importante freno para el correcto desarrollo tecnológico (Cotino y Bauza, 2022).

Con la finalidad de disipar los interrogantes que envuelven el prodigioso avance de las nuevas tecnologías, en especial la IA, es urgente que el poder público ofrezca una respuesta decidida, tendente al establecimiento de un sólido marco ético y normativo que refuerce la protección de los derechos individuales y colectivos, con el propósito de garantizar la inclusión y el bienestar social del conjunto de la ciudadanía. Esta apremiante cuestión exige una minuciosa labor jurídica orientada no solamente a la articulación de garantías que permitan salvaguardar la plena vigencia y efectividad del elenco de derechos fundamentales ya reconocidos, sino también la identificación de reformas legales necesarias, así como de las lagunas jurídicas que requieran una regulación adicional para otorgar seguridad jurídica, elemento indispensable para el fomento de la innovación digital.

Consciente de esta realidad, el viejo continente ha desempeñado en los últimos años una laboriosa tarea, con la finalidad de diseñar el primer texto jurídico verdaderamente relevante encargado de embridar el avance de los sistemas algorítmicos, situando la dignidad de la persona, el respeto de los derechos y libertades fundamentales y los valores esenciales que sustentan la concepción de la ciudadanía europea en el centro gravitacional de la acción normativa, en lo que constituye un auténtico aldabonazo en la travesía hacia una transición digital humanista.

> En palabras de Álvarez y Tahirí (2023: 5) afirman que «[r]esulta difícil enumerar la cantidad tan inmensa de textos de *softlaw* sobre la inteligencia artificial que han generado las Instituciones Europeas durante los últimos años en el seno de la Unión. Nos interesa destacar ahora, no obstante, dos documentos que se integran en el procedimiento legislativo ordinario iniciado por la Comisión para regular este tipo de sistemas, que recordemos son tratados como productos. Nos referimos, en primer término, a la "Propuesta de Reglamento por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de inteligencia artificial)", formulada el 21 de abril de 2021 por la Comisión, que es la Institución que tiene la iniciativa en este tipo de procedimiento legislativo euro-

peo; y, en segundo término, el texto de la propuesta transaccional acordada por el Consejo con fecha de 25 de noviembre de 2022».

A este respecto, igualmente conviene recordar que en otras latitudes ya existen otras propuestas legislativas, de diferente alcance y magnitud, que pretenden normativizar el fenómeno de la IA, a saber: las «Opiniones Orientadoras sobre el Fortalecimiento de la Gobernanza Integral de algoritmos de servicios de información de Internet» de la República Popular China, de 17 de septiembre de 2021; o la Ley de Responsabilidad Algorítmica de Estados Unidos, de 3 de febrero de 2022, entre otras.

II. Concordancias

Para la aplicación del presente artículo, véase también:

– Art. 6 del RIA.

– Arts. 9 a 15 del RIA.

III. Comentario

Como se ha apuntado ya en otros pasajes de esta obra, el capítulo III del RIA se detiene en establecer la clasificación de los conocidos como sistemas de alto riesgo de IA (SARIA), sobre los cuales va a recaer buena parte de la carga regulatoria europea. A tal fin, el art. 6 RIA establece que un sistema de IA se considerará de alto riesgo cuando, con independencia de si se ha introducido en el mercado o se ha puesto en servicio, reúna las dos condiciones que se indican a continuación: (a) que el sistema de IA está destinado a ser utilizado 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, o que el propio sistema de IA sea uno de dichos productos; y (b) que el producto del que el sistema de IA es componente de seguridad, o el propio sistema de IA como producto, debe someterse a una evaluación *ex ante* de la conformidad realizada por un organismo independiente para su introducción en el mercado o puesta en servicio.

También se considerarán de alto riesgo los sistemas de IA que figuran en el anexo III, es decir: Identificación biométrica y categorización de personas físicas: (a) sistemas de IA destinados a utilizarse en la identificación biométrica remota «en tiempo real» o «en diferido» de personas físicas; (b) sistemas de IA destinados a utilizarse como componentes de seguridad en la gestión y funcionamiento del tráfico rodado y el suministro de agua, gas, calefacción y electricidad; (c) sistemas de IA destinados a utilizarse para determinar el acceso o la asignación de personas físicas a los centros de educación y formación profesional; (d) sistemas de IA destinados a utilizarse para evaluar a los estudiantes de centros de educación y formación profesional y para evaluar a los participantes en pruebas generalmente necesarias para acceder a centros de educación; (e) sistemas de IA destinados a utilizarse para la contratación o selección de personas físicas, especialmente para anunciar puestos vacantes, clasificar y filtrar solicitudes o evaluar a candidatos en el transcurso de entrevistas o pruebas; (f) IA destinada a utilizarse para tomar decisiones relativas a la promoción y resolución de relaciones contractuales de

índole laboral, a la asignación de tareas y al seguimiento y evaluación del rendimiento y la conducta de las personas en el marco de dichas relaciones; (g) sistemas de IA destinados a ser utilizados por las autoridades públicas o en su nombre para 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; (h) sistemas de IA destinados a utilizarse para evaluar la solvencia de personas físicas o establecer su calificación crediticia, salvo los sistemas de IA puestos en servicio por parte de proveedores a pequeña escala para su uso propio; (i) sistemas de IA destinados a utilizarse para el envío o el establecimiento de prioridades en el envío de servicios de primera intervención en situaciones de emergencia, por ejemplo bomberos y servicios de asistencia médica; (j) 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; (k) sistemas de IA destinados a utilizarse por parte de las autoridades encargadas de la aplicación de la ley como polígrafos y herramientas similares, o para detectar el estado emocional de una persona física; (l) sistemas de IA destinados a utilizarse por parte de las autoridades encargadas de la aplicación de la ley para detectar ultrafalsificaciones a las que hace referencia el art. 52.3 RIA; (m) sistemas de IA destinados a utilizarse por parte de las autoridades encargadas de la aplicación de la ley para la evaluación de la fiabilidad de las pruebas durante la investigación o el enjuiciamiento de infracciones penales; (n) 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 art. 3.4 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; (ñ) sistemas de IA destinados a utilizarse por parte de las autoridades encargadas de la aplicación de la ley para la elaboración de perfiles de personas físicas, de conformidad con lo dispuesto en el artículo 3, apartado 4, de la Directiva (UE) 2016/680, durante la detección, la investigación o el enjuiciamiento de infracciones penales; (o) 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; (p) sistemas de IA destinados a utilizarse por parte de las autoridades públicas competentes como polígrafos y herramientas similares, o para detectar el estado emocional de una persona física; (q) sistemas de IA destinados a utilizarse por parte de las autoridades públicas competentes para evaluar un riesgo, como un riesgo para la seguridad, la salud o relativo a la inmigración ilegal, que plantee una persona física que pretenda entrar o haya entrado en el territorio de un Estado miembro; (r) sistemas de IA destinados a utilizarse por parte de las autoridades públicas competentes para la verificación de la autenticidad de los documentos de viaje y los documentos justificativos de las personas físicas y la detección de documentos falsos mediante la comprobación de sus elementos de seguridad; (s)

sistemas de IA destinados a ayudar a las autoridades públicas competentes a examinar las solicitudes de asilo, visado y permisos de residencia, y las reclamaciones asociadas con respecto a la admisibilidad de las personas físicas solicitantes; y (t) 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.

En el citado anexo I aparece enumerada la siguiente legislación armonizada: Directiva 2006/42/CE del Parlamento Europeo y del Consejo, de 17 de mayo de 2006, relativa a las máquinas y por la que se modifica la Directiva 95/16/CE; Directiva 2009/48/CE del Parlamento Europeo y del Consejo, de 18 de junio de 2009, sobre la seguridad de los juguetes; 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; Directiva 2014/33/UE del Parlamento Europeo y del Consejo, de 26 de febrero de 2014, sobre la armonización de las legislaciones de los Estados miembros en materia de ascensores y componentes de seguridad para ascensores; Directiva 2014/34/UE del Parlamento Europeo y del Consejo, de 26 de febrero de 2014, sobre la armonización de las legislaciones de los Estados miembros en materia de aparatos y sistemas de protección para uso en atmósferas potencialmente explosivas; Directiva 2014/53/UE del Parlamento Europeo y del Consejo, de 16 de abril de 2014, relativa a la armonización de las legislaciones de los Estados miembros sobre la comercialización de equipos radioeléctricos, y por la que se deroga la Directiva 1999/5/CE; Directiva 2014/68/UE del Parlamento Europeo y del Consejo, de 15 de mayo de 2014, relativa a la armonización de las legislaciones de los Estados miembros sobre la comercialización de equipos a presión; Reglamento (UE) 2016/424 del Parlamento Europeo y del Consejo, de 9 de marzo de 2016, relativo a las instalaciones de transporte por cable y por el que se deroga la Directiva 2000/9/CE; Reglamento (UE) 2016/425 del Parlamento Europeo y del Consejo, de 9 de marzo de 2016, relativo a los equipos de protección individual y por el que se deroga la Directiva 89/686/CEE del Consejo; Reglamento (UE) 2016/426 del Parlamento Europeo y del Consejo, de 9 de marzo de 2016, sobre los aparatos que queman combustibles gaseosos y por el que se deroga la Directiva 2009/142/CE; 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; 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; 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; Reglamento (UE) n.º 168/2013 del Parlamento Europeo y del Consejo, de 15 de enero de 2013, relativo a la homologación de los vehículos de dos o tres ruedas y los cuatriciclos, y a la vigilancia del mercado de dichos vehículos; Reglamento (UE) n.º 167/2013 del Parlamento Europeo y del Consejo, de 5 de febrero de 2013, relativo a la homologación de los vehículos agrícolas o forestales, y a la vigilancia del mercado de

dichos vehículos; 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; Directiva (UE) 2016/797 del Parlamento Europeo y del Consejo, de 11 de mayo de 2016, sobre la interoperabilidad del sistema ferroviario dentro de la Unión Europea; 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, los componentes y las 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 por el que se deroga la Directiva 2007/46/CE; 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 de sus remolques, así como los sistemas, componentes y unidades técnicas independientes destinados a esos vehículos, en lo que respecta a su seguridad general y a la protección de los ocupantes de los vehículos 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 (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 de la Comisión; y el 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 la Unión Europea para la Seguridad Aérea y por el que se modifican los Reglamentos (CE) n.º 2111/2005 (CE) n.º 1008/2008 (UE) n.º 996/2010 (CE) 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, en la medida en la que afecte al diseño, la producción y la comercialización de aeronaves contemplados en el artículo 2, apartado 1, letras a) y b), cuando se refiera a aeronaves no tripuladas y sus motores, hélices, componentes y equipos para controlarlas a distancia.

También se considerarán de alto riesgo los sistemas de IA contemplados en el anexo III, es decir, los sistemas de IA que formen parte de cualquiera de los ámbitos siguientes: (i) biometría; (ii) infraestructuras críticas; (iii) educación y formación profesional; (iv) empleo, gestión de los trabajadores y acceso al autoempleo; (v) acceso a servicios privados esenciales y a servicios y prestaciones públicas esenciales; (vi) garantía del cumplimiento del Derecho; (vii) migración, asilo y gestión del control fronterizo; y (viii) Administración de justicia y procesos democráticos.

De lo anterior se desprende la apuesta del legislador europeo por incorporar un adecuado y marcado enfoque de riesgo como elemento vertebrador del RIA, el cual al igual que ocurre en el 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 (RGPD) acabará cobrando una relevancia capital, toda vez que el instrumento jurídico convierte a los SARIA que plantean riesgos significativos para la salud, la seguridad o los derechos fundamentales de las personas en el centro gravitacional de la regulación.

> De la lectura pausada del Reglamento se deduce el establecimiento de cuatro niveles de riesgo diferenciados, en el primero se prohíben completamente determinadas prácticas en el uso de la IA, el segundo propone una gestión específica de sistemas de alto riesgo a los que se exigen evaluaciones independientes para determinar la conformidad, regula un tercer sistema de transparencia para sistemas con riesgo de manipulación y, finalmente, se permite el libre uso de la IA, y en estos casos, tiene en cuenta situaciones específicas de grupos vulnerables, como los niños, o las personas migrantes. También considera el procedimiento de evaluación aplicable a los sistemas de IA que presenten riesgos a nivel nacional (Garde, 2022: 3).

A grandes rasgos, podemos señalar que el sistema por el que apuesta el legislador europeo consiste en la imposición de unos requisitos mínimos necesarios para abordar los riesgos y problemas vinculados a la IA, con los que se pretende generar confianza y certidumbre, al tiempo que se intenta no limitar u obstaculizar el desarrollo tecnológico y evitar el aumento del coste de la comercialización de soluciones de IA.

> En este punto, conviene precisar que sería conveniente que el nivel de riesgo alto de los SARIA estuviera dividido en subniveles, con la consiguiente subdivisión de requerimientos y obligaciones para los desarrolladores tecnológicos, habida cuenta de que la RIA contempla un amplio listado de casuísticas con implicaciones muy distintas.
>
> Como se expone en la propia RIA, el texto «sigue un enfoque basado en el riesgo e impone cargas reglamentarias sólo cuando es probable que un sistema de IA plantee riesgos elevados para los derechos fundamentales y la seguridad». De este modo, a los sistemas de IA de alto riesgo (SARIA) se determinan «los requisitos de datos de alta calidad, documentación y trazabilidad, transparencia, supervisión humana, exactitud y solidez, son estrictamente necesarios para mitigar los riesgos para los derechos fundamentales y la seguridad que plantea la IA y que no están cubiertos por otros marcos jurídicos existentes».

Sin embargo, cabe señalar que la actual redacción comporta algunos problemas, habida cuenta de que los usos de IA de alto riesgo no incluyen únicamente aquellos sistemas basados en la puesta en marcha de decisiones automatizadas, por lo que las garantías comprendidas en el art. 22 RGPD resultarían claramente insuficientes y no permitirán contener los riesgos que entraña su desarrollo y comercialización. Además, el RGPD tampoco permitiría regular el empleo de aquellos sistemas de IA en los que no existan operaciones de tratamiento de datos personales.

> Junto a estas lagunas normativas y, sin perjuicio de la valoración general del sistema de riesgos, lo cierto es que no queda claro el enfoque sobre los riesgos, es decir, si se hará sobre los riesgos de negocio o sobre los riesgos que el impulso de estos sistemas comporta para el interesado y, en este último caso, si quedan circunscritos a riesgos para los derechos fundamentales o a todos los derechos asociados. Puede tomarse como referencia el RGPD, que define la seguridad como «la capacidad de garantizar la confidencialidad, integridad, disponibilidad y resiliencia permanentes de los sistemas y servicios de tratamiento» [riesgos para la seguridad de la información, art. 32.1.b) RGPD], y que se refiere al riesgo del siguiente modo: «los riesgos de diversa probabilidad y gravedad para los derechos y libertades de las personas físicas» (art. 24.1 RGPD). También entendemos que debería incluirse expresamente la obligatoriedad de que los datos de prueba sean similares en las distintas fases (entrenamiento, validación y prueba). De lo contrario, es difícil alcanzar un resultado fiable.

De conformidad con el art. 8 RIA, precepto cuyo análisis se acomete en las próximas líneas, los sistemas de IA de alto riesgo cumplirán necesariamente, 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, una serie de requisitos entre los que sobresalen las siguientes cuestiones (Tahirí, 2024: 159):

(A) Implantar un sistema de gestión de riesgos. Hace referencia a la obligación del promotor de implementar un proceso iterativo continuo, planificado y ejecutado durante todo el ciclo de vida de un sistema de IA de alto riesgo, que requerirá revisiones y actualizaciones sistemáticas periódicas y que, en todo caso, deberá constar como mínimo de las siguientes etapas:

1. La determinación y el análisis de los riesgos conocidos y previsibles que el sistema de IA de alto riesgo pueda plantear 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.

2. 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.

3. 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 art. 72 RIA.

4. La adopción de medidas adecuadas y específicas de gestión de riesgos diseñadas para hacer frente a los riesgos detectados.

De conformidad con la redacción del propio art. 3.2 RIA, se entiende por riesgo «la combinación de la probabilidad de que se produzca un perjuicio y la gravedad de dicho perjuicio».

(B) Impulsar prácticas de gobernanza y gestión de datos adecuadas. Hace alusión a la necesidad de que el proveedor articule una serie de buenas prácticas sustentadas en criterios de calidad que permitan gestionar los conjuntos de datos empleados para el entrenamiento, validación y prueba de los modelos de IA, las cuales deben permitir una gestión adecuada de la información para la finalidad prevista del sistema de IA de alto riesgo. Dichas prácticas se centrarán, en particular, en los siguientes extremos:

1. Las decisiones pertinentes relativas al diseño.

2. Los procesos de recogida de datos y el origen de los datos y, en el caso de los datos personales, la finalidad original de la recogida de datos.

3. Las operaciones de tratamiento oportunas para la preparación de los datos, como la anotación, el etiquetado, la depuración, la actualización, el enriquecimiento y la agregación.

4. La formulación de supuestos, en particular en lo que respecta a la información que se supone que miden y representan los datos.

5. Una evaluación de la disponibilidad, la cantidad y la adecuación de los conjuntos de datos necesarios.

6. El 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.

7. Medidas adecuadas para detectar, prevenir y mitigar posibles sesgos detectados.

8. La detección de lagunas o deficiencias pertinentes en los datos que impidan el cumplimiento del Reglamento de Inteligencia Artificial, y la forma de subsanarlas.

Con todo ello se establecen una serie de principios inspiradores que deben vehicular y guiar el proceso de entrenamiento de los sistemas de IA de alto riesgo, entre los que destacan la calidad y la finalidad de los conjuntos de datos empleados, así como la tutela jurídica de las categorías especiales de datos personales, cuando sea estrictamente necesario para garantizar la detección y corrección de los sesgos asociados a los sistemas de IA de alto riesgo.

(C) Elaborar la documentación técnica del sistema de IA de alto riesgo con carácter previo a la introducción en el mercado o puesta en servicio de la solución tecnológica. Hace mención a la necesidad de que el proveedor elabore y actualice la documentación técnica del sistema algorítmico de alto riesgo antes de su puesta en marcha. Adicionalmente, el legislador europeo prevé que dicha

documentación técnica permita demostrar de manera clara y completa el cumplimiento de los requisitos objeto de análisis.

> De conformidad con el anexo IV del RIA, la documentación técnica a que se refiere el art. 11.1 RIA incluirá como mínimo la siguiente información, aplicable al sistema de IA pertinente: (i) una descripción general del sistema de IA que incluya su finalidad, el nombre del proveedor, la manera en que este interactúa con otros sistemas de IA, la versión de *software* o firmware, la descripción del hardware, una descripción básica de la interfaz de usuario facilitada al responsable del despliegue, etc.; (ii) una descripción detallada de los elementos del sistema de IA y de su proceso de desarrollo, que necesariamente debe incluir los métodos y las medidas adoptados para el desarrollo del sistema de IA, las especificaciones de diseño del sistema, la arquitectura del sistema, una evaluación de las medidas de supervisión humana articuladas, los procedimientos de validación y prueba utilizados y las medidas de ciberseguridad adoptadas; (iii) información detallada acerca de la supervisión, el funcionamiento y el control del sistema de IA, en particular con respecto a sus capacidades y limitaciones de funcionamiento, incluidos los niveles de precisión para las personas o colectivos de personas específicos en relación con los que está previsto que se utilice el sistema y el nivel general de precisión esperado en relación con su finalidad prevista; los resultados no deseados previsibles y las fuentes de riesgo para la salud y la seguridad, los derechos fundamentales y la discriminación, en vista de la finalidad prevista del sistema de IA; las medidas de supervisión humana necesarias de conformidad con el art. 14 RIA, incluidas las medidas técnicas establecidas para facilitar la interpretación de los resultados de salida de los sistemas de IA por parte de los responsables del despliegue; las especificaciones de los datos de entrada, según proceda; (iv) una descripción de la idoneidad de los parámetros de rendimiento para el sistema de IA concreto; (v) una descripción detallada del sistema de gestión de riesgos con arreglo al art. 9 RIA; (vi) una descripción de los cambios pertinentes realizados por el proveedor en el sistema a lo largo de su ciclo de vida; (vii) una lista de las normas armonizadas, aplicadas total o parcialmente, cuyas referencias se hayan publicado en el Diario Oficial de la Unión Europea; cuando no se hayan aplicado normas armonizadas, una descripción detallada de las soluciones adoptadas para cumplir los requisitos establecidos en el Título III, Capítulo 2, incluida una lista de otras normas y especificaciones técnicas pertinentes que se hayan aplicado; (viii) una copia de la declaración UE de conformidad de conformidad con el art. 47 RIA; y (ix) una descripción detallada del sistema establecido para evaluar el funcionamiento del sistema de IA en la fase posterior a la comercialización, de conformidad con el art. 72 RIA, incluido el plan de vigilancia poscomercialización a que se refiere el art. 72.3 RIA.

(D) Conservar los registros de acontecimientos a lo largo de todo el ciclo de vida del sistema de IA de alto riesgo. Hace referencia a la obligación del proveedor de articular un mecanismo, adecuado a la finalidad del sistema, que permita garantizar la trazabilidad del funcionamiento del sistema de IA de alto riesgo con un triple propósito:

– Detectar situaciones que puedan dar lugar a que el sistema de IA de alto riesgo presente un riesgo para la salud, la seguridad o los derechos fundamentales de las personas, de conformidad con el art. 79 RIA.

– Facilitar la vigilancia poscomercialización del sistema en los términos contemplados en el art. 72 RIA.

– Vigilar la articulación de las medidas técnicas y organizativas adecuadas para garantizar la correcta utilización y funcionamiento de los sistemas de IA de alto riesgo, en virtud de lo preceptuado en el art. 26 RIA.

(E) Articular medidas de transparencia y comunicación de información. Hace alusión a la necesidad de que los sistemas de IA de alto riesgo se diseñen y desarrollen de tal 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. A tal fin, 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.

> En virtud del art. 13.3 RIA, las instrucciones de uso contendrán al menos la siguiente información: (a) la identidad y los datos de contacto del proveedor y, en su caso, de su representante autorizado; (b) las características, capacidades y limitaciones del funcionamiento del sistema de IA de alto riesgo; (c) los cambios en el sistema de IA de alto riesgo y su funcionamiento predeterminado por el proveedor en el momento de efectuar la evaluación de la conformidad inicial; (d) las medidas de supervisión humana, incluidas las medidas técnicas establecidas para facilitar la interpretación de los resultados 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 para garantizar el correcto funcionamiento de dicho sistema; y (f) cuando proceda, una descripción de los mecanismos incluidos en el sistema de IA de alto riesgo que permita a los responsables del despliegue recabar, almacenar e interpretar correctamente registros de acontecimiento del sistema de IA de alto riesgo.

(F) Fomentar medidas que permitan la supervisión humana de los sistemas de IA de alto riesgo. Hace mención a la necesidad de que los sistemas de IA de alto riesgo se diseñen y desarrollen 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. El objetivo de la supervisión 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. Ello exigirá que los responsables del despliegue de la solución

tecnológica encomienden la supervisión humana a personas físicas que tengan la competencia, la formación y la autoridad necesarias y adecuadas.

> Esto es lo que algunos avezados estudiosos han denominado hábilmente «reserva de humanidad». Entre nosotros, Ponce (2019: 7) quien a propósito de la cuestión objeto de estudio afirma que «la IA se desarrolla mediante el uso de algoritmos y datos. En la cocina de la IA, los primeros serían las recetas y los segundos los ingredientes. Los algoritmos son un conjunto de instrucciones para solucionar un problema. Los mismos han ido haciéndose más complejos con el tiempo, pasando de ser estáticos, en el sentido de que los programadores diseñaban ya en los mismos los criterios para tomar las decisiones, a ser dinámicos, en el sentido de que los algoritmos denominados de aprendizaje automático (*machine learning*) tienen la capacidad de aprender con el tiempo de los datos y experiencias para tomar decisiones por sí mismos, generando sus propias instrucciones que ya no son las iniciales del programador. Por otro lado, el llamado aprendizaje profundo *(deep learning)* supone un funcionamiento de la IA emulando redes neuronales complejas. En estos casos, los algoritmos extraen patrones de las masas de datos y los resultados que se obtienen no están relacionados de modo lineal sino complejo, por lo que no es sencillo determinar la causalidad entre datos y decisión adoptada. Por ello, ha sido dicho que en un medio ambiente de aprendizaje automático el problema con estos algoritmos es que los humanos no pueden ya estar en control de qué decisión es tomada y ni siquiera pueden saber o comprender por qué una decisión errónea ha sido tomada dado que están perdiendo de vista la transparencia del proceso desde el principio hasta el final, deviniendo opacos para sus creadores que no entienden la lógica que siguen. De ahí que se hable del proceso de adopción de la decisión por parte de los algoritmos como una caja negra *(black box)* […] Cuando se ejerce discrecionalidad administrativa y se toman en consideración hechos, intereses y derechos de personas, esa empatía debería estar presente, y parece que sólo podría ejercerla un ser humano. Estaríamos ante un elemento de la *phronesis*, la sabiduría práctica o prudencia, aristotélica referida en Ética a Nicómaco, esto es al sabio ejercicio del poder, basado en una ponderación cuidadosa de las circunstancias pertinentes en cada toma de la decisión. Una IA podría imitar cognitivamente la empatía (entrenamiento en gestos faciales, tono de voz, etc.) pero carece de la consciencia, emoción y humanidad precisa para no degenerar en un psicópata de sílice, puesto que los psicópatas humanos carecen de empatía, pero la pueden simular, muy bien, para sus propios intereses […] Si, por tanto, la IA no puede acabar consiguiendo disponer de máquinas con empatía emocional con los humanos, porque hace falta ser humano para ello, entonces IA y empatía emocional sería una *contradictio in terminis,* y el sueño, en el sentido de aspiración, de la razón total (artificial en este caso) podría llegar a producir monstruos burocráticos que no resuenen emocionalmente ni se conecten con los humanos».

(G) Garantizar un nivel adecuado de precisión, solidez y ciberseguridad. De conformidad con este requisito, 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 durante todo su ciclo de vida. Para ello, 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 medirla. Asimismo, los sistemas de IA de alto riesgo serán lo más resistentes posible en lo que respecta a los errores, fallos o incoherencias que pueden surgir en los propios sistemas o en el entorno en el que funcionan, en particular a causa de su interacción con personas físicas u otros sistemas, para lo cual se adoptarán medidas técnicas y organizativas a este respecto. Finalmente, en lo que respecta a la ciberseguridad, los sistemas de IA de alto riesgo serán resistentes a los intentos de terceros no autorizados de alterar su uso, sus resultados de salida o su funcionamiento aprovechando las vulnerabilidades del sistema. A tal fin, 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.

> Para abordar los aspectos técnicos sobre la forma de medir los niveles adecuados de precisión y solidez, así como cualquier otro parámetro de rendimiento pertinente, la Comisión, en cooperación con las partes interesadas y organizaciones pertinentes, como las autoridades de metrología y de evaluación comparativa, fomentará, según proceda, el desarrollo de parámetros de referencia y metodologías de medición (art. 15.2 RIA).
>
> 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 entrenados previamente utilizados en el entrenamiento («envenenamiento de modelos»), la información de entrada diseñada para hacer que el modelo de IA cometa un error («ejemplos adversarios» o «evasión de modelos»), los ataques a la confidencialidad o los defectos en el modelo (art. 15.5 RIA).

De igual forma, el art. 8.2 RIA contempla que cuando un producto contenga un sistema de IA al que se apliquen los requisitos antes enunciados, así como los requisitos de los actos legislativos de armonización de la Unión enumerados en el anexo I, los proveedores serán responsables de garantizar que su producto cumpla plenamente todos los requisitos aplicables en virtud de los actos legislativos de armonización de la Unión que sean aplicables. Para garantizar el cumplimiento de los sistemas de IA de alto riesgo, y con el fin de garantizar la coherencia, evitar duplicidades y reducir al mínimo las cargas adicionales, los proveedores podrán optar por integrar, según proceda, los procesos de prueba y presentación de información necesarios, y la información y la documentación que faciliten con respecto a su producto en documentación y procedimientos que ya existan y exijan los actos legislativos de armonización de la Unión enumerados en el anexo I.

IV. Bibliografía

Álvarez García, V. y Tahirí Moreno, J. (2023). La regulación de la inteligencia artificial en Europa a través de la técnica armonizadora del nuevo enfoque. *Revista General de Derecho Administrativo*, núm. 63.

Belloso Martín, N. (2020). «La problemática de los sesgos algorítmicos (con especial referencia a los de género). ¿Hacia un derecho a la protección contra los sesgos?», en Llano Alonso, F. H. (dir.). *Inteligencia artificial y filosofía del derecho*. Ediciones Laborum, Murcia.

Bueno De Mata, F. (2020). Macrodatos, inteligencia artificial y proceso: luces y sombras. *Revista General de Derecho Procesal*, núm. 51.

Capdeferro Villagrasa, O. (2020). La inteligencia artificial del sector público: desarrollo y regulación de la actuación administrativa inteligente en la cuarta revolución industrial. *IDP: Revista e Internet, Derecho y Política,* núm. 30.

Cerrillo i Martínez, A. (2023). Lección 21. Actuación automatizada, robotizada e inteligente. *Revista de Derecho Público: Teoría y método*, vol. 2023.

Cerrillo i Martínez, A. (2019). El impacto de la inteligencia artificial en el Derecho administrativo ¿Nuevos conceptos para nuevas realidades técnicas?. *Revista General de Derecho Administrativo*, núm. 50.

Cotino Hueso, L. y Bauza Reilly, M. (2022). *Derechos y garantías ante la inteligencia artificial y las decisiones automatizadas*. Editorial Thomson-Reuters-Aranzadi, Cizur Menor.

Fernández de Gatta Sánchez, D. (2020). El ambicioso Pacto Verde Europeo. *Actualidad Jurídica Ambiental*, núm. 101.

Garde Roca, J. A. (2022). Europa: Oportunidades y retos en inteligencia artificial, mercados y servicios digitales. *Encuentros Multidisciplinares*, núm. 70.

Guilabert Vital, M. R. (2022). «La protección de los datos biométricos por parte de la UE especial consideración al Shared Biometric Matching Service (SBMS) dispuesto por la Agencia EU-LISA para el control de las fronteras exteriores a partir de 2022», en AAVV. *Sistemas jurídicos en Europa e Iberoamérica: tendencias actuales*. Editorial Diké S.A.S., Bogotá.

Lavanda Oliva, M. (2022). Deepfake: Cuando la inteligencia artificial amenaza el Derecho y la Democracia. *Revista de Derecho y Tecnología,* núm. 2.

Márquez Díaz, J. (2020). Inteligencia artificial y Big Data como soluciones frente a la COVID-19. *Revista de Bioética y Derecho*, núm. 50.

Menéndez Sebastián, E. M. (2021). «Buena administración, algoritmos y perspectiva de género», en Bonorino Ramírez, P. R., Fernández Acevedo, R. y Valcárcel Fernández, P. (coords.). *Nuevas normatividades. Inteligencia artificial, derecho y género*. Editorial Thomson Reuters-Aranzadi, Cizur Menor.

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, Inteligencia artificial y educación: guía para las personas a cargo de formular políticas, Paris, 2021.

Ponce Solé, J. (2019). Inteligencia artificial, Derecho administrativo y reserva de humanidad: algoritmos y procedimiento administrativo debido tecnológico. *Revista General de Derecho Administrativo*, núm. 50.

Tahirí Moreno, J. A (2024). Una panorámica de los sistemas de inteligencia artificial desde la perspectiva del Derecho administrativo. *Revista Aragonesa de Administración Pública*, núm. 61.

Agradecimientos

La presente contribución científica forma parte de los resultados del proyecto de I+D+i «Desafíos de una Administración tributaria más eficaz: digitalización inclusiva, compliance y mejoras en seguridad jurídica», liderado por la Profa. Dra. María Ángeles Guervós Maíllo. Así mismo, el presente capítulo de libro se enmarca en la estancia de investigación desarrollada por el autor en la Faculdade de Direito da Universidade do Porto bajo la dirección de la Profa. Dra. Juliana Manuela Alves Ferraz Coutinho.

Artículo 9. ***Sistema de gestión de riesgos***

1. Se establecerá, implantará, documentará y mantendrá un sistema de gestión de riesgos en relación con los sistemas de IA de alto riesgo.

2. El sistema de gestión de riesgos se entenderá como un proceso iterativo continuo planificado y ejecutado durante todo el ciclo de vida de un sistema de IA de alto riesgo, que requerirá revisiones y actualizaciones sistemáticas periódicas. Constará de las siguientes etapas:

a) la determinación y el análisis de los riesgos conocidos y previsibles que el sistema de IA de alto riesgo pueda plantear 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).

3. Los riesgos a que se refiere el presente artículo son únicamente aquellos que pueden mitigarse o eliminarse razonablemente mediante el desarrollo o el diseño del sistema de IA de alto riesgo o el suministro de información técnica adecuada.

4. Las medidas de gestión de riesgos mencionadas en el apartado 2, letra d), tendrán debidamente en cuenta los efectos y la posible interacción derivados de

la aplicación combinada de los requisitos establecidos en la presente sección, con vistas a reducir al mínimo los riesgos de manera más eficaz al tiempo que se logra un equilibrio adecuado en la aplicación de las medidas para cumplir dichos requisitos.

5. Las medidas de gestión de riesgos mencionadas en el apartado 2, letra d), considerarán aceptables los riesgos residuales pertinentes asociados a cada peligro, así como el riesgo residual general de los sistemas de IA de alto riesgo.

A la hora de determinar las medidas de gestión de riesgos más adecuadas, se procurará:

a) eliminar o reducir los riesgos detectados y evaluados de conformidad con el apartado 2 en la medida en que sea técnicamente viable mediante un diseño y un desarrollo adecuados del sistema de IA de alto riesgo;

b) implantar, cuando proceda, unas medidas de mitigación y control apropiadas que hagan frente a los riesgos que no puedan eliminarse;

c) proporcionar la información requerida conforme al artículo 13 y, cuando proceda, impartir formación a los responsables del despliegue.

Con vistas a eliminar o reducir los riesgos asociados a la utilización del sistema de IA de alto riesgo, se tendrán debidamente en cuenta los conocimientos técnicos, la experiencia, la educación y la formación que se espera que posea el responsable del despliegue, así como el contexto en el que está previsto que se utilice el sistema.

6. 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 establecidos en la presente sección.

7. Los procedimientos de prueba podrán incluir pruebas en condiciones reales de conformidad con el artículo 60.

8. 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.

9. Cuando se implante el sistema de gestión de riesgos previsto en los apartados 1 a 7, los proveedores prestarán atención a si, en vista de su finalidad prevista, es probable que el sistema de IA de alto riesgo afecte negativamente a las personas menores de dieciocho años y, en su caso, a otros colectivos vulnerables.

10. 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.

José Luis Domínguez Álvarez
Profesor de Derecho Administrativo de la Universidad de Salamanca

I. Consideraciones previas

Cuando nos encontramos al borde de la quinta revolución industrial, fruto de la combinación del abundante ecosistema de datos existente, la creciente potencia de los sistemas algorítmicos y el fortalecimiento de la capacidad informática, la Unión Europea se prepara para «establecer un marco normativo horizontal orientado al futuro, favorable a la innovación para el desarrollo y el uso de la IA, garantizando al mismo tiempo la salvaguarda del Estado de Derecho y de los derechos fundamentales de los ciudadanos de la Unión».

> A tal fin, el Parlamento Europeo señala que, «para que los agentes europeos tengan éxito en la era digital y se conviertan en líderes tecnológicos en el ámbito de la IA, es necesario un marco normativo claro, un compromiso político y una mentalidad más prospectiva, a menudo inexistentes en la actualidad; concluye que, sobre la base de ese enfoque, tanto los ciudadanos como las empresas de la Unión pueden sacar provecho de la IA y de la gran oportunidad que ofrece para impulsar la competitividad, también en materia de prosperidad y bienestar; subraya que los marcos reglamentarios han de configurarse de manera que no impongan obstáculos injustificados para impedir que los agentes europeos tengan éxito en la era digital, en particular las empresas emergentes y las pequeñas y medianas empresas (pymes); destaca que se deben incrementar sustancialmente las inversiones públicas y privadas para crear un clima en el que surjan y se desarrollen en nuestro continente más historias de éxito europeas». Estas son las razones por las que la Institución defiende el establecimiento de una nueva hoja de ruta para convertir al viejo continente en líder mundial en materia tecnológica, bajo el lema «Una Europa adaptada a la era digital»; edificada sobre las siguientes premisas: (a) marco normativo favorable; (b) culminación del mercado único digital; (c) infraestructura verde digital; (d) ecosistema de excelencia; (e) ecosistema de confianza; (f) estrategia industrial; y (g) seguridad (Parlamento Europeo, 2022: 12).

Si bien es cierto que el viejo continente no lidera en estos momentos el desarrollo tecnológico de la IA, la Unión Europea lleva trabajando casi un lustro en el diseño de una suerte de RGPD en materia de inteligencia artificial, bajo la siguiente premisa: «el establecimiento del primer marco regulador del mundo en materia de IA podría suponer una ventaja para la Unión, como pionera, para establecer normas internacionales de IA basadas en los derechos fundamentales, así como para exportar con éxito la "IA de confianza" centrada en el ser humano a todo el mundo».

> La regulación de la inteligencia artificial cuyos trabajos preparatorios inició la Comisión en el año 2018, se ven ahora reforzados y dirigidos a conseguir una

> regulación de la IA en el conjunto de la Unión de forma irremediable; una futura ordenación que complementará a otra anterior y relevante, próxima en la materia, como es la relativa a la protección de datos, consumidores o la ciberseguridad (García, 2022: 304).
>
> En este punto, conviene recordar la Resolución del Parlamento Europeo, de 3 de mayo de 2022, sobre la inteligencia artificial en la era digital, Estrasburgo, 2022, p. 31 [2020/2266(INI)], donde afirma con rotundidad que el establecimiento del primer marco regulador del mundo en materia de IA podría suponer una ventaja para la Unión, como pionera, para establecer normas internacionales de IA basadas en los derechos fundamentales, así como para exportar con éxito la «IA de confianza» centrada en el ser humano a todo el mundo; subraya que es necesario apoyar este planteamiento mediante la coordinación regulatoria y la convergencia con socios democráticos afines.

De esta forma, el pasado 21 de abril de 2021, la Comisión Europea presentó un conjunto de medidas con el objeto de avanzar hacia la regulación jurídica de la inteligencia artificial en el viejo continente. Desde la perspectiva regulatoria, sobresale su 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

> Asimismo, conviene subrayar que algunas de las ideas que aparecen esbozadas en el texto del Reglamento de Inteligencia Artificial (RIA) ya se podían entrever con claridad en el Plan coordinado sobre la inteligencia artificial de la Comisión Europea [COM(2018) 795 final], donde las Instituciones europeas mostraban la importancia de avanzar hacia el establecimiento de una legislación que ofrezca el marco adecuado para la innovación impulsada por la IA y la aceptación de las soluciones que ofrece la IA, al tiempo que aborda los posibles riesgos derivados del uso y las interacciones con la tecnología, incluidas las cuestiones de seguridad cibernética.
>
> A esta norma hay que sumar la Propuesta de Reglamento en materia de responsabilidad civil por el uso de inteligencia artificial, del Parlamento Europeo, de 20 de octubre de 2020 (Plaza, 2022: 7).

De esta forma, irrumpe en el panorama internacional la primera norma jurídica *stricto sensu* propuesta en el campo de la inteligencia artificial, lo que representa un importante salto cualitativo, al superar las limitaciones propias del enfoque ético que se había impuesto hasta la fecha.

> Así, el Reglamento Europeo de Inteligencia Artificial pretende tanto garantizar la tutela de los derechos fundamentales de las personas situadas en la Unión Europea frente a las amenazas y riesgos ligados al desarrollo de herramientas de inteligencia artificial, como reforzar la innovación en el seno de la Unión Europea. De esta forma, se aspira a convertir a Europa en el centro mundial de una «IA digna de confianza» (Barrio, 2022: 82).

En efecto, conviene reseñar que la propuesta se construye en buena parte sobre el modelo de la legislación europea preexistente relativa a la seguridad de los productos y a las disposiciones contempladas en el Libro Blanco sobre la inteligencia artificial: un enfoque europeo orientado a la excelencia y la confianza [COM(2020) 65 final]. Seguidamente, entre los meses de febrero y junio del mismo año, se celebró una consulta pública en la que se recibieron más de 1.200 contribuciones que implicaron 6.667 respuestas de texto libre y 408 documentos presentados analizados por la Comisión. Por su parte, el Parlamento Europeo en octubre 2020 procedió a la elaboración de dos importantes recomendaciones que incluyeron propuestas de regulación de la IA con texto articulado. La más cercana al texto de la PRIA fue 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)]. A esta se sumaron poco después la Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un régimen de responsabilidad civil en materia de inteligencia artificial [2020/2014(INL)] y la Resolución del Parlamento Europeo, de 20 de octubre de 2020, sobre los derechos de propiedad intelectual para el desarrollo de las tecnologías relativas a la inteligencia artificial [2020/2015(INI)], materias estas dos últimas que no se abordan en la PRIA.

No obstante, en el contexto internacional comienzan a vislumbrarse otra serie de intentos normativos, de distinta magnitud y alcance, que conviene igualmente examinar, con el objetivo de dominar la amalgama de iniciativas reguladoras que aspirar a ordenar el vertiginoso despliegue de los sistemas algorítmicos en las diferentes economías y sociedades. Nos estamos refiriendo, como no podía ser de otra manera a los esfuerzos acometidos por las dos grandes superpotencias tecnológicas del momento: la República Popular China y Estados Unidos.

Según el último ranking elaborado por «SCImago Journal & Country Rank», los 15 países del mundo que están a la vanguardia en IA son los siguientes: (i) China; (ii) Estados Unidos; (iii) India; (iv) Reino Unido; (v) Japón; (vi) Alemania; (vii) Francia; (viii) Italia; (ix) Irán; (x) Corea del Sur; (xi) Australia; (xii) Turquía; (xiii) Canadá; (xiv) España; y (xv) Brasil.

Sin embargo, conviene ahora detenerse en el examen de algunos de los pasajes más relevantes del esperado y ansiado RIA. A este respecto, conviene precisar que han transcurridos ya más de tres años desde que la Presidenta Von der Leyen, anunciara que la Comisión Europea procedería a la presentación de una propuesta legislativa encargada de articular un enfoque europeo coordinado sobre las implicaciones éticas y humanas de la IA. Así las cosas, con fecha 21 de abril de 2021, la Comisión hizo pública su propuesta de marco reglamentario sobre inteligencia artificial con los siguientes objetivos específicos: (i) garantizar que los sistemas de IA introducidos y usados en el mercado de la UE sean seguros y respeten la legislación vigente en materia de derechos fundamentales y valores de la Unión; (ii) garantizar la seguridad jurídica para facilitar la inversión e inno-

vación en IA; (iii) mejorar la gobernanza y la aplicación efectiva de la legislación vigente en materia de derechos fundamentales y los requisitos de seguridad aplicables a los sistemas de IA; y (iv) facilitar el desarrollo de un mercado único para hacer un uso legal, seguro y fiable de las aplicaciones de IA y evitar la fragmentación del mercado.

Para alcanzar dichos objetivos, la propuesta presenta un enfoque normativo horizontal, equilibrado y proporcionado, para la IA, que se limita a establecer los requisitos mínimos necesarios para subsanar los riesgos y problemas vinculados a la IA, sin obstaculizar ni impedir indebidamente el desarrollo tecnológico y sin aumentar de un modo desproporcionado el coste de introducir soluciones de IA en el mercado.

> Recuérdese que la base jurídica de la propuesta es, en primer lugar, el artículo 114 del TFUE, que trata de la adopción de medidas para garantizar el establecimiento y funcionamiento del mercado interior, en la medida en que la propuesta tiene como objetivo primordial 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. En este punto, conviene reseñar que 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. Al igual que ocurría en el estadio regulatorio anterior en el campo de la protección de datos de carácter personal, 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.

Al igual que ocurriría en el supuesto concreto del Reglamento General de Protección de Datos, la propuesta se caracteriza por establecer un marco jurídico sólido, pero marcadamente flexible. Por un lado, las opciones reglamentarias fundamentales que plantea, incluidos los requisitos basados en principios que deben cumplir los sistemas de IA (especialmente en lo que atañe a las soluciones tecnológicas de alto riesgo), son amplias y pueden resistir el paso del tiempo. Por otro lado, establece un sistema regulatorio proporcionado centrado en un enfoque normativo basado en los riesgos y claramente definido que no impone restricciones innecesarias al comercio, en el que la intervención jurídica se adapta a aquellas situaciones concretas en las que existe un motivo de preocupación justificado o en las que es posible anticipar razonablemente que se producirá un problema en un futuro próximo. Al mismo tiempo, el marco jurídico

incluye mecanismos flexibles que le permiten adaptarse de manera dinámica a medida que evoluciona la tecnología y surgen nuevas situaciones preocupantes, en lo que representa un claro esfuerzo por contrarrestar la obsolescencia normativa a la que están expuestos aquellos textos que intentan embridar el avance digital.

De esta forma, el Reglamento Europeo de Inteligencia Artificial apuesta abiertamente por el establecimiento de normas armonizadas para el desarrollo, la introducción en el mercado y la utilización de sistemas de IA en la Unión a partir de un enfoque proporcionado basado en los riesgos.

> Como subraya Bini (2022:83-84), la noción de «riesgo» resulta central en la construcción de una «Europa adaptada a la era digital», según el planteamiento presentado por la Comisión Europea 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. Dicha norma representa una contribución particularmente significativa, que aspira a «convertir a Europa en el centro mundial de una inteligencia artificial (IA) digna de confianza». Para que este objetivo pueda ser en concreto alcanzado, se elabora un planteamiento basado en el concepto de «riesgo», clasificado en el marco de una taxonomía que diferencia y ordena los diferentes sistemas de inteligencia artificial según el nivel de riesgo que llevan consigo (*unacceptable, high, limited, minimal*). Sin tomar en consideración todas las diferentes categorías de riesgo, basta con destacar que, si es «inadmisible» («*unacceptable*») el riesgo que conllevan los sistemas de inteligencia artificial que representen una amenaza para los derechos y la seguridad de la persona, incluso llegando a concretar formas de manipulación del comportamiento humano, se considera, en cambio, «alto» el riesgo que caracteriza los sistemas de inteligencia artificial utilizados en sectores neurálgicos como, por ejemplo, las infraestructuras críticas, la formación educativa y profesional, el empleo, la migración, la Administración de justicia y los procesos democráticos.

También propone una definición única de la IA que puede resistir el paso del tiempo. Asimismo, prohíbe determinadas prácticas particularmente perjudiciales de IA por ir en contra de los valores de la Unión y propone restricciones y salvaguardias específicas en relación con determinados usos de los sistemas de identificación biométrica remota con fines de aplicación de la ley. La propuesta establece una sólida metodología de gestión de riesgos para definir aquellos sistemas de IA que plantean un «alto riesgo» para la salud y la seguridad o los derechos fundamentales de las personas (Soriano, 2021:5). Dichos sistemas de IA tendrán que cumplir una serie de requisitos horizontales obligatorios que garanticen su fiabilidad y ser sometidos a procedimientos de evaluación de la conformidad antes de poder introducirse en el mercado de la Unión. Del mismo modo, se imponen obligaciones previsibles, proporcionadas y claras a los proveedores y los usuarios de dichos sistemas, con el fin de garantizar la seguridad y el respeto de la legislación vigente protegiendo los derechos fundamentales

durante todo el ciclo de vida de los sistemas de IA, aspectos todos ellos que aspira a clarificar el presente estudio.

> En el caso de determinados sistemas de IA, solo se proponen obligaciones mínimas en materia de transparencia, en particular cuando se utilizan robots conversacionales o ultrafalsificaciones. Aspectos todos ellos sumamente relevantes que pasamos a examinar en las próximas páginas con mayor grado de detalle. A este respecto, conviene subrayar que, de conformidad con el texto articulado del Reglamento de Servicios Digitales, en relación a las ultrafalsificaciones establece que cuando una plataforma en línea de muy gran tamaño tenga conocimiento de que un contenido es una imagen, un audio o un vídeo generado o manipulado que se parece a personas, objetos, lugares u otras entidades o sucesos existentes y, de manera falsa, parezca auténtico, deberá etiquetar de forma visible el contenido a fin de informar que el contenido no es auténtico.

II. Concordancias

Para la aplicación del presente artículo, véase también:

– Art. 6 del RIA.

– Art. 8 del RIA.

– Arts. 10 a 15 del RIA.

III. Comentario

Los derechos fundamentales son el fundamento básico para garantizar la «primacía del ser humano» en un contexto de cambio tecnológico. Las nuevas tecnologías basadas en los datos han impulsado el desarrollo de la IA, en particular, mediante la creciente automatización de tareas que tradicionalmente eran realizadas por humanos. La rápida transformación digital de las diferentes estructuras sociales y económicas ha favorecido la adopción de la IA y el intercambio de datos, generando así nuevas oportunidades de futuro, como ocurre en el supuesto concreto de la inteligencia artificial generativa (IAG), pero también presentando novedosos retos y amenazas para los derechos humanos y los derechos fundamentales.

> El interés y el bienestar del ser humano deberán prevalecer sobre el interés exclusivo de la sociedad o de la ciencia. Esta idea aparece reflejada con clarividencia en el art. 2 del Instrumento de Ratificación del Convenio para la protección de los derechos humanos y la dignidad del ser humano con respecto a las aplicaciones de la Biología y la Medicina (Convenio relativo a los derechos humanos y la biomedicina), hecho en Oviedo el 4 de abril de 1997.

Como es sabido por todos, el uso de la IA afecta a la plena efectividad de varios derechos fundamentales constitucionalmente reconocidos (Parlamento Europeo, 2017: 7). Más allá de los aspectos específicos contextuales que afectan a una pluralidad de derechos en distinto grado, con carácter general, la Agencia de los Derechos Fundamentales de la Unión Europea (2021: 10), así como dis-

tintas investigaciones coinciden en señalar que esta injerencia de la IA se produce de manera más acuciante en relación a la necesidad de garantizar el uso no discriminatorio de la IA (derecho a no ser discriminado); el requisito de tratar los datos legalmente (derecho a la protección de los datos personales); y la posibilidad de presentar una reclamación frente a aquellas decisiones basadas en soluciones tecnológicas que impliquen el uso de IA y buscar resarcimiento (derecho a un recurso efectivo y a un juicio justo).

> Si bien es cierto que la protección de datos y la no discriminación son dos de los derechos fundamentales en juego ante el auge imparable de la IA, debe recordarse que «existen otros derechos incididos a consecuencia del uso del *big data* y del desarrollo de la inteligencia artificial y sus avanzados algoritmos para la toma de decisiones como el derecho a la tutela judicial, el derecho a la libertad de información, el derecho de sufragio o el derecho de acceso a la información pública». (Gómez, 2022: 93).

En este sentido, uno de los primeros desafíos a los que se enfrenta la doctrina *iuspublicista* en el estado regulatorio actual consiste precisamente en garantizar la completa protección de los derechos fundamentales ya reconocidos en nuestro *corpus* normativo, identificar reformas legales necesarias, así como lagunas jurídicas que requieran una regulación adicional ante los envites del avance digital y, especialmente, ante el desarrollo exponencial de la automatización y la IA. A partir de la constatación de esta realidad, cobran relevancia dos cuestiones: por un lado, la de si el actual marco de los derechos humanos es suficiente o si es necesario reconocer nuevos derechos; por otro, la de si, además de los derechos, es necesario utilizar otras herramientas.

> Un examen de las diferentes propuestas de nuevos derechos, tanto en el ámbito de los derechos digitales como en el de los neuroderechos, vuelve a poner de manifiesto cómo los problemas vinculados a la libertad de pensamiento, privacidad e integridad siguen siendo los grandes temas a discutir. Por su parte, la discusión relativa a la necesidad o no de reconocer nuevos derechos en el marco de la reflexión sobre la suficiencia del discurso de los derechos, ha ido acompañada de otra que trata sobre la oportunidad de utilizar otras herramientas, más allá de los derechos. Entre estas medidas o herramientas, con carácter general la doctrina identifica las siguientes: (a) la creación de una Comisión de Expertos en Derecho y Ciencia Internacional sobre Neuroderechos en Naciones Unidas; (b) el nombramiento por Naciones Unidas de expertos altamente calificados para servir como asesores especiales sobre neuroderechos a organizaciones, instituciones e industria; (c) el mantenimiento de consultas periódicas con países clave por parte de los asesores y la Comisión; (d) la creación de un nuevo tratado o de un protocolo adicional a los tratados existentes para incorporar los neuroderechos; (e) la elaboración de Comentarios generales sobre neuroderechos por parte de los Comités de seguimiento de los tratados; (f) el nombramiento de un Relator especial sobre el impacto de la neurotecnología en los derechos humanos; (g) la creación de una agencia especializada para coordinar las actividades globales de

> neuroderechos y; (h) ayudar a codificar los neuroderechos en un tratado internacional de derechos humanos (De Asís, 2022: 36-37).

En lo que sí parece existir un consenso generalizado entre la doctrina es en la premisa de que el uso de sistemas de IA compromete a una gran diversidad de derechos fundamentales, independientemente del campo de aplicación. Entre ellos se incluyen, entre otros, la privacidad, la no discriminación y el acceso a la justicia. Además de estos derechos, podrían considerarse otros como, por ejemplo, la dignidad humana, el derecho a la seguridad social y la asistencia social, el derecho a una buena administración (especialmente relevante para el Sector público) o la protección de los consumidores (particularmente importante para las empresas).

> A este respecto, Castellanos y Montero (2020: 75) afirman que en el «sistema de justicia, la IA puede ser un motor de mejora o un elemento distorsionador, de modo que existen elementos que favorecen el avance, pero de igual manera generan una dificultad. De ahí la problemática del debate jurídico en la implantación de elementos de IA en el progreso de la justicia».

Por esta razón, la Unión Europea ha considerado que es una obligación impostergable poner en marcha nuevas políticas públicas y adoptar nuevas disposiciones legislativas en materia de IA. Tanto el legislador comunitario como los Estados miembros deben garantizar que se tiene en cuenta en todo momento el respeto de toda la variedad de derechos fundamentales y libertades públicas consagrados tanto en la Carta de Derechos Fundamentales de la Unión Europea (CDFUE) como en los Tratados de la Unión Europea.

> Así las cosas, el considerando 1 del Reglamento Europeo de Inteligencia Artificial consigna la siguiente premisa: «[e]l objetivo del presente Reglamento es mejorar el funcionamiento del mercado interior mediante el establecimiento de un marco jurídico uniforme, en particular para el desarrollo, la introducción en el mercado, la puesta en servicio y la utilización de sistemas de inteligencia artificial (en lo sucesivo, "sistemas de IA") en la Unión, de conformidad con los valores de la Unión, a fin de 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 […] incluidos la democracia, el Estado de Derecho y la protección del medio ambiente, proteger frente a los efectos perjudiciales de los sistemas de IA en la Unión, así como brindar apoyo a la innovación. El presente Reglamento garantiza la libre circulación transfronteriza de mercancías y servicios basados en la IA, con lo que impide que los Estados miembros impongan restricciones al desarrollo, la comercialización y la utilización de sistemas de IA, a menos que el presente Reglamento lo autorice expresamente».

Es necesario, por tanto, que la ley prevea las salvaguardias y garantías pertinentes al objeto de proteger de manera eficaz contra todas aquellas injerencias arbitrarias que amenazan con lesionar los derechos fundamentales de la ciuda-

danía (Sancho, 2019: 1), lo que contribuirá a su vez a proporcionar la siempre necesaria seguridad jurídica, elemento indispensable tanto para los desarrolladores tecnológicos como para los propios usuarios de soluciones de IA. Seguridad jurídica que, por otra parte, no olvidemos reviste importantes dosis de esencialidad, habida cuenta de que una adecuada regulación del fenómeno tecnológico, además de maximizar la confianza ciudadana en estos nuevos productos y servicios alumbrados al calor de la (r)evolución tecnológica, puede ser un importante motor de la innovación digital.

En este sentido, dentro del actual proceso regulador del fenómeno de la IA emprendido en el seno de la Unión Europea se establece la obligatoriedad de incorporar un sistema de gestión de riesgos que debe consistir en un proceso iterativo continuo que sea planificado y ejecutado durante todo el ciclo de vida del sistema de IA de alto riesgo. Dicho proceso debe tener por objeto detectar y mitigar los riesgos pertinentes de los sistemas de IA para la salud, la seguridad y los derechos fundamentales.

> Una cuestión a resolver es identificar los derechos fundamentales que se pueden ver afectados por las tecnologías inteligentes (dignidad, vida, integridad física, igualdad, privacidad, imagen, protección de datos, educación, salud, etc.), lo que exigiría por un lado analizar las eventuales utilizaciones de estas tecnologías y su posible repercusión en los distintos derechos fundamentales; lo que llevaría por otro lado a la necesidad de delimitar su contenido y garantías necesarias. Se ha llegado a afirmar incluso la necesidad de avanzar hacia «un nuevo paradigma en la protección de los derechos fundamentales […] una reorientación a consecuencia de la naturaleza de las tecnologías [disruptivas]» (Sarrión, 2020: 328).

El sistema de gestión de riesgos debe revisarse y actualizarse periódicamente para garantizar su eficacia continua, así como la justificación y documentación de cualesquiera decisiones y acciones significativas adoptadas con arreglo al citado Reglamento. Este proceso debe garantizar que el proveedor determine los riesgos o efectos negativos y aplique medidas de mitigación de los riesgos conocidos y razonablemente previsibles de los sistemas de IA para la salud, la seguridad y los derechos fundamentales, habida cuenta de su finalidad prevista y de su uso indebido razonablemente previsible, incluidos los posibles riesgos derivados de la interacción entre el sistema de IA y el entorno en el que opera (cdo. 65 RIA). El sistema de gestión de riesgos debe adoptar las medidas de gestión de riesgos más adecuadas a la luz del estado actual de la técnica en materia de IA. Al determinar las medidas de gestión de riesgos más adecuadas, el proveedor debe documentar y explicar las elecciones realizadas y, cuando proceda, contar con la participación de expertos y partes interesadas externas. Al determinar el uso indebido razonablemente previsible de los sistemas de IA de alto riesgo, el proveedor debe tener en cuenta los usos de los sistemas de IA que, aunque no estén directamente cubiertos por la finalidad prevista ni establecidos en las instrucciones de uso, cabe esperar razonablemente que se deriven de un compor-

tamiento humano fácilmente previsible en el contexto de las características específicas y del uso de un sistema de IA concreto.

> Junto a ellas, cobra igual sentido la aplicación de sistemas de responsabilidad eficaces para vigilar y, cuando sea necesario, abordar de manera eficaz cualquier externalidad o impacto negativo derivado de la implementación de los sistemas de IA sobre los derechos fundamentales (Terrón y Domínguez, 2022: 153).
>
> Este enfoque normativo basado en el riesgo está ya presente en otras regulaciones europeas con extraordinarios resultados, como pone de relieve el art. 35 RGPD cuando establece la obligatoriedad de llevar a cabo una evaluación de impacto de protección de datos en función del riesgo del tratamiento (Lazcoz, 2020: 29 y Demetzou, 2019).

De esta forma, el art. 9 RIA no deja lugar a dudas cuando establece abiertamente la obligatoriedad de implantar, documentar y mantener un sistema de gestión de riesgos en relación con los sistemas de IA de alto riesgo, el cual debe concebirse como un proceso iterativo continuo, planificado y ejecutado durante todo el ciclo de vida de un sistema de IA de alto riesgo, que requerirá revisiones y actualizaciones sistemáticas periódicas.

> De conformidad con la redacción del propio art. 3.2 RIA, se entiende por riesgo «la combinación de la probabilidad de que se produzca un perjuicio y la gravedad de dicho perjuicio».

Así las cosas, antes de proceder al análisis de las diferentes etapas y requerimientos que debe reunir este sistema de gestión de riesgos, en primer término, debemos tener presente que el ciclo de vida de un sistema de IA, desde su génesis hasta su descarte, experimenta distintas fases que, a grandes rasgos, resultan comunes a todos los procesos de desarrollo tecnológico. No obstante, conviene precisar que, en función de la tecnología de IA que se pretenda implementar podrían existir algunos matices o particularidades, habida cuenta de la pluralidad y variedad de tecnologías subyacentes que envuelven este poderoso fenómeno. Sin embargo, para el objeto de este trabajo, solamente nos detendremos en el análisis de las siguientes etapas:

i. Concepción y análisis: en esta primera fase se fijarán los requisitos funcionales y no funcionales de la solución IA. Estos vendrán determinados por los objetivos de negocio derivados del tratamiento de datos en donde se incorporará o del mercado donde se pretende comercializar el componente. Incluirá los planes de proyecto, las restricciones normativas, cumplimiento de principios éticos, etc.

> La primera etapa en el ciclo de vida de la IA es la conceptualización y diseño del proyecto, la cual estará a cargo de su director, es decir, la persona responsable de tomar las decisiones y dirigir el esfuerzo que será ejecutado de manera colaborativa con un equipo multidisciplinar de la institución. Una buena conceptualización y diseño del proyecto asegurará su viabilidad, sostenibilidad y valor

público, además de que contribuirá a mitigar los riesgos que conlleva la aplicación de la herramienta de IA. En el proceso de conceptualización y diseño se pueden identificar una serie de pasos claves y preguntas esenciales a las que debe responder el responsable de la toma de decisiones antes de ejecutar un proyecto de IA, como son: (i) definición del problema: el primer paso en todo proyecto es definir claramente el problema del cual se pretende dar respuesta con la implementación de una herramienta de toma y/o soporte de decisiones basada en IA; (ii) análisis de prefactibilidad: después de definir el problema y determinar si la IA es la herramienta correcta para apoyar la solución, y antes de continuar con el proyecto, hay otras preguntas claves a las que se deberá responder con el objetivo de asegurar la viabilidad del proyecto para no desperdiciar los recursos de la institución; (iii) definición de objetivos: una vez que el proyecto se declara factible, es el momento de fijar los objetivos y sus métricas o indicadores, los cuales servirán para medir sus logros. Tales métricas deben reflejar el impacto esperado de la aplicación de la herramienta en la población destinataria. El logro de los objetivos deberá ayudar a solucionar el problema identificado; (iv) descripción de acciones: las acciones son las actividades que realiza la entidad y que permitirán articular la respuesta para resolver el problema; (v) mapeo de datos: se debe investigar si existen los datos necesarios y suficientes para llevar a cabo el proyecto, si la institución tiene acceso a las bases de datos o si se necesitarán convenios para obtenerlos. Un proyecto de IA se puede basar en datos tanto internos como externos, ya sean públicos o privados; (vi) definición del análisis y sus herramientas pertinentes: en esta etapa, el director del proyecto debe identificar preliminarmente qué tipo de análisis se requerirá para solucionar el problema. El tipo de análisis o la herramienta a implementar dependerá de la naturaleza del desafío y ayudará a mejorar los procesos de atención o respuesta necesarios. En esta etapa se trata de lograr una aproximación inicial que posteriormente debe consensuarse con el equipo técnico de proyecto; (vii) consideraciones éticas, legales y de gobernanza: incluso antes de comenzar con la ejecución del proyecto, el director deberá tener claros los desafíos éticos y legales que pueden surgir durante la implementación. Esto le permitirá adelantarse a posibles situaciones que puedan poner en riesgo la iniciativa y tomar las medidas de mitigación oportunas; y (viii) conformación del equipo responsable: en su calidad de director del proyecto, la persona responsable de la toma de decisiones debe conformar el equipo que se encargará de llevarlo a cabo. Los proyectos de IA no solo comprometen a quien decide y al equipo técnico; se requiere la participación de una variedad de áreas de la institución (por ejemplo, el equipo legal) e incluso de instituciones externas (por ejemplo, quienes poseen bases de datos útiles para el proyecto). Por consiguiente, el proceso de concepción y análisis de un proyecto debe ser iterativo. Aunque la idea es comenzar con una definición sólida del alcance del problema, esto puede cambiar si, por ejemplo, la institución no tiene la capacidad necesaria para actuar sobre él, o debe replantearse si los datos requeridos por no estar disponibles o ser insuficientes (Banco Interamericano de Desarrollo, 2021: 13-14).

Aunque la IA tiene un importante potencial para agilizar procesos y ampliar la capacidad del Estado, también hay que señalar que no es una bala de plata. Una vez definidos el problema y el tipo de intervención, es necesario contextua-

lizar y replantear el uso de la IA y el aprendizaje automático en consonancia con los principios de la IA aprobados por la OCDE, a saber: (a) la IA debe estar al servicio de las personas y del planeta, impulsando un crecimiento inclusivo, el desarrollo sostenible y el bienestar; (b) los sistemas de IA deben diseñarse de manera que respeten el Estado de derecho, los derechos humanos, los valores democráticos y la diversidad, e incorporar salvaguardias adecuadas, por ejemplo, permitiendo la intervención humana cuando sea necesario, con miras a garantizar una sociedad justa y equitativa; (c) los sistemas de IA deben estar presididos por la transparencia y una divulgación responsable a fin de garantizar que las personas sepan cuándo están interactuando con ellos y puedan oponerse a los resultados de esa interacción; (d) los sistemas de IA han de funcionar con robustez, de manera fiable y segura durante toda su vida útil, y los potenciales riesgos deberán evaluarse y gestionarse en todo momento; y (e) las organizaciones y las personas que desarrollen, desplieguen o gestionen sistemas de IA deberán responder de su correcto funcionamiento en consonancia con los principios precedentes.

Con la finalidad de dotar de efectividad estos principios, la OCDE recomienda a los gobiernos: facilitar una inversión pública y privada en investigación y desarrollo que estimule la innovación en una IA fiable; fomentar ecosistemas de IA accesibles con tecnologías e infraestructura digitales, y mecanismos para el intercambio de datos y conocimientos; desarrollar un entorno de políticas que allane el camino para el despliegue de unos sistemas de IA fiables; capacitar a las personas con competencias de IA y apoyar a los trabajadores con miras a asegurar una transición equitativa; y cooperar en la puesta en común de información entre países y sectores, desarrollar estándares y asegurar una administración responsable de la IA (Organización para la Cooperación y el Desarrollo Económico, 2019:3).

ii. Desarrollo: incluye diversas actuaciones, tales como la investigación, el prototipado, el diseño, la realización de pruebas, entrenamiento y validación. No todas las etapas estarán siempre presentes y su existencia quedará supeditado a la solución de IA concreta adoptada.

A la hora de evaluar si una solución de IA cumple los requisitos necesarios para efectuar un tratamiento de datos personales con todas las garantías, hay ciertos parámetros comunes a cualquier solución técnica que deberán quedar especificadas, como son: (a) precisión, exactitud o medidas de error requeridos por el tratamiento; (b) requisitos de calidad en los datos de entrada al componente IA; (c) precisión, exactitud o medidas de error efectivas de la solución IA en función de la métrica adecuada para medir la bondad de esta; (d) convergencia del modelo, cuando nos encontremos con entrenamiento y soluciones adaptativas; (e) consistencia entre los resultados del proceso de inferencia; y (f) predictibilidad del algoritmo, entre otros parámetros. Una solución técnica que no tenga respuesta a estas preguntas de una forma acreditable, no se podría considerar basada en una tecnología madura, sino en una tecnología sin capacidad de cumplir con los requisitos básicos de *accountability*, transparencia y legalidad exigidos en la actual regulación europea de protección de datos de carácter personal. Asimismo, el análisis de los parámetros antes enunciados permitirá identificar otros requisitos esenciales desde el punto de vista de protección de datos como puede ser, por

ejemplo, la aplicación del principio de minimización o el principio de calidad de los datos empleados (Latorre, 2021: 75-76).

iii. Explotación: comprende la ejecución de distintas acciones, algunas de las cuales podrán ejecutarse en paralelo: integración, producción, despliegue, inferencia, decisión, mantenimiento y evolución.

En todo caso, durante esta fase el responsable deberá: (i) garantizar la capacitación de los servidores que interactúan con el modelo de IA para que la herramienta sea sostenible en el tiempo; (ii) elaborar un manual de usuario dirigido a las personas que van a interactuar con el modelo; (iii) establecer mecanismos de retroalimentación para las personas que interactúan con el modelo; (iv) diseñar e implementar procesos simples para corregir aquellos errores presentes en el modelo que afecten a los usuarios; (v) establecer sistemas automatizados, o al menos periódicos, de monitoreo del modelo; (vi) mantener un registro de los resultados del modelo teniendo en cuenta las restricciones de acceso y seguridad; (vii) implementar mejoras necesarias al modelo y al proceso a partir de los hallazgos del monitoreo y la evaluación, y (viii) destinar los recursos necesarios para mantener la herramienta en el tiempo.

iv. Retirada final del tratamiento o componente.

En función del tipo de aplicación, algunas de las etapas anteriores podrían estar solapadas. Por ejemplo, la validación se podría solapar durante las etapas de desarrollo y explotación, o la etapa de evolución podría desarrollarse de manera simultánea a la etapa de inferencia.

Sentado lo anterior, y una vez clarificado el ciclo de vida de una solución de IA, conviene precisar las diferentes etapas de las que constará necesariamente este sistema de gestión de riesgos, el cual representa la piedra angular del instrumento jurídico europeo encargado de embridar el despliegue de esta poderosa herramienta:

a) la determinación y el análisis de los riesgos conocidos y previsibles que el sistema de IA de alto riesgo pueda plantear 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 art. 72 RIA;

d) la adopción de medidas adecuadas y específicas de gestión de riesgos diseñadas para hacer frente a los riesgos detectados.

Ahora bien, en este punto, conviene subrayar que los riesgos a los que se refiere el art. 9 RIA son únicamente aquellos que pueden mitigarse o eliminarse razonablemente mediante el desarrollo o el diseño del sistema de IA de alto riesgo o el suministro de información técnica adecuada.

Asimismo, conviene tener presente que la adopción de medidas adecuadas y específicas de gestión de riesgos diseñadas para hacer frente a los riesgos detectados, tal y como establece la literalidad del texto disponible del RIA, tendrán debidamente en cuenta los efectos y la posible interacción derivados de la aplicación combinada de los requisitos establecidos en materia de gobernanza y gestión de los datos, elaboración de documentación técnica, conservación de registros de acontecimientos, fomento de la transparencia, la supervisión humana y la garantía de un nivel adecuado de precisión, solidez y ciberseguridad; todo ello con vistas a reducir al mínimo los riesgos de manera más eficaz, al tiempo que se logra un equilibrio adecuado en la aplicación de las medidas para cumplir dichos requisitos.

Sin embargo, las medidas de gestión de riesgos mencionadas anteriormente considerarán aceptables los riesgos residuales pertinentes asociados a cada peligro, así como el riesgo residual general de los sistemas de IA de alto riesgo. A la hora de determinar las medidas de gestión de riesgos más adecuadas, se procurará:

a) eliminar o reducir los riesgos detectados y evaluados en la medida en que sea técnicamente viable mediante un diseño y un desarrollo adecuados del sistema de IA de alto riesgo;

b) implantar, cuando proceda, unas medidas de mitigación y control apropiadas que hagan frente a los riesgos que no puedan eliminarse;

c) proporcionar la información requerida conforme al art. 13 RIA y, cuando proceda, impartir formación a los responsables del despliegue.

Hace alusión a la necesidad de que los sistemas de IA de alto riesgo se diseñen y desarrollen de tal 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. A tal fin, 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.

Con vistas a eliminar o reducir los riesgos asociados a la utilización del sistema de IA de alto riesgo, se tendrán debidamente en cuenta los conocimientos técnicos, la experiencia, la educación y la formación que se espera que posea el responsable del despliegue, así como el contexto en el que está previsto que se utilice el sistema.

De igual forma, se preceptúa la obligatoriedad de que los sistemas de IA de alto riesgo sean sometidos a pruebas destinadas a determinar cuáles son las medidas de gestión de riesgos más adecuadas y específicas, requerimiento que permitirá comprobar que los sistemas de IA de alto riesgo funcionan de manera coherente con su finalidad prevista y cumplen los requisitos establecidos en la sección II del capítulo III de la norma objeto de estudio.

Paralelamente, se reconoce la posibilidad de que estos procedimientos de prueba puedan incluir pruebas en condiciones reales de conformidad con lo establecido en el art. 60 RIA, es decir, se dispone la oportunidad de que los proveedores o proveedores potenciales de sistemas de IA de alto riesgo puedan realizar pruebas de sistemas de IA de alto riesgo en condiciones reales fuera de los espacios controlados de pruebas para la IA cuando se cumplan de manera acumulativa todas las condiciones que se enumeran a continuación:

a) El proveedor o proveedor potencial ha elaborado un plan de la prueba en condiciones reales y lo ha presentado a la autoridad de vigilancia del mercado del Estado miembro en que se vayan a realizar las pruebas en condiciones reales.

b) La autoridad de vigilancia del mercado del Estado miembro en que se vayan a realizar las pruebas en condiciones reales ha aprobado las pruebas en condiciones reales y el plan de la prueba en condiciones reales; si la autoridad de vigilancia del mercado no responde en un plazo de treinta días, se entenderá que las pruebas en condiciones reales y el plan de la prueba en condiciones reales han sido aprobados.

c) El proveedor o proveedor potencial, con excepción de los proveedores o proveedores potenciales de sistemas de IA de alto riesgo mencionados en el anexo III, puntos 1, 6 y 7, en los ámbitos de la garantía del cumplimiento del Derecho, la migración, el asilo y la gestión del control fronterizo, así como de los sistemas de IA de alto riesgo mencionados en el punto 2 del anexo III ha registrado las pruebas en condiciones reales de conformidad con el art. 71.3 RIA, con un número de identificación único para toda la Unión y la información indicada en el anexo IX.

d) El proveedor o proveedor potencial que realiza las pruebas en condiciones reales está establecido en la Unión o ha designado a un representante legal que está establecido en la Unión.

e) Los datos recabados y tratados a efectos de las pruebas en condiciones reales únicamente se transferirán a terceros países si se aplican las garantías adecuadas y aplicables en virtud del Derecho de la Unión.

f) Las pruebas en condiciones reales no duran más de lo necesario para lograr sus objetivos y, en cualquier caso, no más de seis meses, que podrán prorrogarse por un período adicional de seis meses, con sujeción al envío de

una notificación previa por parte del proveedor o proveedor potencial a la autoridad de vigilancia del mercado, acompañada por una explicación de la necesidad de dicha prórroga.

g) Los sujetos de las pruebas en condiciones reales que sean personas pertenecientes a colectivos vulnerables debido a su edad o a una discapacidad cuentan con protección adecuada.

Cuando se implante el sistema de gestión de riesgos, los proveedores prestarán atención a si, en vista de su finalidad prevista, es probable que el sistema de IA de alto riesgo afecte negativamente a las personas menores de dieciocho años y, en su caso, a otros colectivos vulnerables.

h) Cuando un proveedor o proveedor potencial organice las pruebas en condiciones reales en cooperación con uno o varios responsables del despliegue o responsables del despliegue potenciales, estos últimos habrán sido informados de todos los aspectos de las pruebas que resulten pertinentes para su decisión de participar y habrán recibido las instrucciones de uso pertinentes del sistema de IA a que se refiere el art. 13 RIA; el proveedor o proveedor potencial y el responsable del despliegue o responsable del despliegue potencial alcanzarán un acuerdo en que se detallen sus funciones y responsabilidades con vistas a garantizar el cumplimiento de las disposiciones relativas a las pruebas en condiciones reales con arreglo al citado Reglamento y a otras disposiciones de Derecho de la Unión y nacional aplicable.

i) Los sujetos de las pruebas en condiciones reales han dado su consentimiento informado de conformidad con el art. 61 RIA o, en el ámbito de la garantía del cumplimiento del Derecho, en el que intentar obtener el consentimiento informado impediría que se probara el sistema de IA, las pruebas en sí y los resultados de las pruebas en condiciones reales no tendrán ningún efecto negativo sobre los sujetos, cuyos datos personales se suprimirán una vez realizada la prueba.

j) Las pruebas en condiciones reales son supervisadas de manera efectiva por el proveedor o el proveedor potencial y por los responsables del despliegue o los responsables del despliegue potenciales mediante personas adecuadamente cualificadas en el ámbito pertinente y con la capacidad, formación y autoridad necesarias para realizar sus tareas.

k) Se pueden revertir y descartar de manera efectiva las predicciones, recomendaciones o decisiones del sistema de IA.

A mayor abundamiento, se preconiza 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 pro-

babilidades previamente definidos que sean adecuados para la finalidad prevista del sistema de IA de alto riesgo.

Con todo ello, se articulan un conjunto de medidas y actuaciones extremadamente importantes que dan forma a un sistema integral de gestión de riesgos en lo que constituye, salvando las distancias, una suerte de Evaluación de Impacto Algorítmica, herramienta muy similar a la articulada y contemplada en el art. 35 RGPD, cuando se establece que «[c]uando sea probable que un tipo de tratamiento, en particular si utiliza nuevas tecnologías, por su naturaleza, alcance, contexto o fines, entrañe un alto riesgo para los derechos y libertades de las personas físicas, el responsable del tratamiento realizará, antes del tratamiento, una evaluación del impacto de las operaciones de tratamiento en la protección de datos personales. Una única evaluación podrá abordar una serie de operaciones de tratamiento similares que entrañen altos riesgos similares».

IV. Bibliografía

Agencia de los Derechos Fundamentales de la Unión Europea, *Construir correctamente el futuro: la inteligencia artificial y los derechos fundamentales,* Luxemburgo, 2021.

Banco Interamericano de Desarrollo, *Uso responsable de IA para política pública: manual de formulación de proyectos*, Washington DC, 2021.

Barrio Andrés, M. (2022). *Manual de Derecho Digital*. Editorial Tirant lo Blanch, Valencia, 2.ª edición.

Bini, S. (2022). «Reflexiones sobre justicia, humanidad y digitalización», en Llano Alonso, F. H. (dir.), *Inteligencia artificial y filosofía del derecho*. Editorial Laborum, Murcia.

Castellanos Claramunt, J. y Montero Caro, M. D. (2020). Perspectiva constitucional de las garantías de aplicación de la inteligencia artificial: la ineludible protección de los derechos fundamentales. *Ius et Scientia*, vol. 6, núm. 2.

Comisión Europea, *Plan coordinado sobre la inteligencia artificial,* Bruselas, 2018 [COM(2018) 795 final].

De Asís Roig, R. (2022). «Ética, tecnología y derechos», en Llano Alonso, F. H. (dir.). *Inteligencia artificial y filosofía del derecho*. Ediciones Laborum, Murcia.

Demetzou, K. (2019). Data Protection Impact Assessment: A Tool for Accountability and the Unclarified Concept of «High Risk» in the General Data Protection Regulation. *Computer Law & Security Review,* vol. 35, núm. 6.

García García, S. (2022). Una aproximación a la futura regulación de la inteligencia artificial en la Unión Europea. *Revista de Estudios Europeos,* vol. 79, núm. 1.

Gómez Abeja, L. (2022). «Inteligencia artificial y derechos fundamentales», en Llano Alonso, F. H. (dir.). *Inteligencia artificial y filosofía del derecho*. Editorial Laborum, Murcia.

Latorre, J. I. (2021). «Philosophical and ethical challenges of AI. The importance of awereness», en García Mexía, P. y Pérez Bes, F. (eds.). *Artificial Intelligence and the Law*. Editorial Wolters Kluwer, Madrid.

Lazcoz Moratinos, G. (2020). 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 Scientia,* vol. 6, núm. 2.

Organización para la Cooperación y el Desarrollo Económico, *Principios sobre la Inteligencia Artificial,* París, 2019 [OCDE/LEGAL/0449].

Parlamento Europeo, *Informe sobre la inteligencia artificial en la era digital,* Estrasburgo, 2022 [A9-0088/2022].

Parlamento Europeo, *Resolución del Parlamento Europeo, de 3 de mayo de 2022, sobre la inteligencia artificial en la era digital*, Estrasburgo, 2022 [2020/2266(INI)].

Parlamento Europeo, *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*, Estrasburgo, 2017, p. 7 [2016/2225(INI)].

Plaza Penadés, J. (2022). Fundamentos y principios jurídicos básicos de la inteligencia artificial y el «big data». *Revista Aranzadi de derecho y nuevas tecnologías*, núm. 59.

Sancho López, M. (2019). Estrategias legales para garantizar los derechos fundamentales frente a los desafíos del *big data*. *Revista General de Derecho Administrativo*, núm. 50.

Sarrión Esteve, J. (2020). «El derecho constitucional en la era de la inteligencia artificial, los robots y los drones», en Pérez Miras, A., Teruel Lozano, G. M., Raffiotta, E. y Pia Aidicicco, M. (dirs.). *Setenta años de la Constitución Italiana y cuarenta años de Constitución: Retos en el siglo XXI* (vol. 5). Editorial BOE-Centro de Estudios Políticos y Constitucionales, Madrid.

Soriano Arnanz, A. (2021). La propuesta de Reglamento de Inteligencia Artificial de la Unión Europea y los sistemas de alto riesgo. *RSR: Revista General de Derecho de los Sectores Regulados,* núm. 8.

Terrón Santos, D. y Domínguez Álvarez, J. L. (2022). *i-Administración pública, sistemas algorítmicos y protección de datos*. Editorial Iustel, Madrid.

Artículo 10. ***Datos y gobernanza de datos***

1. Los sistemas de IA de alto riesgo que utilizan técnicas que implican el entrenamiento de modelos de IA con datos se desarrollarán a partir de conjuntos de datos de entrenamiento, validación y prueba que cumplan los criterios de calidad a que se refieren los apartados 2 a 5 siempre que se utilicen dichos conjuntos de datos.

2. Los conjuntos de datos de entrenamiento, validación y prueba se someterán a prácticas de gobernanza y gestión de datos adecuadas para la finalidad prevista del sistema de IA de alto riesgo. Dichas prácticas se centrarán, en particular, en lo siguiente:

a) las decisiones pertinentes relativas al diseño;

b) los procesos de recogida de datos y el origen de los datos y, en el caso de los datos personales, la finalidad original de la recogida de datos;

c) las 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 supuestos, en particular en lo que respecta a la información que se supone que miden y representan los datos;

e) una evaluación de la disponibilidad, la cantidad y la adecuación de los conjuntos de datos necesarios;

f) el 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 posibles sesgos detectados con arreglo a la letra f);

h) la detección de lagunas o deficiencias pertinentes en los datos que impidan el cumplimiento del presente Reglamento, y la forma de subsanarlas.

3. Los conjuntos de datos de entrenamiento, validación y prueba serán pertinentes, suficientemente representativos y, en la mayor medida posible, carecerán de errores y estarán completos en vista de su finalidad prevista. Asimismo, tendrán las propiedades estadísticas adecuadas, por ejemplo, cuando proceda, en lo que respecta a las personas o los colectivos de personas en relación con los cuales está previsto que se utilice el sistema de IA de alto riesgo. Los conjuntos de datos podrán reunir esas características para cada conjunto de datos individualmente o para una combinación de estos.

4. Los conjuntos de datos tendrán en cuenta, en la medida necesaria para la finalidad prevista, las características o elementos particulares del entorno geográfico, contextual, conductual o funcional específico en el que está previsto que se utilice el sistema de IA de alto riesgo.

5. En la medida en que sea estrictamente necesario para garantizar la detección y corrección de los sesgos asociados a los sistemas de IA de alto riesgo de conformidad con lo dispuesto en el apartado 2, letras f) y g), del presente artículo, los proveedores de dichos sistemas podrán tratar excepcionalmente las categorías especiales de datos personales siempre que ofrezcan las garantías adecuadas en relación con los derechos y las libertades fundamentales de las personas físicas. Además de las disposiciones establecidas en los Reglamentos (UE) 2016/679 y (UE) 2018/1725 y la Directiva (UE) 2016/680, para que se produzca dicho tratamiento deben cumplirse todas las condiciones siguientes:

a) que el tratamiento de otros datos, como los sintéticos o los anonimizados, no permita efectuar de forma efectiva la detección y corrección de sesgos;

b) que las categorías especiales de datos personales estén sujetas a limitaciones técnicas relativas a la reutilización de los datos personales y a medidas punteras en materia de seguridad y protección de la intimidad, incluida la seudonimización;

c) que las categorías especiales de datos personales estén sujetas a medidas para garantizar que los datos personales tratados estén asegurados, protegidos y sujetos a garantías adecuadas, incluidos controles estrictos y documentación del acceso, a fin de evitar el uso indebido y garantizar que solo las personas autorizadas tengan acceso a dichos datos personales con obligaciones de confidencialidad adecuadas;

d) que las categorías especiales de datos personales no se transmitan ni transfieran a terceros y que estos no puedan acceder de ningún otro modo a ellos;

e) que las categorías especiales de datos personales se eliminen una vez que se haya corregido el sesgo o los datos personales hayan llegado al final de su período de conservación, si esta fecha es anterior;

f) que los registros de las actividades de tratamiento con arreglo a los Reglamentos (UE) 2016/679 y (UE) 2018/1725 y la Directiva (UE) 2016/680 incluyan las razones por las que el tratamiento de categorías especiales de datos personales era estrictamente necesario para detectar y corregir sesgos, y por las que ese objetivo no podía alcanzarse mediante el tratamiento de otros datos.

6. Para el desarrollo de sistemas de IA de alto riesgo que no empleen técnicas que impliquen el entrenamiento de modelos de IA, los apartados 2 a 5 se aplicarán únicamente a los conjuntos de datos de prueba.

José López Calvo
Doctor en Derecho. Administrador Civil del Estado

I. Consideraciones previas

En palabras de Inés HUERTAS, experta en *Big Data* y *Machine Learning*: «En esto de la IA no sirve el "Si no sabes adónde vas, cualquier camino te llevará allí" como decía Alicia en el País de las Maravillas».

La IA nos permite tomar decisiones basadas en datos, en cantidades masivas de datos. No en intuiciones subjetivas. Debe rehuirse la tentación de humanizarla o antropomorfizarla. Su funcionamiento se basa en que es un gran cerebro analítico con capacidad de procesar grandes volúmenes de datos que se convierten, así, en la clave. No tienen objetivos, ni creencias, ni deseos, ni comprensión. Es un resultado de su diseño de programación y de la información que se le suministra en forma de datos, que únicamente son útiles si permiten derivar en conclusiones adecuadas.

A los principios y exigencias en materia de gobernanza de datos en la IA se remite el artículo 10 del RIA, que se complementa —y viceversa— con el artículo 5 del RGPD, intitulado «principios relativos al tratamiento de datos personales», con el que se hermana, convirtiéndose en la práctica en una «ley especial» que concreta, matiza y, en algunos casos, redimensiona e incluso modifica sus previsiones para los tratamientos de datos realizados por la IA.

II. Concordancias

Para la aplicación del presente artículo, véase también:

– Principios del art. 5 del RGPD.

– Art. 22 del RGPD.

– Derecho de explicación de las decisiones automatizadas recogido en el RGPD en el considerando 63 y vinculado al principio de transparencia (arts. 2.f), 14.2.g) y 15.1.h) RGPD).

III Comentario

1. El RIA como complemento y especificación de los principios del artículo 5 del RGPD

A) Artículo 5.b) del RGPD: principio de limitación de la finalidad

El principio de limitación de la finalidad del RGPD exige que el tratamiento de datos no se realice de manera incompatible con los fines para los que fueron captados.

Con frecuencia, el principio está claramente salvaguardado por la IA. Los datos se tratan para la misma finalidad por Netflix recomendando qué otras películas pueden ver el usuario en base a su histórico de películas vistas y a su valoración; Spotify por su estilo de música y Amazon recomendando productos en base a los patrones de compra del cliente. Se utilizan los datos con una finalidad manifiestamente compatible.

Pero dudosamente se concilia sin esfuerzos con técnicas habituales en la IA. Como el denominado *web scraping*, técnica que consiste en automatizar la extracción de datos presentes en Internet. Ya hay demandas contra el sistema ChatGPT de OpenAI por vulneración de datos personales, en cuanto

que el sistema habría «escrapeado» datos personales de manera no autorizada. También es suficientemente reveladora la suspensión reciente, por otras agencias además de la AEPD, de la práctica de Worldcoin de escanear el iris de personas voluntarias a cambio de una compensación económica para «entrenar nuestros modelos de IA», sin cumplir con el principio de limitación y finalidad, por no ser determinada y específica. Y ya han recaído sanciones, como la que en marzo de 2024 se impuso a Google por la autoridad de competencia francesa por 250 millones de euros por entrenar sus modelos de IA con contenidos de medios y agencias de prensa francesas en lo que puede ser un precedente extrapolable.

Cuantos más datos disponga más exacta será la IA, pero la normativa de protección de datos no contempla un uso incondicionado. Como es sabido, el RGPD suprime el, hasta su aprobación vigente, concepto de «fuentes de acceso público» de uso sin límites ni condiciones (censo promocional, medios de comunicación social…). Aun tratándose de información recogida en fuentes públicas, el RGPD impone respetar el principio de limitación de la finalidad y atender a las razones concretas que han determinado la publicación de la información. Y lo corrobora el artículo 10.4 del RIA que vincula la finalidad al entorno geográfico, contextual, conductual o funcional específico del sistema de IA.

Lo que crea una intersección que, como otras, deberá requerir una consolidación técnica, doctrinal, jurisprudencial e institucional para que se despejen dudas y límites. Superando situaciones como las que declaraba en una entrevista la Vicepresidenta de Tecnología de OpenAI manifestando que ni siquiera ella sabe que datos se usan para alimentar Sora o Shutterstrock. Quedando así a la espera de la interpretación que se haga del «interés legítimo» —que se ha asumido como base legitimadora para el tratamiento de datos por los buscadores sin consentimiento de los «buscados» (STJUE, Gran Sala, de 13 de mayo de 2014, asunto C-131/12, caso *Google. AEPD. Mario Costeja*)—, la consideración como compatible del tratamiento ulterior de los datos personales con fines de archivo en interés público o fines de investigación científica (art. 5.1 b RGPD) y, en definitiva, del diseño de un equilibrio entre regulación e innovación y servicios.

B) Artículo 5.1.c) del RGPD: principio de exactitud

Según el principio de calidad del dato del RGPD, los datos personales han de ser exactos, veraces, actuales y no excesivos. Principio no aplicable en exclusiva a la IA, pero que, de nuevo, debe dimensionarse con la exigencia en la gobernanza de la RIA del uso de datos pertinentes, suficientemente representativos y actualizados.

El RIA abunda en tal previsión y la perfecciona al exigir su dimensionamiento respecto al proyecto en cuestión que se acometa. Sin datos adecuados, no solo no se cumple el principio de exactitud del RGPD (y otros como minimización o

protección de datos por diseño o por defecto) sino que no hay inteligencia artificial adecuada. La disponibilidad y la calidad del dato es imprescindible para comenzar un proyecto. No únicamente por imperativo del RGPD, sino por la lógica de la IA recogida en el RIA. Para lo que el RIA da criterios sobre las características que deben tener los datos:

– Actualizados: una IA, por ejemplo, para predecir la probabilidad de un caso judicial no puede entrenarse con sentencias judiciales basadas en leyes obsoletas. Y debe someterse a actualización constante, evitando el efecto «fosilización» de sus predicciones.

– Representativos: por ejemplo, una IA que brinde recomendaciones personalizadas sobre la base de análisis de los hábitos y comportamientos de los usuarios (p. ej. relevancia de productos ecológicos o hábitos de compra *on line*) solo puede sostenerse sobre la disposición ingente de información sobre los mismos.

– Etiquetado de datos: esencial para el buen funcionamiento, en especial en los casos de *machine learning* en que el algoritmo aprende sobre la base de las información clasificada que se le suministra: para identificar imágenes de perros y gatos, si un correo electrónico es *spam* o no, etc. Crítico y complicado en la información sanitaria recogida en historiales médicos en que la supervisión humana por sanitarios desbordados de trabajo es imprescindible para clasificar la información: tipo de tumor, síntomas, etc.

La exigencia del artículo 10 del RIA de que los datos con los que se entrene la IA deben ser veraces, exactos y representativos de la realidad la complementa el propio artículo al recordar que debe evitar sesgos (art. 10.2.f) RIA) y alucinaciones:

– Sesgos estadísticos: que no reflejen la distribución de la población relevante. Por ejemplo, si la IA realiza un análisis de incidencia de Alzheimer basándose principalmente en datos de población menor de 50 años. O la propia base de datos Imagenet, el principal algoritmo de visión artificial, alimentado en un 45 % de imágenes procedentes de Estados Unidos, país que representa el 4 % de la población. O algoritmos médicos que analizan una insuficiente base de datos de historiales médicos estructurados.

– Sesgos motivados por el suministro de información incorrecta: la Comisión en marzo de 2024 ha requerido información a Bing, Google Search (VLOSEs), Facebook, Instagram, Snapchat, TikTok, YouTube, and X (VLOPs) sobre sus medidas para mitigar riesgos de la IA generativa con respecto a las alucinaciones de la misma. Errores que hacen que estos sistemas y conjuntos algorítmicos interpreten datos o señales de manera incorrecta, lo que se puede traducir en la entrega de información inexacta. Como ocurrió en mayo de 2023, cuando un abogado estadounidense presentó en Estados Unidos

escritos con múltiples referencias y sentencias falsas creadas por la inteligencia artificial que había utilizado para redactar sus escritos de defensa.

Evitando la información inexacta que derive, por ejemplo, en resultados de tanta gravedad como diagnósticos médicos erróneos, manipulación o usos malintencionados creando *deep fakes*. O riesgos para la integridad física por accidentes provocados por una mala configuración del algoritmo de un coche autónomo que interprete indebidamente los objetos de la carretera.

– Sesgos sociales: por no adaptar los datos a considerar a los grupos sociales afectados, derivando en resultados poco representativos. Como ocurrió en algunas IA de COVID que no consideraban su especial magnitud en zonas de menor renta, no solo por hacinamiento, sino por mayor presencia de enfermedades como diabetes o neumonía.

Cuya posible trascendencia política se ha puesto de manifiesto con la dimisión en 2021 del Gobierno de los Países Bajos por un escándalo en ayudas sociales en que miles de familias extranjeras fueron acusadas de fraude de forma injusta y obligadas a devolver importantes cantidades, sumiendo a núcleos familiares en la más absoluta ruina.

Lo que puede llevar a verdaderas «alucinaciones» de la IA, como ocurre con el nuevo programa de IA de Adobe, Firefly, que no ha comprendido correctamente la diversidad racial aplicada artificiosamente y muestra imágenes de nazis negros.

– Sesgos de género: sin tener en cuenta las diferencias de género a la hora de entrenar a la IA con datos de procedencia indistinta sin tener en cuenta aspectos diferenciadores por sexo: por ejemplo, en los diferentes síntomas en hombres y mujeres de un ataque de corazón, o sin tener en cuenta que las mujeres padecen más trastornos de las extremidades superiores relacionados con los movimientos repetitivos, el cáncer ocupacional es más común entre los hombres que entre las mujeres, el asma y las alergias suelen ser más comunes entre las mujeres, etc.

C) Artículo 5.1.e) del RGPD: limitación del plazo de conservación

Es claro que la opacidad de las grandes empresas —no consta ninguna inspección en profundidad en los servidores de ninguna de las componentes de GAFAM— impide la comprobación fehaciente de que se cumple el principio de «limitación del plazo de conservación» del RGPD reiterado en el artículo 10.5.e) del RIA: supresión de los datos cuando sean obsoletos. Límites de comprobación también aplicables a la exigencia en el mismo apartado de no transmisión a terceros sin base legítima.

D) Artículo 5.1.f) del RGPD: principio de integridad y confidencialidad. En especial, seudonimización y anonimización

Principalmente recogidos en el artículo 10.5 del RIA en los apartados a) a f) referido a categorías especiales de datos, aunque los declara «además» de los previstos en el RGPD.

Porque la IA introduce matices.

Así, los principios de seudonimización y anonimización colisionan con la realidad de que información como el ADN presenta problemas de anonimización. Aunque el ADN de un posible criminal no se encuentre incluido en las grandes bases de datos, se puede identificar (se calcula que con un 60 % de los casos) al constar en las mismas la de familiares que permiten la identificación mediante un análisis del perfil genético. También se ha demostrado posible a partir de resonancias craneales magnéticas empleando un algoritmo de identificación de caras gratuito basado en IA. La IA, en definitiva, es un instrumento de enorme potencia, también para desvelar identidades. Creando una encrucijada que también emplaza al Reglamento del Espacio Europeo de Datos sanitarios que permite que datos de salud anonimizados o seudonimizados se compartan por interés público por fines (el llamado uso secundario) de investigación, innovación, formulación de políticas, educación y seguridad del paciente.

Lo que exige profundizar en técnicas como el «aprendizaje federado» que entrena IA por separado en centros distintos combinando la información *a posteriori* de manera que los datos nunca abandonan el centro donde se encuentran, en favor de la privacidad.

2. El RIA y el registro de actividades de tratamiento

El artículo 10.5.f) del RIA introduce la exigencia, no recogida en el RGPD (art. 30 RGPD), de incluir en el registro de actividades del tratamiento (RAT) las razones por las que el tratamiento de categorías especiales de datos personales era estrictamente necesario para detectar y corregir sesgos, y por las que ese objetivo no podía alcanzarse mediante el tratamiento de otros datos.

El RIA profundiza de manera exigente en el requerimiento del RGPD de que el RAT refleje los «fines del tratamiento» y traslada al mismo la exigencia de acreditar los principios de gobernanza de la RIA.

3. RGPD, RIA, derecho de explicación y derecho a no someterse a decisiones automatizadas

El Magistrado Manuel Marchena manifestaba, con ocasión de la aprobación del RIA, lo siguiente: «Tenemos que ser conscientes de que todo lo que está regulado por la inteligencia artificial está sometido a un fenómeno de obsolescencia jurídica que neutraliza la validez de nuestras fórmulas jurídicas. Siempre vamos a llegar tarde».

El artículo 10.2.a) del RIA recoge, entre las «prácticas de gobernanza y gestión de datos adecuadas para la finalidad», la siguiente: «a) las decisiones pertinentes relativas al diseño».

Pues bien, los atisbos de alarma social ante una innovación, en este caso la IA, y su implantación van a requerir tiempo hasta que se ultime el aserto de M. Weiser: «las tecnologías más destacadas son las que desaparecen», pasan a formar parte de nuestra vida inadvertidamente. Proceso indudablemente en marcha en que ya coexistimos con la IA que nos enfrenta a múltiples dilemas.

Tal desafío de conciliación se encontrará, entre otras, en dos previsiones: la obligación de explicar las decisiones algorítmicas y de no ser sometidos a decisiones exclusivamente algorítmicas que colisionan con la opacidad algorítmica:

A) Derecho de explicación

El artículo 86 del RIA y varios considerandos del mismo (cdos. 93, 107, 171) prevén que «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» (cdo. 171 RIA).

El RIA insiste y refuerza el derecho de explicación de los tratamientos automatizados recogido en el RGPD, que implica que aquel que realice tal tipo de análisis debe encontrar formas sencillas de informar al titular de los datos sobre los motivos subyacentes, sobre los criterios en los que se basa para tomar la decisión.

B) Derecho a no ser sometido a decisiones automatizadas o scoring

La aplicabilidad del régimen de protección de datos cuando haya un tratamiento de datos mediante sistemas de IA no excluye el particular régimen y garantías del artículo 22 del RGPD: 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.

Derecho «ampliado» por la STJUE de 7 de diciembre de 2023, asunto C-634/21, caso *Schufa*, que llegaba a incluir las «predicciones de probabilidad» aunque no sean «decisiones» propiamente dichas, sino pronósticos, aunque no tengan efectos directos y exista una mayor o menor intervención humana, si el tercero que «decide» lo hace conforme a la predicción; es decir, la respeta.

Sentencia que parece anticiparse al RIA en que muchas de los productos IA no adoptarán la forma de decisiones sino de «aproximaciones».

El RIA confirma que la IA basada en la puntuación estará prohibida y se acompasa con la decisión del TJUE en el citado caso *Schufa*.

Los sistemas de IA basados o que deriven en puntuaciones individuales —como el Sistema de Crédito Social del gobierno de China para rastrear y evaluar el comportamiento de individuos y empresas— quedan expresamente prohibidos por el RIA: el *scoring* social, la evaluación o clasificación de individuos o grupos de individuos en función de su comportamiento social o de sus características personales, causándoles un trato perjudicial o desfavorable (art. 5.1.c) RIA).

C) Derechos y opacidad

Cuando la IA decide —como ya hace— la lista de espera hospitalaria y quien recibe tratamiento médico preferente, se convierten en verdaderas programaciones normativas al adoptar decisiones con trascendencia frente a terceros. Incluso en el ámbito judicial, como ya se planteó en el conocido caso *State versus Loomis* en 2013 en la Corte Suprema del Estado de Wisconsin. Ante la alegación de que era secreto el algoritmo que atribuía al señor Loomis «riesgo elevado de reincidencia» y de cometer actos violentos, que solo lo conocía la empresa que lo había desarrollado, los jueces argumentaron que, en definitiva, el programa informático se había basado únicamente en los factores habituales para medir la peligrosidad criminal futura como, por ejemplo, huir de la policía y el historial delictivo previo.

Los algoritmos se convierten así en una posible y novedosa fuente del Derecho, por lo que se les exige las mismas garantías de cualquier otro acto administrativo: control, motivación o transparencia. Extrapolable a lo privado por imperativo del RGPD y el RIA.

Lo que encuentra una nueva encrucijada en la actual prevalencia del principio de opacidad algorítmica (caja negra). Como resolvió en España el Consejo de Transparencia respecto al programa BOSCO, un *software* desarrollado por orden del Gobierno que las eléctricas utilizan para decidir quién es beneficiario del bono social que aplica un descuento en la factura de la luz a las familias vulnerables.

La fundación Civio planteó una reclamación alegando que se trataba de un programa incorrectamente diseñado que negaba ayudas a personas que tendrían derecho, en especial jubilados y viudas que se quedaban sin bono por decisión del algoritmo.

Pues bien, el Consejo de Transparencia no estimó en su resolución la remisión del código fuente, el programa que permitiría conocer su funcionamiento, al entender que, en este caso, se aplicaba el artículo que establece el secreto profesional y la propiedad intelectual e industrial como uno de los límites al derecho de acceso a la información amparada por transparencia. Resolución confirmada en sede judicial donde a la propiedad intelectual añade razones de «seguridad pública y defensa nacional».

Opacidad que introduce claros elementos de tensión, en especial para poder comprobar el grado de cumplimiento de los principios de gobernanza de datos del artículo 10 del RIA en cada sistema IA. Creándose una de las principales encrucijadas y retos al que se enfrenta y nos enfrenta la IA.

IV. Recursos complementarios

Gobierno Federal de Estados Unidos. *Memorandum sobre gobernanza, innovación y control de riesgos en el uso por las Agencias de Inteligencia Artificial,* de 28 de marzo de 2024.

V. Bibliografía

Cobo Cano, M. y Lloret Iglesias, L. (2023). *Inteligencia Artificial y Medicina*. Editorial Catarata/CSIC, Madrid.

Degli-Esposti, S. (2023) *La ética de la Inteligencia Artificial*. Editorial Catarata/CSIC, Madrid.

Peralta Gutiérrez, A. (2023). La estrecha conexión entre la inteligencia artificial y la privacidad. *Revista LA LEY Derecho digital e innovación*, núm. 18.

Artículo 11. ***Documentación técnica***

1. 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 conformidad del sistema de IA con dichos requisitos. Contendrá, como mínimo, los elementos contemplados en el anexo IV. Las pymes, incluidas las empresas emergentes, podrán facilitar los elementos de la documentación técnica especificada en el anexo IV de manera simplificada. A tal fin, la Comisión establecerá un formulario simplificado de documentación técnica orientado a las necesidades de las pequeñas empresas y las microempresas. Cuando una pyme, incluidas las empresas emergentes, opte por facilitar la información exigida en el anexo IV de manera simplificada, utilizará el formulario a que se refiere el presente apartado. Los organismos notificados aceptarán dicho formulario a efectos de la evaluación de la conformidad.

2. Cuando se introduzca en el mercado o se ponga en servicio un sistema de IA de alto riesgo asociado a un producto que entre en el ámbito de aplicación de los actos legislativos de armonización de la Unión mencionados en el anexo I, sección A, se elaborará un único conjunto de documentos técnicos que contenga toda la información mencionada en el apartado 1, así como la información que exijan dichos actos legislativos.

3. La Comisión estará facultada para adoptar actos delegados con arreglo al artículo 97 al objeto de modificar el anexo IV, cuando sea necesario, para garantizar que, en vista de los avances técnicos, la documentación técnica proporcione toda la información necesaria para evaluar si el sistema cumple los requisitos establecidos en la presente sección.

Javier Puyol Montero
Doctor en Derecho. Abogado. Magistrado Excedente. Letrado del Tribunal Constitucional

I. Consideraciones previas

En el presente apartado se hace referencia introductoria al concepto de documentación técnica con relación a los sistemas de IA, así como a las exigencias formales de dicha documentación, que debe acompañar a la puesta en marcha de cualquier sistema de IA, y al mismo tiempo, al contenido material que debe reunir dicha documentación de conformidad con las exigencias establecidas en el presente artículo 11 del Reglamento Europeo de Inteligencia Artificial, en relación con las especificaciones contenidas en el anexo IV del mismo. Al mismo tiempo, se hace una especial mención a las previsiones regulatorias especificas establecidas en dicho Reglamento con relación a las Pymes y a las Micropymes, a los efectos de facilitar el efectivo cumplimiento de estos requisitos. Y, por último, se hace referencia a determinadas previsiones normativas vinculadas a la necesidad de armonizar la regulación en materia de documentación técnica derivada de los sistemas de IA, a los efectos de hacer compatible la misma, dada la diversidad existente de dicha normativa.

II. Concordancias

Para la aplicación del presente artículo, véase también:

– Art. 73 del RIA.

– Anexo II, parte I del RIA, que versa sobre la lista de la legislación de armonización de la Unión basada en el nuevo marco legislativo.

– Anexo IV del RIA, que precisa la documentación técnica relacionada en el artículo 11 del RIA.

III. Comentario

1. Introducción

En una aproximación al contenido del artículo 11 del Reglamento Europeo de Inteligencia Artificial, debe partirse del concepto de «documentación técnica» obrante, precisamente, en los sistemas de inteligencia artificial, la cual puede ser considerada como el conjunto de documentos que describen en detalle todos los aspectos técnicos, éticos y legales de un sistema de inteligencia artificial.

La documentación técnica de un sistema de IA es un componente esencial para comprender su funcionamiento, sus capacidades, pero también las limitaciones con las que cuenta dicho sistema. Esta documentación tiene que reunir una serie de requisitos esenciales, en el sentido de que la misma tiene ser necesariamente clara, completa, y precisa, y a través de la misma debe permitir a las autoridades competentes poder evaluar adecuadamente la conformidad del sistema, con los requisitos establecidos en el citado RIA. Consecuentemente con ello, la documentación técnica puede ser considerada como un componente fundamental para garantizar la seguridad y los estándares éticos con los que debe contar un sistema de inteligencia artificial de alto riesgo, entendiendo como tales aquellos que pueden generar un impacto significativo en la vida y en los derechos de las personas, o en la sociedad en general, y, por ello, es sumamente importante que estos sistemas se regulen de forma adecuada para garantizar que se utilizan de forma segura, responsable y ética.

Debe tenerse presente que la responsabilidad de la elaboración de la documentación técnica recae en el fabricante o desarrollador del sistema de IA de alto riesgo. En algunos casos, puede ser necesario incluso recurrir a un experto en IA para la elaboración de la indicada documentación. Su redacción debe ser efectiva en un momento anterior a su introducción en el mercado o puesta en servicio. Del mismo modo, la documentación técnica debe ser actualizada de forma regular para reflejar cualquier cambio en el sistema, como por ejemplo aquellas modificaciones que puedan producirse en el diseño, en los datos utilizados o en el entorno operativo utilizado. También es necesario actualizar la documentación en caso de que se identifiquen nuevos riesgos o se produzcan cambios significativos en la normativa que regule el funcionamiento del sistema de IA.

2. La documentación técnica

La documentación técnica, según se desprende del apartado 1.º de dicho artículo 11 del RIA, relativa a un sistema de IA de alto riesgo, tiene que elaborarse de manera necesaria antes de su introducción en el mercado, incluso previamente a su comercialización o puesta en servicio, y se tiene que mantener actualizada, tal como se ha indicado anteriormente. Ello es consecuencia de un principio de seguridad jurídica que vincula tanto a dichos sistemas, como al hecho consistente en que, de manera efectiva, se dé cumplimiento a dichas prescripciones de carácter técnico, de modo y manera que el sistema responda plenamente a las mismas y a sus características previamente definidas.

A tal efecto, el Reglamento, en el apartado 2.º de dicho precepto, señala la necesidad de que dicha documentación técnica se elabore de forma que demuestre de manera real y efectiva, que el sistema de IA de alto riesgo cumpla con los requisitos establecidos en el Reglamento, y que simultáneamente, también proporcione a las autoridades nacionales competentes y a los organismos notificados, aquella información necesaria para evaluar la conformidad del sis-

tema de IA con el cumplimiento de los requisitos indicados. Dicha información tiene que ser proporcionada, de acuerdo con el indicado texto normativo, de manera clara y completa, ya que su finalidad es poder contrastar la misma con los requisitos establecidos al efecto en el anexo IV del citado Reglamento, y consecuentemente con ello, verificar su cumplimiento. En este sentido, debe indicarse que dichos requisitos técnicos tienen un carácter de mínimo, a los efectos de poder justificar de manera ajustada a Derecho, que el sistema de IA cumple efectivamente con las prescripciones establecidas en el RIA, a los efectos de su homologación, una vez que se haya procedido a evaluar su conformidad con los indicados requisitos de carácter técnico.

En otro orden de cosas, dentro de la documentación técnica de un sistema de IA se debe contener e incluir, al menos, la siguiente información, que haga referencia a los siguientes aspectos relativos a la misma:

a) En primer término, la información general del sistema, que describe los aspectos básicos del sistema, como su propósito, su funcionamiento y los actores involucrados en su desarrollo y uso. Esta información tiene un carácter fundamental, a los efectos de que por parte de las autoridades competentes puedan evaluar efectivamente la conformidad del sistema con los requisitos legales y también con los de carácter éticos. La información general del sistema es importante por el hecho de que permite a las autoridades competentes comprender el sistema y su funcionamiento, facilitando al mismo tiempo la evaluación de los riesgos asociados al sistema. También debe destacarse que, a través de la misma, se ayuda a determinar si el sistema cumple con los requisitos legales y éticos. En este mismo sentido, permite la trazabilidad del desarrollo del sistema y su evolución, y por último, se suele indicar que contribuye a la transparencia y la responsabilidad en el uso de la IA.

De manera concreta, la información general del sistema tiene que hacer una puntual referencia a determinados aspectos del mismo, entre los que cabe destacar los siguientes:

a.1) La identificación del sistema, donde se contemplan aspectos tales como: (i) nombre; (ii) versión; y (iii) fecha de desarrollo o de última actualización.

a.2) Su propósito o finalidad, en la que se hace referencia a: (i) la descripción de los objetivos que persigue el sistema; y (ii) los casos de uso previstos por el mismo.

a.3) La descripción técnica del sistema, en la que se tiene que indicar. (i) la arquitectura del sistema, incluyendo sus componentes principales (ii) las tecnologías utilizadas para el desarrollo del mismo; y finalmente (iii) los datos utilizados por dicho sistema.

a.4) La determinación de los actores que se encuentran involucrados, entre los que cabe mencionar: (i) la identificación del fabricante o desarrollador del

sistema; (ii) la identificación de los propietarios o usuarios del mismo; y (iii) la identificación de los responsables de la gestión de dicho sistema.

b) El segundo aspecto que debe incluir la documentación técnica es el atinente a la llamada «evaluación de riesgos», la cual es importante ya que permite a las autoridades competentes evaluar la seguridad y la ética del sistema. Además, ayuda a los desarrolladores y usuarios del sistema a tomar decisiones informadas sobre su uso, y contribuye a la prevención de daños y a la protección de los derechos de las personas.

La evaluación de riesgos debe ser realizada por un equipo multidisciplinar en la que tengan la intervención de profesionales con experiencia en el ámbito de la IA, en el ámbito de la seguridad, y en aspectos vinculados con la ética y el derecho. Al mismo tiempo, es necesario poner de manifiesto que la utilización de métodos y herramientas adecuados para poder llevar a cabo la identificación, el análisis y la evaluación de los riesgos. También debe tenerse presente que la evaluación de riesgos debe ser un proceso continuo, que se debe actualizar de manera periódica.

A los efectos del efectivo cumplimiento de estos objetivos, la evaluación de riesgos tiene que contener los elementos que se determinan a continuación:

b.1) La identificación de riesgos, la cual se refiere a la necesidad de proceder a la identificación de todos los riesgos potenciales asociados al sistema de IA, entre los que se debe incluir: (i) los riesgos para la seguridad, como por ejemplo los fallos en el sistema que puedan causar daños físicos o económicos; (ii) los riesgos de discriminación que conllevan sesgos en el sistema que puedan discriminar a ciertos grupos de personas; (iii) los riesgos para la privacidad, que implican la posibilidad de la recopilación y el uso indebido de datos personales; y, finalmente (iv) los riesgos para la ética, que representan el uso del sistema de IA para fines propiamente contrarios a la ética o los derechos humanos.

b.2) El análisis de los riesgos, que determina la necesidad de realizar un análisis de cada riesgo identificado, con la finalidad de poder determinar su probabilidad de ocurrencia y su impacto potencial.

b.3) La evaluación de los riesgos, que consiste en valorar la importancia de cada riesgo, en función de su probabilidad de ocurrencia y lo que lleva consigo su impacto potencial.

b.4) Las medidas de mitigación para aquellos riesgos que previamente hayan sido identificados como tales. Así, dentro de las posibles medidas a adoptar, cabe citar principalmente las siguientes: (i) las medidas técnicas, donde es especialmente importante traer a colación el uso de técnicas caracterizadas por su seguridad y por su robustez; (ii) las medidas de tipo organizativo, tales como la formación y capacitación de los usuarios, así como la definición de políticas de uso de la IA. (iii) Finalmente, hay que hacer especial referencia a las llamadas «medi-

das éticas», las cuales se encuentran dirigidas a la definición de principios éticos para el desarrollo y uso del sistema.

b.5) Por último, dentro de esta evaluación de riesgos, es determinante llevar a cabo un análisis del impacto que tiene el sistema de IA, tanto desde una perspectiva de carácter social, como puramente ética, a los efectos de determinar las implicaciones que el mismo conlleva en estos ámbitos de actuación

c) El tercer aspecto que debe incluir la documentación técnica del sistema de IA es la que se refiere al uso y manejo tanto de los datos, como de los algoritmos por parte del mismo. En lo que hace referencia a la importancia de los datos para el funcionamiento de dicho sistema, debe tenerse presente que los datos hoy en día ya representan la base fundamental de cualquier sistema de IA. La calidad y cantidad de los datos utilizados en el entrenamiento del sistema van a determinar en gran medida su rendimiento y su evolución. Los beneficios de documentar técnicamente los datos y los algoritmos que se van a emplear en el sistema de IA se concretan en su mayor facilidad de comprensión del funcionamiento del mismo. A ello debe serle añadido la posibilidad de reproducir los resultados de dicho sistema También representan beneficios muy ostensibles, la existencia de una mayor facilidad para el mantenimiento y la mejora del sistema de IA, así como la generación de confianza en la comunidad sobre el indicado sistema.

Consiguientemente con ello, la documentación técnica debe incluir información detallada sobre los datos utilizados, incluyendo aspectos, tales como:

c.1) El origen, relativo a donde se obtuvieron los datos que alimentan el sistema.

c.2) Las características, las cuales hacen referencia a qué tipos de variables y valores se encuentran en los datos.

c.3) El llamado «preprocesamiento», que implica las transformaciones que se han se aplicado a los datos antes de su uso por parte del sistema de IA.

c.4) Los sesgos, en el sentido de que los datos contienen tendencias o inclinaciones ideológicas conocidas, que deben ser documentadas y explicadas de manera adecuada.

Dentro de este tipo de documentación técnica, también tiene que hacerse una referencia puntual a los algoritmos, que representan las instrucciones que le permiten al sistema de IA poder procesar los datos y generar los resultados. La documentación técnica debe explicar en detalle los algoritmos utilizados, incluyendo la siguiente información al respecto con relación a los mismos: (i) el tipo de algoritmo que se va a utilizar, así como qué tipo de aprendizaje automático (supervisado, no supervisado, etc.) se utiliza; (ii) su arquitectura, es decir, cómo se estructura el algoritmo (por ejemplo, redes neuronales, árboles de decisión), lo que determina los métodos de entramiento a los que ha sido sometido

el sistema; (iii) los parámetros a los que el mismo está sujeto, concretando que valores se han configurado para los diferentes parámetros del algoritmo; y, finalmente, (iv) la interpretabilidad del mismo, lo que lleva consigo si el contenido del algoritmo es interpretable, la documentación técnica, la cual en tal caso, debe explicar cómo se pueden interpretar sus resultados.

c.5) Las métricas que se van a utilizar a los efectos de medir y cuantificar el rendimiento del sistema.

d) Las pruebas y validación. En el contexto de las pruebas y la validación del sistema, debe tenerse presente que la documentación técnica debe abordar algunos aspectos muy concretos y determinados con relación a la misma, y que hace referencia a los siguientes aspectos:

d.1) En primer término, hay que hacer referencia al llamado «plan de pruebas», el cual debe tener como finalidad la definición clara de los objetivos que se persiguen con las pruebas, *v. gr.*: (i) la verificación de la precisión del sistema; (ii) la robustez en su funcionamiento; y (iii) la seguridad que proporciona y se deriva del uso del mismo. En este orden de cosas, hay que hacer referencia también a la metodología de pruebas, que consiste en una descripción detallada de las técnicas y herramientas que se utilizarán para realizar las pruebas, incluyendo las siguientes: las de carácter unitario, las pruebas de integración, las pruebas de sistema, y, por último, las pruebas de aceptación. También es necesario traer a colación en este momento los criterios de aceptación, los cuales hacen referencia, precisamente, a la especificación de los criterios que se deben cumplir para considerar que las pruebas llevadas a cabo han tenido un resultado exitoso.

d.2) En segundo lugar, hay que hacer expresa mención a la «ejecución de las pruebas». A tales efectos, se exige un registro de pruebas, donde tienen que contemplarse con cierto detalle los casos de prueba llevados a cabo, con especial indicación de: (i) los datos utilizados; (ii) los resultados obtenidos; y (iii) la existencia de cualquier anomalía que haya sido detectada durante la ejecución de tales diligencias de prueba. También debe tenerse en consideración los análisis de los resultados obtenidos como consecuencia de las pruebas llevadas a cabo. Ello exige una interpretación de los resultados de las pruebas con la finalidad de poder identificar adecuadamente los posibles errores, los sesgos o la existencia de vulnerabilidades contenidas en el funcionamiento del sistema.

d.3) Seguidamente es procedente llevar a cabo la validación del sistema. Ello requiere, en primer término, la obtención de la evidencia de la propia validación realizada, que conlleva la exigencia de la recopilación de aquellas pruebas que demuestren que el sistema cumple con los requisitos especificados en la normativa contenida en el anexo IV del RIA, y que al mismo tiempo funciona según las indicaciones proporcionadas para su funcionamiento. Complementariamente a ello, es necesario llevar a cabo un «informe de validación», el cual ha

de contener el resumen de los resultados de las pruebas, así como la validación realizada, incluyéndose en el mismo aquellas conclusiones y recomendaciones que se consideren pertinentes.

En definitiva, con ello se pretende fundamentalmente llevar a cabo el análisis de la robustez y la seguridad en el funcionamiento del sistema de IA.

e) La gestión de la calidad representa un proceso fundamental para garantizar que la información sea precisa, completa, confiable y fácil de entender. En este sentido, la misma exige el cumplimiento de algunos aspectos clave en el contenido de la documentación técnica de los sistemas de IA, que son los que se citan seguidamente:

e.1) Una adecuada planificación del sistema de IA, lo que lleva consigo la necesidad de establecer y definir objetivos en el ámbito de la documentación técnica, que tienen que ser claros, tales como: (i) proporcionar la información necesaria para el desarrollo, la operación, el mantenimiento y la evaluación del sistema de IA; (ii) identificar a las partes interesadas en dicho sistema de IA, lo que representa: (i) la determinación de la persona o personas que necesita/n la documentación técnica, y, de manera concreta, qué información específica necesitan al efecto; y definir el alcance y el contenido de dicha información, en el sentido de especificar qué información se incluirá en la documentación técnica y cómo se organizará la misma.

e.2) Con relación a la creación y formalización de la documentación técnica exigida por el sistema de IA, se exige, en primer término, que la misma sea escrita en un lenguaje claro y conciso. En este orden de cosas, se hace preciso utilizar un lenguaje que sea fácil de entender para las partes interesadas, evitando tecnicismos innecesarios. También debe tenerse presente la necesidad de utilizar un formato consistente, y ello determina la necesidad de aplicar un formato consistente y de carácter uniforme a lo largo de la documentación técnica precisamente con la finalidad de facilitar su consulta. Finalmente, también se hace preciso incluir principalmente aquella información que sea precisa y completa, y que, además, se encuentre conveniente actualizada.

e.3) Con relación a la revisión y al debido control de la documentación técnica, se establece la necesidad de implementar un proceso de revisión adecuado, que tenga como finalidad el hecho de garantizar la calidad de la documentación técnica que se aporte. Del mismo modo, es importante implementar un sistema de control de versiones para documentar los cambios realizados en la documentación técnica; y, finalmente, acompañar a todo ello de un proceso para gestionar los comentarios y las sugerencias de las partes interesadas.

e.4) En lo que atañe a las herramientas y recursos utilizados para la elaboración de la documentación técnica, se hace necesario utilizar herramientas de gestión de contenido con la finalidad de poder facilitar la creación, la revisión y la publicación de la documentación técnica.

Adicionalmente, es recomendable seguir los estándares comúnmente admitidos al efecto, y las mejores prácticas para la documentación técnica de los sistemas de IA; y, en este sentido, cada vez es más necesario poner más énfasis en el hecho de capacitar efectivamente al personal en la creación y la gestión de la documentación técnica de calidad.

En todo caso, y como colofón a lo expuesto con relación a los requisitos exigidos con relación a la documentación técnica que conlleva la gestión de la calidad de la misma, debe tenerse presente que dicha documentación constituye un proceso continuo que requiere un esfuerzo constante.

3. La documentación técnica y las Pymes y Micropymes

El RIA presta una especial atención a la situación tanto de las Pymes, como de las Micropymes. Es cierto que las Pymes pueden verse ciertamente muy comprometidas por las exigencias derivadas de la documentación técnica exigida y vinculada a la puesta en marcha de los sistemas de IA. En este sentido, hay que reconocer que las pequeñas y medianas empresas (Pymes) tienen un papel determinante en la adopción e implementación de sistemas basados en IA. Sin embargo, la complejidad de la documentación técnica para dichos sistemas de IA puede ser un obstáculo para su implementación.

Ello se debe básicamente a la existencia de unos recursos de todo orden que son de manera evidente limitados. Así, no hay que olvidar que las Pymes suelen tener recursos humanos y financieros escasos, lo que dificulta en gran manera la elaboración de dicha documentación técnica, y al mismo tiempo que la misma sea efectivamente completa, y, además, detallada. Complementariamente a ello, no puede pasarse por alto que en el ámbito de las Pymes normalmente se evidencia una considerable falta de experiencia. Por ello, es importante considerar que la complejidad de la IA y sus requisitos técnicos pueden ser un desafío y un obstáculo de difícil superación para las Pymes, que no tienen la suficiente experiencia en este ámbito de actuación. Y, por último, tampoco puede soslayarse el hecho de que por parte de esta tipología de empresas existe un accedo reducido a la información. En este sentido, la información sobre los requisitos legales y las mejores prácticas para la documentación técnica de la IA, en muchas ocasiones, puede ser difícil de encontrar y comprender.

Por todo ello, en el citado apartado 1.º del artículo 11 del RIA se prevé un régimen jurídico adaptado a las necesidades específicas de las Pymes. De este modo, se hace especial alusión al hecho de que las Pymes, entre las que se incluyen normativamente aquellas que sean de nueva creación, pueden facilitar o proporcionar los elementos de la documentación técnica especificados en el anexo IV del RIA, pero en este caso, haciéndolo de manera simplificada, lo que contribuirá tanto al cumplimiento de dichos requisitos técnicos en los términos normativamente exigidos, y al mismo tiempo, la puesta en marcha de dichos sistemas de IA.

Para facilitar dicho proceso, se prevé expresamente en el texto reglamentario de referencia, que sea la Comisión, en el ejercicio de sus facultades y competencias, quien prepare un texto predeterminado, a modo de formulario simplificado de documentación técnica, el cual se ha de encontrar dirigido y orientado a las necesidades concretas y específicas en este ámbito de las pequeñas empresas y microempresas, quienes vendrán obligadas a la utilización de dicho formulario cuando pretendan hacer uso de esta documentación técnica de forma simplificada. En el establecimiento de esta obligación, ello va a ser especialmente útil con relación a las Pymes que lo sean de nueva creación (las *startups*).

También, debe tenerse en cuenta la obligación que se establece de manera general con relación a los organismos, que hayan sido notificados de la existencia de dichos sistemas de IA, a los efectos de que acepten los indicados formularios simplificados con relación a la documentación técnica exigida, a los efectos de la evaluación de la conformidad de la misma de acuerdo con las prescripciones contenidas en el anexo IV del Reglamento.

Consecuentemente con ello, si bien es cierto que existen importantes desafíos para las Pymes, también es verdad que se han puesto soluciones efectivas y eficientes para ayudar a dichas empresas, precisamente para que puedan cumplir con los requisitos legales y mejorar la transparencia y la seguridad de sus sistemas de IA.

4. Otras consideraciones vinculadas a la armonización normativa

Debe indicarse que, en el apartado 2.º de este artículo 11 del RIA, se regulan determinados aspectos de los sistemas de IA, cuando se proceda a la comercialización o se ponga en servicio un sistema de IA de alto riesgo relacionado con un producto al que se apliquen los actos jurídicos enumerados en la parte 1.ª del anexo II del Reglamento, se elaborará a los efectos de coordinar de manera adecuada la exigencia de dicha documentación técnica una única documentación de estas características, que contenga toda la información establecida en el apartado 1.º del citado artículo 11 del Reglamento, al que se ha hecho puntual referencia, así como aquella información exigida en virtud de dichos actos jurídicos contenidos en la indicada sección a del anexo II del mismo.

Ello tiene como finalidad proceder en los diversos supuestos contenidos en las indicadas previsiones normativas, a unificar la información a proporcionar, de tal modo que la misma reúna los requisitos de claridad y que sea completa, pese a la variedad normativa legislativa existente en los diversos supuestos a los que la misma hace referencia.

Finalmente, debe tenerse en cuenta que en el apartado 3.º de dicho artículo 11 del RIA se prevé expresamente la posibilidad de que la Comisión Europea esté expresamente facultada para adoptar actos delegados con arreglo al artículo 73 del Reglamento a fin de modificar el anexo IV relativo a la documentación de carácter técnico que es necesario aportar con la implementación de los sis-

temas de IA en los términos indicados, cuando ello sea necesario para garantizar que, a la luz del progreso técnico que se produzca en el ámbito de dichos sistemas, sea necesario que la indicada documentación técnica se actualice y, al mismo tiempo, proporcione toda la información necesaria para evaluar la conformidad del sistema con los requisitos establecidos en el Reglamento, en los términos a los que se ha hecho referencia en el presente comentario.

Artículo 12. ***Conservación de registros***

1. Los sistemas de IA de alto riesgo permitirán técnicamente el registro automático de acontecimientos (en lo sucesivo, «archivos de registro») a lo largo de todo el ciclo de vida del sistema.

2. Para garantizar un nivel de trazabilidad del funcionamiento del sistema de IA de alto riesgo que resulte adecuado para la finalidad prevista del sistema, las capacidades de registro permitirán que se registren acontecimientos pertinentes para:

a) la detección de situaciones que puedan dar lugar a que el sistema de IA de alto riesgo presente un riesgo en el sentido del artículo 79, apartado 1, o a una modificación sustancial;

b) la facilitación de la vigilancia poscomercialización a que se refiere el artículo 72, y

c) la vigilancia del funcionamiento de los sistemas de IA de alto riesgo a que se refiere el artículo 26, apartado 5.

3. En el caso de los sistemas de IA de alto riesgo mencionados en el anexo III, punto 1, letra a), las capacidades de registro incluirán, como mínimo:

a) un registro del período de cada uso del sistema (la fecha y la hora de inicio y la fecha y la hora de finalización de cada uso);

b) la base de datos de referencia con la que el sistema ha cotejado los datos de entrada;

c) los datos de entrada con los que la búsqueda ha arrojado una correspondencia;

d) la identificación de las personas físicas implicadas en la verificación de los resultados que se mencionan en el artículo 14, apartado 5.

Jimena Campuzano Gómez-Acebo
Jesús Sieira Gil
Registradores de la Propiedad y Mercantiles

I. Consideraciones previas

Los sistemas de inteligencia artificial (en adelante, IA) de alto riesgo están sujetos al cumplimiento de una serie de requisitos y obligaciones para poder ser

comercializado o desplegados en el ámbito de la Unión Europea. Estos requerimientos están enumerados en el capítulo III de la sección 2 del RIA, y deberán ser cumplidos teniendo en cuenta la finalidad prevista para el sistema de IA y el estado actual de la técnica. Entre ellos se encuentra la conservación de registros de actividad de los sistemas de IA de alto riesgo, definida y perfilada esta obligación en este artículo 12 del RIA.

II. Concordancias

Para la aplicación del presente artículo, véase también:

– Cdos. 71, 73 y 91 del RIA.

– Arts. 13, 14, 19, 21, 22, 23, 26, 59, 72 y 79, y anexo III del RIA.

– Art. 11 del 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.

– Art. 9 del RGPD.

III. Comentario

1. Fundamento del requisito de conservar archivos

El objetivo principal del RIA es reducir al máximo los riesgos que pueden presentar, para los usuarios y personas afectadas, los sistemas de IA que se introducen en el mercado o que se ponen en servicio en la Unión, por ello subordina su utilización al cumplimiento de ciertos requisitos obligatorios.

El requisito relativo a la conservación de registros por parte de los sistemas de IA de alto riesgo, a que hace referencia este artículo 12 del RIA, responde al mandato general al que están sujetos los sistemas de IA de transparencia, trazabilidad, y explicabilidad. Resulta esencial disponer de información sobre el modo en que se han desarrollado los sistemas de IA de alto riesgo y sobre su funcionamiento durante todo su ciclo de vida. Además, para establecer las medidas de gestión de riesgos más adecuadas se debe documentar y motivar las elecciones realizadas. A tal fin, es preciso llevar y conservar registros de los eventos acaecidos durante todo su ciclo de vida, para evaluar si el sistema de IA en cuestión cumple los requisitos pertinentes.

La Comisión en el considerando 71 establece que, para (i) permitir la trazabilidad de los sistemas de IA de alto riesgo (ii) verificar si cumplen los requisitos previstos en el Reglamento, así como (iii) vigilar su funcionamiento y llevar a cabo la vigilancia poscomercialización, resulta esencial disponer de información comprensible sobre el modo en que se han desarrollado y sobre su funcionamiento durante toda su vida útil. Los sistemas de IA de alto riesgo deben per-

mitir técnicamente el registro automático de acontecimientos, mediante archivos de registro, durante toda la vida útil del sistema.

El RIA es plenamente coherente con las recomendaciones y principios fijados internacionalmente, lo que permite garantizar la compatibilidad del marco propuesto para la IA con los adoptados por los socios comerciales internacionales de la UE. Uno de los objetivos europeos es que esta normativa pueda servir para, *de facto*, establecer unas prácticas y estándares comunes internacionales, de una forma similar a como ocurrió con la normativa europea sobre protección de datos.

Así, los Principios de la OCDE sobre Inteligencia Artificial, adoptados por la OCDE y los países socios en la reunión del consejo de ministros de la organización de mayo 2019, para velar por el diseño de sistemas de IA robustos, seguros, imparciales y fiables, incluyen como tales los de *«transparency, traceability and explainability»*.

En el RIA se han plasmado los requisitos esenciales fijados por la Comisión Europea en su Comunicación de 2019 al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo, sobre la generación de confianza en la IA centrada en el ser humano. Para lograr una IA fiable, basada en valores europeos, la Comisión respalda las Directrices éticas para una IA fiable establecidas por el Grupo independiente de expertos de alto nivel sobre IA, entre los que se encuentra la necesidad de transparencia.

2. Diseño y desarrollo del sistema de conservación de registros

Consecuentemente, con la finalidad de garantizar la trazabilidad y transparencia, es requisito para la puesta en marcha de un sistema de IA de alto riesgo que esté diseñado técnicamente para permitir el registro automático de acontecimientos, lo que el artículo denomina «capacidades de registro». Esto es, todo sistema de IA de alto riesgo debe contar con la aptitud de constatar y conservar los distintos eventos generados durante su funcionamiento. Es requerimiento esencial que desde su concepción se desarrolle el sistema de IA con «capacidades de registro», *«logging capabilities»*.

Debe estar integrado técnicamente en la propia estructura del sistema de IA. No puede configurarse como un programa aparte, o recurrirse a agentes externos, sino que el propio sistema debe estar proyectado para incluir entre sus funcionalidades un método para el *«logging»,* para almacenar y consignar los diferentes eventos acaecidos.

El diseño exige automatismo, es decir debe ser un registro automático, no debe precisar intervención humana para la constancia de los diferentes acontecimientos producidos durante el funcionamiento del sistema de IA. Además, debe de estar operativo durante todo el ciclo de vida del sistema, durante toda su vida útil. Sin perjuicio de las obligaciones que existen de conservar estos

archivos de registro, puede que incluso más allá del tiempo de uso del sistema de IA.

Un adecuado cumplimiento de las finalidades previstas para este archivo de registros, para que pueda examinar la trazabilidad del sistema de IA de alto riesgo, exige que se diseñe de tal forma que cuente con herramientas de seguimiento y verificación. No es suficiente con hacer una recopilación de eventos, pues el solo archivo de acontecimientos de por sí no es suficiente para cumplir con los objetivos fijados. Deberá contar el sistema con instrumentos para el análisis y monitorización, que alerten de la existencia de riesgos o, en general, de una desviación del cumplimiento de los requisitos exigidos. Para ello deberá configurarse el sistema con tales herramientas técnicas y dotarlas de indicadores o reglas para analizar la información obtenida.

Adicionalmente, dadas las obligaciones que existen de conservar estos archivos de registro, el diseño de los mismos debe prever sistemas de almacenamiento seguros, resistentes y escalables, y que permitan el acceso posterior para, si fuera necesario, realizar análisis o auditorías. Deben establecerse también medidas de protección, para controlar el acceso a esos registros, y evitar manipulaciones.

En el caso de que los ficheros de estos registros de actividad incluyan datos personales, deberá tenerse en cuenta, para su diseño, la normativa sobre protección de datos, y prestar particular atención a la justificación legal existente para el tratamiento de datos personales en estos registros de *log*, al interés legítimo del responsable limitado a los propósitos previstos en la norma (considerando 49 del Reglamento General de Protección de Datos).

Las personas obligadas principalmente a verificar la existencia de estos archivos de registro en los sistemas de IA de alto riesgo son los proveedores. Además de su obligación genérica de velar porque cumplan los requisitos definidos en la sección 2.ª —artículo 16 apartado a) del RIA—, expresamente el artículo 17, en su apartado primero, les impone establecer un sistema de gestión de la calidad que incluya las técnicas, los procedimientos y las actuaciones sistemáticas que se utilizarán en el diseño, control, y desarrollo del sistema, y los sistemas y procedimientos para llevar un registro de toda la documentación e información pertinente —artículo 17 apartado primero k) del RIA—.

En España 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 establece en su artículo 11 que los sistemas de inteligencia artificial deberán técnicamente permitir el registro automático de eventos (logs) a lo largo del ciclo de vida del sistema. Estos registros serán guardados por el participante.

3. Obligaciones de conservación y de dar acceso

Los proveedores de sistemas de IA de alto riesgo también tienen la obligación de conservar estos archivos de registro generados automáticamente por sus sistemas de IA, siempre que estén bajo su control —artículo 16 apartado e) del RIA—, al menos durante seis meses —artículo 19 apartado primero del RIA—. Todo ello 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. El Reglamento (UE) 2016/679 General de Protección de Datos requiere, en particular, garantizar que se limite a un mínimo estricto el plazo de conservación de los datos personales. La normativa sobre prevención del blanqueo de capitales y de la financiación del terrorismo también establece obligaciones sobre la conservación y tratamiento de los registros de actividad.

Los proveedores que sean entidades financieras sujetas a requisitos relativos a su gobernanza, sus sistemas o sus procesos internos, en virtud del Derecho de la Unión en materia de servicios financieros, mantendrán los archivos de registro generados automáticamente por sus sistemas de IA de alto riesgo como parte de la documentación conservada, en virtud del Derecho pertinente en materia de servicios financieros.

Por último, respecto de estos archivos de registro, y siempre que estén bajo su control, también deben los proveedores, si les fuera solicitado, cooperar con las autoridades competentes, permitiéndoles el acceso —artículo 21 del RIA—. En el caso de proveedores de sistemas de IA de alto riesgo que están establecidos en terceros países, la obligación de permitir el acceso se desplaza a los representantes autorizados que hayan nombrado —artículo 22 apartado 3 c) del RIA—. Contempla este precepto que los archivos estén bajo control del proveedor, pero que es el representante nombrado el que debe de dar acceso a las autoridades competentes previa solicitud motivada, de dichos archivos de registro generados automáticamente.

Si el control de los archivos de registro lo tienen los responsables de despliegue, a ellos les corresponderá la obligación de conservación de los mismos, durante un período de tiempo adecuado para la finalidad prevista del sistema de IA de alto riesgo, de al menos seis meses, salvo que se disponga otra cosa en el Derecho de la Unión o nacional aplicable, en particular el Derecho de la Unión en materia de protección de datos personales —artículo 26 apartado sexto del RIA—. En todo caso, aun cuando no tengan el control de estos archivos, debe permitirse a estos responsables de despliegue recabar, almacenar e interpretar correctamente los archivos de registro. Así, el artículo 13 apartado tercero f) del RIA exige que dentro de las instrucciones de uso que se proporcionan a los responsables de despliegue conste una descripción de los mecanismos incluidos en el sistema de IA de alto riesgo que les permita realizar estas acciones.

No establece el RIA obligación específica alguna respecto de la conservación o acceso de estos archivos para los importadores o distribuidores, sin perjuicio de su obligación genérica de cerciorarse que cumplen con los requisitos establecidos en la sección segunda del RIA. Con la excepción de que se trate de un distribuidos o importador, o en general responsable del despliegue o usuario, en el que concurra alguna de las circunstancias previstas en el artículo 25 del RIA —que ponga su nombre o marca o modifique sustancialmente el sistema de IA—, en cuyo caso será considerado proveedor de dicho sistema, y por l o tanto estará sujeta a las obligaciones del artículo 16, y entre ellas la de conservar los archivos de registro.

4. Contenido de los «archivos de registro»

El RIA no desciende a detallar cuales son los acontecimientos concretos que deben recogerse en estos archivos, sólo exige que sean aquellos adecuados para garantizar un nivel de trazabilidad del funcionamiento del sistema de IA de alto riesgo que resulte oportuno para la finalidad prevista. Por lo que deberán consignarse todos los hechos que permitan comprobar que el sistema de IA está operando de una forma acorde con el uso para el que el proveedor lo concibió —artículo 3 apartado 12) del RIA—.

El RIA no define los incidentes específicos que deben registrarse en estos archivos, no identifica hechos concretos, sólo establece este artículo 12, en su apartado segundo, con carácter general para todos los sistemas de IA de alto riesgo, que deben ser aquellos apropiados para cumplir unos objetivos determinados. En su apartado tercero sí identifica, para ciertos tipos de sistemas de IA de alto riesgo, los eventos concretos que deben consignarse como mínimo.

El apartado segundo de este artículo hace referencia a varios tipos de eventos, que podríamos catalogar como eventos de seguridad —aquellos que permiten detectar un riesgo o anomalía— y eventos de control —aquellos que permiten la vigilancia del sistema, que funciona correctamente—.

En el apartado a), este artículo 12 se refiere al registro de los acontecimientos pertinentes para detectar situaciones que puedan dar lugar a que el sistema presente un riesgo en el sentido del artículo 79, apartado primero. El empleo del verbo detectar denota que este sistema de «archivos de registro» no puede ser simplemente un sistema de «archivo y conservación», sino que debe estar diseñado, como decíamos anteriormente, no sólo para constatar un hecho, sino para alertar de la existencia de cualquiera de estos eventos de seguridad.

Por lo tanto, deberán registrarse los acontecimientos pertinentes para detectar situaciones que puedan dar lugar a que el sistema de IA presente un riesgo que afecte a la salud, seguridad o derechos fundamentales de las personas. Es necesario detectar y alertar de esta situaciones, porque la consideración de que un sistema de IA de alto riesgo efectivamente supone un riesgo de estas características va a tener, aparte de otras consecuencias, la de pasar a ser considerado

como «producto que presenta un riesgo» a los efectos 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) 765/2008 y (UE) 305/2011 de producto de riesgo. Un producto de riesgo a los efectos de este Reglamento es, según establece el artículo 3 punto 19 del mismo, aquel producto que pueda 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 considera 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.

También se refiere este artículo 12, en su número segundo apartado a), a eventos relevantes para la detección de situaciones que puedan dar lugar a que el sistema de IA de alto riesgo experimente una modificación sustancial. Debe consignar y reflejar incidentes que puedan suponer, por implicar un cambio en el sistema de IA que no haya sido previsto o una modificación de la finalidad inicial, una modificación sustancial del mismo —según la definición del artículo 3 apartado 23 del RIA—. La razón fundamental es que la modificación sustancial de un sistema de IA de alto riesgo supone la consideración de la creación de un nuevo sistema de IA de alto riesgo, y por lo tanto la generación de la exigencia del cumplimiento de los requisitos y obligaciones previstos en el Reglamento.

Los «archivos de registro» de los sistemas de IA de alto riesgo deben también comprender ciertos eventos de control, por ejemplo, tal como dispone el apartado b) del número segundo de este artículo, los acontecimientos generados que faciliten la vigilancia poscomercialización a que se refiere el artículo 72 del RIA. El cumplimiento de todos los requisitos legales previstos debe producirse, no sólo en el momento de diseño y puesta en funcionamiento del sistema de IA de alto riesgo, sino durante todo el tiempo que este esté en funcionamiento. De ahí que el RIA establece la obligación de los proveedores de realizar una vigilancia poscomercialización, que recoja y examine la experiencia obtenida con el uso y que permita evaluar la necesidad de aplicar medidas correctoras o preventivas —artículo 3 número 25 del RIA—. El artículo 72 del RIA prevé que ese sistema de vigilancia recopile, documente y analice los datos pertinentes que puedan recopilarse sobre el funcionamiento de los sistemas de IA de alto riesgo durante toda su vida útil. Para proporcionar estos datos, y con el fin de valorar el cumplimiento permanente y constante de los requisitos establecidos, resultan esenciales los «archivos de registro», el registro automático de los acontecimientos ocurridos durante todo el ciclo de vida del sistema de IA de alto riesgo.

Por último, también deben conservarse los acontecimientos pertinentes para facilitar el control y seguimiento del sistema de IA de alto riesgo por parte de los responsables del despliegue. Así, conforme al apartado c) del punto segundo de este artículo, también deben conservarse los acontecimientos oportunos para que los responsables del despliegue de sistemas de IA de alto riesgo puedan vigilar el funcionamiento del mismo, tal como les encomienda el artículo 26 del RIA. Fundamentalmente deben asegurarse de que se utilizan los sistemas con arreglo a sus instrucciones de uso y, en particular, y en la medida en que ejerzan control sobre ellos, que los datos de entrada sean pertinentes y suficientemente representativos en vista de la finalidad prevista.

Tal como hemos expuesto, fuera de la cita de estos objetivos que se pretenden cumplir con el archivo de acontecimientos, el RIA no enumera hechos concretos que deban consignarse, pero podemos señalar algunos tipos de eventos que, con carácter general, podrían recopilarse, como por ejemplo, fallos en las comunicaciones, si se trata de sistemas cuyo uso requiere autenticación, la identificación del uso incorrecto de usuarios y contraseñas, accesos fallidos, rendimiento, actividad de los usuarios, etc.

5. Sistemas de IA de alto riesgo de identificación biométrica remota

Para una categoría concreta de sistema de IA de alto riesgo, sí establece el RIA, en el apartado tercero de este artículo, unos incidentes específicos, que operan como mínimos, que deben registrarse en estos archivos. Son los sistemas de IA de alto riesgo mencionados en el anexo III, punto 1, letra a) del RIA, es decir, los sistemas para la identificación biométrica remota. Están excluidos los de verificación biométrica cuya única finalidad sea confirmar que una persona física concreta es la persona que afirma ser.

Ha optado el RIA por permitir excepcionalmente los sistemas de IA de identificación biométrica remota, pero consciente de los riesgos que pueden suponer para los derechos fundamentales de las personas, ha sido especialmente cuidadoso en perfilar los requisitos, limitaciones y obligaciones de dichos sistemas, para que se utilicen de manera responsable y proporcionada y sin vulnerar derechos fundamentales —artículo 5 apartado primero letra h) y apartado segundo del RIA—.

Así, para los sistemas de IA de alto riesgo cuya finalidad sea la identificación biométrica remota, y cuyo uso se ajuste a las previsiones del RIA, el artículo 12 en su apartado tercero sí establece expresamente los acontecimientos que, como mínimo, que deben incluirse en estos «archivos de registros».

En primer lugar, debe recogerse el período de uso del sistema, las fechas y horas de inicio y fin de cada uno de los usos del sistema. El consignar el tiempo de uso es fundamental en los sistemas de identificación biométrica remota, especialmente para valorar la limitación temporal a que están sujetos, y por el distinto tratamiento que tienen según se trate de sistemas de identificación «en

tiempo real» o en diferido. El uso de sistemas de IA para la identificación biométrica remota «en tiempo real» invade de forma especialmente grave los derechos y las libertades de las personas, y la inmediatez de las consecuencias y las escasas oportunidades para realizar comprobaciones en los sistemas que operan «en tiempo real», acrecientan estos riesgos. De ahí que sólo se consideren admisibles, y con la debida autorización, los sistemas de identificación biométrica remota «en tiempo real» en espacios de acceso público con fines de garantía del cumplimiento del Derecho con el objetivo de identificar personas en los casos de víctimas de determinados delitos, de prevención de amenazas reales, presentes o previsibles, como los atentados terroristas, y de búsqueda de personas sospechosas de los delitos más graves —artículo 5.1, párrafo 1, letra h)—.

En los sistemas «en tiempo real» la recogida y comparación se produce de manera instantánea o casi instantánea, casi en directo, mientras que en los sistemas «en diferido» hay una importante demora entre la captación de los datos biométricos y la comparación e identificación. Tiene la precaución el RIA de definir qué se entiende por «tiempo real» y advirtiendo que no se puede eludir las precauciones, generando «demoras mínimas» —artículo 3 apartado 42 del RIA—. Consecuentemente, resulta esencial que se conserven los datos relativos a los períodos de uso de estos sistemas.

En segundo lugar, según el apartado b) del punto tercero, debe consignarse la base de datos de referencia con la que el sistema ha cotejado los datos de entrada. En los sistemas de identificación biométrica se determina la identidad de una persona comparando sus datos biométricos con los almacenados en base de datos de referencia, es decir bases de datos que almacenan datos biométricos. Estas bases de datos de personas de referencia deben ser, según dispone el considerando 34 del RIA, adecuadas para cada supuesto de uso en cada una de las situaciones enumeradas de manera limitativa y definidas con precisión. La existencia de bases de datos biométricos de personas es excepcional, y debe siempre y en todo caso estar amparada en alguna de las excepciones a la prohibición general de tratamiento de datos biométricos que contempla el Reglamento general de protección de datos (artículo 9 del 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). Por todas las precauciones y cautelas que han de observarse, resulta fundamental que se identifiquen las bases de datos utilizadas, y que quede constancia de este hecho en los «archivos de registro» del sistema de IA.

En tercer lugar, los archivos de registro incluirán los datos de entradas con los que la búsqueda ha arrojado una correspondencia. Los datos de entrada, conforme a la definición del artículo 3 apartado 33 del RIA, son aquellos datos proporcionados a un sistema de IA, u obtenidos directamente por él, a partir de los cuales produce un resultado de salida.

La constancia de estos datos de entrada en el archivo de registros en estos sistemas de IA de identificación biométrica probablemente responda al efecto previsto en el RIA de suprimirlos si la autorización para el uso de sistema de IA de identificación biométrica remota es denegada, ya se trate de sistemas «en tiempo real» o en diferido —artículo 5.3 y artículo 26.10 del RIA—. Por lo tanto, el sistema debe identificar estos datos de entrada, para tenerlos localizados en el caso de que deban ser eliminados.

Por último, estos archivos de registros deberán identificar a las personas físicas implicadas en la verificación de resultados. En los sistemas de IA de alto riesgo de identificación biométrica remota, tal como establece el artículo 14 apartado 5 del RIA, el responsable del despliegue no puede emprender ninguna acción basada en la identificación generada por el sistema, sin que antes, al menos dos personas físicas, la verifiquen y confirmen por separado. Si fuera necesario comprobar la existencia de estas personas y valorar su competencia, formación y autoridad necesaria, resulta imprescindible dejar constancia de su identificación en estos «archivos de registro» del sistema de IA de identificación biométrica remota.

IV. Bibliografía

Barrio Andrés, M. (2022). *Manual de Derecho digital*. Editorial Tirant lo Blanch, Valencia, 2.ª edición.

Barrio Andrés, M. (dir.) (2023). *Legal Tech. La transformación digital de la abogacía*. Editorial La Ley, Madrid, 2.ª edición.

Barrio Andrés, M. (dir.) (2024). *El Reglamento Europeo de Inteligencia Artificial*. Editorial Tirant lo Blanch, Valencia.

Artículo 13. ***Transparencia y comunicación de información a los responsables del despliegue***

1. Los 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.

2. 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.

3. Las instrucciones de uso contendrán al menos la siguiente información:

a) la identidad y los datos de contacto del proveedor y, en su caso, de su representante autorizado;

b) las características, capacidades y limitaciones del funcionamiento del sistema de IA de alto riesgo, con inclusión de:

i) su finalidad prevista,

ii) el nivel de precisión (incluidos los parámetros para medirla), 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,

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 sus resultados de salida,

v) cuando proceda, su funcionamiento con respecto a determinadas personas o determinados colectivos de personas 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 de alto riesgo,

vii) en su caso, información que permita a los responsables del despliegue interpretar los resultados 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 supervisión humana a que se hace referencia en el artículo 14, incluidas las medidas técnicas establecidas para facilitar la interpretación de los resultados 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;

f) cuando proceda, una descripción de los mecanismos incluidos en el sistema de IA de alto riesgo que permita a los responsables del despliegue recabar, alma-

cenar e interpretar correctamente los archivos de registro de conformidad con el artículo 12.

María Concepción Campos Acuña
Doctora en Derecho. Codirectora de la Cátedra de Buen Gobierno Local (Universidad de Vigo). Experta en Gestión Pública. Funcionaria de Habilitación Nacional (en excedencia). Profesora Asociada de Derecho Administrativo de la Universidad Rovira i Virgili

I. Consideraciones previas

El artículo 13 del RIA establece el criterio básico de transparencia algorítmica en el uso de los sistemas de IA de alto riesgo, es decir aquéllos calificados como tales, de conformidad con lo establecido en el anexo III del RIA. Pero a diferencia de su uso común y generalizado, pensando en los destinatarios finales, y en evitar las conocidas *black-box,* en este caso dicha transparencia tiene como destinatarios a los responsables del despliegue, con la finalidad de que éstos puedan interpretar y usar correctamente los resultados de salida.

Debemos recordar que con el enfoque basado en la gestión de riesgos que ha adoptado el RIA, siguiendo las técnicas de *compliance*, adquiere una importancia clave el uso ético y responsable de la IA, conectado con las Directrices éticas para una IA fiable (2019) y, por tanto, de los siete principios éticos no vinculantes para la IA que tienen por objeto contribuir a garantizar la fiabilidad y el fundamento ético de la IA. Los siete principios 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 ambiental; y rendición de cuentas.

Principio de transparencia que se concreta en la obligación de que los sistemas de IA se desarrollen y utilicen 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 informen 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 este punto, simplemente hay que matizar que el RIA no se aplicará a las obligaciones de los responsables del despliegue que sean personas físicas que utilicen sistemas de IA en el ejercicio de una actividad puramente personal de carácter no profesional.

II. Concordancias

Para la aplicación del presente artículo, véase también:

– Art. 5.1.a) y arts. 12 y siguientes del RGPD.

– Art. 11 de la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.

– Art. 41 de la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público.

– Art. 23 de la Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación.

– Arts. 8 y 11 del 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.

III. Comentario

El precepto objeto de análisis se estructura en tres apartados en los que se va desgranando esa obligación de transparencia y comunicación de información a los responsables del despliegue, responsables que podrán ser tanto del sector público, como privado, y en el primer caso, en los diferentes niveles de administración territorial, con independencia de su tamaño. Lo determinante será el uso de sistemas de IA calificados como de alto riesgo.

1. Obligación general de un nivel de transparencia suficiente

En relación con la obligación general enunciada en el apartado 1, debe ponerse de relieve la utilización de dos calificativos: suficiente y adecuados, con relación al grado de transparencia que los responsables del diseño y desarrollo de los sistemas de IA deben ofrecer a los responsables del sistema. A tales efectos, ofrece un elemento de referencia como es que se cumplan las obligaciones pertinentes previstas en la sección 3, relativa a las obligaciones de proveedores y responsables del despliegue de sistemas de IA de alto riesgo y de otras partes.

Se contemplan específicamente las responsabilidades a lo largo de toda la cadena de valor de la IA (art. 25), que debe completarse con las obligaciones específicas para los responsables del despliegue (art. 26), que deberán adoptar las medidas técnicas y organizativas adecuadas para garantizar su uso conforme a las instrucciones de uso. Para ello, deberán encomendar la supervisión humana a personas físicas que tengan la competencia, la formación y la autoridad necesarias, y garantizando, cuando así proceda, que los datos de entrada sean pertinentes y suficientemente representativos. En idéntica línea, el art. 11 del 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, cuyo apartado 1.g) recoge la necesidad de que los sistemas de IA hayan sido diseñados y desarrollados de tal modo que puedan ser supervisados por personas físicas.

2. Instrucciones de uso

Para poder asegurar la conexión transparente entre el diseño y el desarrollo del sistema de IA y su despliegue a través de los resultados de salida, el apartado 2 del art. 13 establece un mecanismo concreto a tal efecto: las instrucciones de uso.

Estas instrucciones generan una obligación de doble impacto. Por una parte, para responsables del diseño y desarrollo que deben proporcionar información concisa, completa, correcta y clara que sea pertinente, accesible y comprensible para los responsables del despliegue, que se concreta con un estándar de mínimos en el apartado 3 del precepto. Por otra, para los responsables del despliegue que, tal y como hemos visto, deberán adoptar las medidas adecuadas, tanto técnicas como organizativas, para que el despliegue se produzca conforme a las instrucciones de uso.

3. Contenido mínimo de las instrucciones de uso

La información concisa, completa, correcta y clara sobre el sistema se concreta en el apartado 3 con una enumeración descriptiva del contenido de las instrucciones de uso, engarzada, por remisión, en otros preceptos del RIA, asegurando así un control de seguridad en el uso de la IA.

Este escenario nos sitúa ante el reto y desafío para las entidades, tanto del sector público como del sector privado, de contar con el capital humano con la cualificación profesional suficiente para poder comprobar adecuadamente dicha información. Sólo de este modo se podrá verificar en todo momento la protección de derechos y libertades que el RIA fija como objetivo ante el uso de sistemas de IA de alto riesgo y que adquieren especial relevancia en el caso del sector público, que deberá enlazarse a través del marco de contratación pública, delimitando adecuadamente estas relaciones.

A estos efectos, cabe recordar que el art. 23 de la Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación, contempla ya la obligación (para las administraciones públicas), de favorecer la puesta en marcha de mecanismos para que los algoritmos involucrados en la toma de decisiones que se utilicen en las administraciones públicas tengan en cuenta criterios de minimización de sesgos, transparencia y rendición de cuentas, siempre que sea factible técnicamente. Mecanismos que deberán incluir en su diseño y datos de entrenamiento, y abordar el potencial impacto discriminatorio, a través de, entre otros medios, la realización de evaluaciones de impacto que determinen el posible sesgo discriminatorio.

La transparencia se presenta así como un principio fundamental para permitir el conocimiento y comprobación del buen funcionamiento del sistema de IA por los sujetos de la cadena de valor, en todas su diferentes posiciones, así como todos aquellos que tienen que verificar, comprobar el mismo, en el caso que nos

ocupa, responsables del despliegue. Así se recoge en la Carta de Derechos Digitales de 2021, en el numeral XVI, relativo a los derechos ante la inteligencia artificial, al imponer el establecimiento de «condiciones de transparencia, auditabilidad, explicabilidad, trazabilidad, supervisión humana y gobernanza».

Para finalizar, hay que recordar que, tal y como señala Fry (2019), esta tecnología, los algoritmos no son ni buenos ni malos, sino que será su uso, y, por tanto, los efectos que produzcan, los que determinarán los beneficios o perjuicios que puedan causar. Estas observaciones adquieren especial relevancia en el caso de uso de la IA por los poderes públicos, que nos sitúa ante los principios de buena administración y de transparencia.

IV. Recursos complementarios

AEPD. *Inteligencia Artificial: Transparencia*, 2023.

Digital Future Society. *Hacia un uso responsable de los algoritmos: métodos y herramientas para su auditoría y evaluación*, 2024.

Gobierno de España. *Carta de Derechos Digitales*, 14 de julio de 2021.

OCDE. *Recommendation of the Council on Artificial Intelligence*, 2019, OECD/LEGAL/0449.

UNESCO. *Recomendación Ética en la Inteligencia Artificial*.

V. Bibliografía

Araya Paz, C. F. (2021). Transparencia algorítmica ¿un problema normativo o tecnológico? *CUHSO*, vol. 31, núm. 2.

Barrio Andrés, M. (2022). *Manual de Derecho digital*. Editorial Tirant lo Blanch, Valencia, 2.ª edición.

Barrio Andrés, M. (2023). El cumplimiento basado en el riesgo o *risk-based compliance*, pieza cardinal del nuevo Derecho digital europeo. *Análisis del Real Instituto Elcano*, núm. 34.

Barrio Andrés, M. (dir.) (2023). *Legal Tech. La transformación digital de la abogacía*. Editorial La Ley, Madrid, 2.ª edición.

Campos Acuña, M.ª C. (2021). El derecho a una buena administración digital en la Carta de derechos digitales. *Revista LA LEY de Derecho digital e innovación*, núm. 6.

Cotino Hueso, L. (2022). «Nuevo paradigma en la garantía 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*. Editorial Thomson Reuters, Aranzadi, FIADI, Cizur.

Cotino Hueso, L. (2023). Qué concreta transparencia e información de algoritmos e inteligencia artificial es la debida. *Revista española de la transparencia,* núm. 16.

Fry, H. (2019). *Hola mundo. Cómo seguir siendo humanos en la era de los algoritmos.* Editorial Blackiebooks, Barcelona.

Gallego Córcoles, I. (2023). «La contratación de soluciones de inteligencia artificial», en Gamero Casado, E. (dir.). *Inteligencia artificial y sector público. Retos, límites y medios*. Editorial Tirant lo Blanch, Valencia.

Gamero Casado, E. (dir.) (2023). *Inteligencia artificial y sector público. Retos, límites y medios*. Editorial Tirant lo Blanch, Valencia.

Garriga Domínguez, A. (2023). Las exigencias de transparencia para los sistemas algorítmicos de recomendación, selección de contenidos y publicidad en línea en el nuevo Reglamento Europeo de Servicios Digitales. *Revista española de la transparenci*a, núm. 17.

Jiménez-Castellanos, I. (2023). Decisiones automatizadas y transparencia administrativa. Nuevos retos para los derechos fundamentales. *Revista Española de la Transparencia*, núm. 16.

Martín Delgado, I. (2023). «La aplicación del principio de transparencia a la actividad administrativa automatizada», en Gamero Casado, E. (dir.). *Inteligencia artificial y sector público. Retos, límites y medios*. Editorial Tirant lo Blanch, Valencia.

Martínez Gutiérrez, R. (2023). «Responsabilidad administrativa por el uso de la Inteligencia Artificial», en Gamero Casado, E. (dir.). *Inteligencia artificial y sector público. Retos, límites y medios*. Editorial Tirant lo Blanch, Valencia.

Pasquale, F. (2015). *The Black Box Society. The Secret Algorithms That Control Money and Information*. Editorial Harvard University Press, Cambridge.

Rodríguez Martínez, G. (2020). La transparencia algorítmica en el tratamiento de los datos personales. *Revista de privacidad y derecho digital*, vol. 5, núm. 15.

Sangüesa, R. (2018). Inteligencia artificial y transparencia algorítmica: «It's complicated». *BID: textos universitarios de biblioteconomía y documentación*, núm. 41.

Vestri, G. (2021). La inteligencia artificial ante el desafío de la transparencia algorítmica. Una aproximación desde la perspectiva jurídico-administrativa. *Revista Aragonesa de Administración Pública*, núm. 56.

Artículo 14. ***Supervisión humana***

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 supervisión 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.

3. Las medidas de supervisión serán proporcionales a los riesgos, al nivel de autonomía y al contexto de uso del sistema de IA de alto riesgo, y se garantizarán bien mediante uno de los siguientes tipos de medidas, bien mediante ambos:

a) las medidas que el proveedor defina y que integre, cuando sea técnicamente viable, en el sistema de IA de alto riesgo antes de su introducción en el mercado o su puesta en servicio;

b) las medidas que el proveedor defina antes de la introducción del sistema de IA de alto riesgo en el mercado o de su puesta en servicio y que sean adecuadas para que las ponga en práctica el responsable del despliegue.

4. A efectos de la puesta en práctica de lo dispuesto en los apartados 1, 2 y 3, el sistema de IA de alto riesgo se ofrecerá al responsable del despliegue de tal modo que las personas físicas a quienes se encomiende la supervisión humana puedan, según proceda y de manera proporcionada a:

a) entender adecuadamente las capacidades y limitaciones pertinentes del sistema de IA de alto riesgo y poder vigilar debidamente su funcionamiento, por ejemplo, con vistas a detectar y resolver anomalías, problemas de funcionamiento y comportamientos inesperados;

b) ser conscientes de la posible tendencia a confiar automáticamente o en exceso en los resultados de salida generados por un sistema de IA de alto riesgo («sesgo de automatización»), en particular con aquellos sistemas que se utilizan para aportar información o recomendaciones con el fin de que personas físicas adopten una decisión;

c) 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;

d) decidir, en cualquier situación concreta, no utilizar el sistema de IA de alto riesgo o descartar, invalidar o revertir los resultados de salida que este genere;

e) intervenir en el funcionamiento del sistema de IA de alto riesgo o interrumpir el sistema pulsando un botón de parada o mediante un procedimiento similar que permita que el sistema se detenga de forma segura.

5. En el caso de los sistemas de IA de alto riesgo mencionados en el anexo III, punto 1, letra a), las medidas a que se refiere el apartado 3 del presente artículo garantizarán, además, que el responsable del despliegue no actúe ni tome ninguna decisión basándose en la identificación generada por el sistema, salvo si al menos dos personas físicas con la competencia, formación y autoridad necesarias han verificado y confirmado por separado dicha identificación.

El requisito de la verificación por parte de al menos dos personas físicas por separado no se aplicará a los sistemas de IA de alto riesgo utilizados con fines de garantía del cumplimiento del Derecho, de migración, de control fronterizo o de asilo cuando el Derecho nacional o de la Unión considere que la aplicación de este requisito es desproporcionada.

Fernando H. Llano Alonso
Catedrático de Filosofía del Derecho y Decano de la Facultad de Derecho de la Universidad de Sevilla

I. Consideraciones previas

El artículo 14 del RIA es un precepto clave en la regulación de los sistemas calificados de alto riesgo. A este respecto, resulta muy significativa su ubicación destacada entre los requisitos de los sistemas de IA de alto riesgo que se recogen en la sección 2 del capítulo III, dedicado por completo a los sistemas de IA de alto riesgo, es decir, aquellos que comprometen la salud, la seguridad o los derechos fundamentales (art. 6.3 RIA). Esta situación de algo riesgo puede surgir cuando un sistema de IA de estas características se utiliza conforme a su finalidad prevista o cuando se le da un uso indebido razonablemente previsible.

Consciente de la casuística existente por los perjuicios y problemas ocasionados por el uso de algoritmos de alto riesgo en las tres grandes áreas a las que se refiere el art. 6.2 RIA: salud, seguridad y derechos fundamentales, el legislador europeo introduce en el art. 14 RIA la vigilancia humana para mantener el control humano de la interacción entre las personas físicas y las máquinas inteligentes. A este modelo de interacción hombre-máquina se le denomina, por parte de la doctrina, *Man in the Loop*: el humano es asistido por una máquina; es decir, el hombre toma las decisiones y la máquina sólo apoya o automatiza parcialmente algunas de ellas.

A través de la vigilancia humana del uso de los sistemas de IA de alto riesgo se somete a los mismos a una supervisión efectiva realizada por personas físicas durante su fase de diseño, desarrollo y uso posterior. Por otra parte, la vigilancia humana requiere la dotación de una herramienta interfaz humano-máquina adecuada, de forma proporcionada, a los riesgos asociados a dichos sistemas (art. 14.1 RIA).

El objetivo de la vigilancia humana no es otro que el de prevenir o reducir al mínimo los potenciales riesgos contra la salud, la seguridad, los derechos fundamentales y el *medio ambiente*. Se trata de riesgos derivados del uso inde-

bido o razonablemente previsible de los sistemas de IA de alto riesgo, y cuando las decisiones basadas exclusivamente en el procesamiento automatizado producen efectos jurídicos o de otro tipo a las personas con los que se utiliza el sistema.

A fin de procurar una vigilancia humana eficiente y rigurosa, a los proveedores y a los responsables del despliegue de sistemas de IA se les pide que garanticen «un nivel suficiente de *alfabetización en materia de IA*», de conformidad con los establecido en el art. 4 RIA. A propósito de la alfabetización, el Reglamento no solo encomienda a la Unión Europea y a los Estados miembros la promoción de medidas para el desarrollo de un nivel *suficiente* de alfabetización en materia de IA en todos los sectores, sino que también exige a los proveedores y responsables del despliegue de estos sistemas inteligentes que adopten medidas para garantizar un nivel suficiente de alfabetización en materia de IA entre su personal y otras personas que se ocupen en su nombre de la operación y el uso de los sistemas de IA.

II. Concordancias

Para la aplicación del presente artículo, véase también:

– Art. 22 del RGPD.

– Arts. 11.2, 18, 20.1.c) y 22.4 de la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.

– Art. 9.2 del 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.

– Art. 23 de la Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación.

III. Comentario

En los últimos años ha habido casos paradigmáticos y problemáticos derivados del uso de algoritmos de alto riesgo que han afectado precisamente a las tres grandes áreas a las que se refiere el mencionado precepto de la propuesta de Reglamento UE de IA: seguridad, educación y sanidad. Veamos un ejemplo representativo por cada una de estas áreas.

En relación con la seguridad: el miércoles 19 de diciembre de 2018, a las 21:03 horas, el aeropuerto de Gatwick se vio obligado a cerrar sus vuelos de entrada y salida durante 32 horas debido al avistamiento cerca del aeródromo de varios drones. Más de 120.000 personas se vieron afectadas por este cierre en el segundo aeropuerto londinense. Tan solo unas semanas más tarde, el 8 de enero de 2019, el aeropuerto de Heathrow, el primero en volumen de tráfico aéreo del Reino Unido, también suspendió durante una hora sus vuelos tras

detectar la torre de control, a las 17.12 horas, un dron sobrevolando su espacio aéreo. Desde un punto de vista legal y judicial estos incidentes generaron cambios (el 31 de diciembre de 2020 entró en vigor una nueva normativa para propietarios de drones, clasificando tipos de drones en diversas categorías, introduciendo requisitos de registro y seguro para cada clase). Por otra parte, se suscitó la cuestión a propósito de las reclamaciones de indemnización a las compañías de seguros por parte de los viajeros, y determinar sobre quién recaería la responsabilidad civil en ambos incidentes, se presentaría también la ocasión de plantearse la oportunidad de reconocer la personalidad electrónica de los robots o máquinas inteligentes entrenadas con IA para responder por los perjuicios causados a terceros por su mal funcionamiento o por sus erróneas tomas de decisión automatizadas.

Respecto a la vigilancia humana, la reserva de humanidad y el control de las decisiones administrativas basadas en algoritmos predictivos cuando limitan el disfrute de los derechos fundamentales, en el ámbito de la educación (art. 27 CE) resulta ilustrativo el caso de los exámenes de acceso a las universidades británicas en el verano de 2020, el primero de la pandemia del Covid-19, cuyos resultados no se obtuvieron a partir de una prueba presencial, sino excepcionalmente mediante un algoritmo que combinó la calificación obtenida por los alumnos en sus colegios o institutos en los exámenes realizados en los tres últimos cursos y una calificación estimada de cada alumno en comparación con sus compañeros de su centro.

El resultado de las predicciones del algoritmo utilizado para determinar las notas GCSE y A-Levels de los alumnos de Inglaterra, Gales y Escocia fue, anunciado el 13 de agosto, fue decepcionante: las calificaciones asignadas por el algoritmo fueron, en general, inferiores a las previsiones realizadas por los profesores de los colegios e institutos; por otra parte, se apreció también en la decisión automatizada un sesgo discriminatorio hacia los alumnos de las escuelas públicas situadas en zonas humildes y que, por el contrario, favorecía a los estudiantes de colegios privados o centros públicos de alto rendimiento. A raíz de la polémica causada por la prueba de acceso al sistema universitario británico mediante el empleo de un algoritmo predictivo, el gobierno anunció el 17 de agosto que los resultados del nivel A se modificarían para introducir en ellos como criterio corrector las estimaciones originalmente realizadas por los profesores de los centros de procedencia de cada alumno.

Por último, respecto a la problemática que se origina a partir de la aplicación de los sistemas expertos de IA y los algoritmos en el ámbito de la salud, debe conjugarse su empleo con el marco legal que protege los derechos de los pacientes (consentimiento informado, privacidad y confidencialidad de los datos médicos, acceso al historial clínico, respeto de la voluntad anticipada por el paciente respecto a los cuidados y tratamientos de salud que desea recibir, derecho a presentar reclamaciones y sugerencias).

V. Bibliografía

Arellano Toledo, W. (dir.) (2023). *Derecho, Ética e Inteligencia Artificial*. Editorial Tirant lo Blanch, Valencia.

Barrio Andrés, M. (2022). *Manual de Derecho digital*. Editorial Tirant lo Blanch, Valencia, 2.ª edición.

Barrio Andrés, M. (dir.) (2024). *El Reglamento Europeo de Inteligencia Artificial*. Editorial Tirant lo Blanch, Valencia.

Gamero Casado, E. (dir.) (2023). *Inteligencia artificial y sector público. Retos, límites y medios*. Editorial Tirant lo Blanch, Valencia.

García Mexía, P. (dir.) (2022). *Claves de Inteligencia Artificial y Derecho*. Editorial La Ley, Madrid.

López Calvo, J. (dir.) (2019). *La adaptación al nuevo marco de protección de datos tras el RGPD y la LOPDGDD*. Editorial Bosch-Wolters Kluwer, Madrid.

Artículo 78. *Confidencialidad*

1. La Comisión, 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 o 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, incluido el código fuente, salvo en los casos mencionados en el artículo 5 de la Directiva (UE) 2016/943 del Parlamento Europeo y del Consejo (57);

b) la aplicación eficaz del presente Reglamento, en particular a efectos de investigaciones, inspecciones o auditorías;

c) los intereses de seguridad pública y nacional;

d) el desarrollo de las causas penales o los procedimientos administrativos;

e) la información clasificada con arreglo al Derecho de la Unión o nacional.

2. Las autoridades involucradas en la aplicación del presente Reglamento de conformidad con el apartado 1 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 de conformidad con el presente Reglamento y con el Reglamento (UE) 2019/1020. Establecerán medidas adecuadas y eficaces en materia de ciberseguridad a fin de proteger la seguridad y la confidencialidad de la información y los datos obtenidos, y suprimirán los datos recopilados tan pronto como dejen de ser necesarios para los fines para los que se obtuvieron, de conformidad con el Derecho de la Unión y nacional aplicable.

3. Sin perjuicio de lo dispuesto en los apartados 1 y 2, la información intercambiada de forma confidencial entre las autoridades nacionales competentes o entre estas y la Comisión no se revelará sin consultar previamente a la autoridad nacional competente de origen y al responsable del despliegue cuando las autoridades garantes del cumplimiento del Derecho, del control de fronteras, de la inmigración o del asilo utilicen los sistemas de IA de alto riesgo a que se refiere el anexo III, puntos 1, 6 o 7, y dicha divulgación comprometería los intereses de seguridad pública y nacional. Este intercambio de información no comprenderá los datos operativos sensibles relativos a las actividades de las autoridades garantes del cumplimiento del Derecho, del control de fronteras, de la inmigración o del asilo.

Cuando las autoridades garantes del cumplimiento del Derecho, de la inmigración o del asilo sean proveedores de sistemas de IA de alto riesgo a que se refiere el anexo III, puntos 1, 6 o 7, la documentación técnica mencionada en el anexo IV permanecerá dentro de las instalaciones de dichas autoridades. Dichas autoridades velarán por que las autoridades de vigilancia del mercado a que se refiere el artículo 74, apartados 8 y 9, según proceda, puedan, previa solicitud, acceder inmediatamente a la documentación u obtener una copia de esta. Tan solo se permitirá acceder a dicha documentación o a cualquier copia de esta al personal de la autoridad de vigilancia del mercado que disponga de una habilitación de seguridad del nivel adecuado.

4. Los apartados 1, 2 y 3 no afectarán a los derechos u obligaciones de la Comisión, los Estados miembros y sus autoridades pertinentes, ni a los derechos u obligaciones de los organismos notificados en lo que se refiere al intercambio de información y la difusión de advertencias, también en el contexto de la cooperación transfronteriza, ni a las obligaciones de facilitar información en virtud del Derecho penal de los Estados miembros que incumban a las partes interesadas.

5. Cuando sea necesario y con arreglo a las disposiciones pertinentes de los acuerdos internacionales y comerciales, la Comisión y los Estados miembros podrán intercambiar información confidencial con autoridades reguladoras de terceros países con las que hayan celebrado acuerdos de confidencialidad bilaterales o multilaterales que garanticen un nivel de confidencialidad adecuado.

María Jesús Blanco Sánchez
Profesora de Derecho Mercantil de la Universidad Pablo de Olavide de Sevilla

I. Consideraciones previas

El art. 78 del RIA se encuadra en el capítulo IX, dedicado a la *vigilancia poscomercialización, intercambio de información, vigilancia del mercado*, y particularmente en su sección tercera atinente a *ejecución*. En él se establecen, tras la relación normativa que rige en materia de regulación relativa a vigilancia de mercado y control de los sistemas de IA en el mercado de la Unión, asistencia mutua, vigilancia del mercado y control de sistemas de IA de uso general, super-

visión de las pruebas en condiciones reales por las autoridades de vigilancia del mercado y poderes de las autoridades encargadas de proteger los derechos fundamentales (arts. 74 a 77 RIA, respectivamente), las garantías de confidencialidad de la información y los datos obtenidos por la Comisión, autoridades de vigilancia del mercado, organismos notificados y otras personas físicas o jurídicas que participen en la aplicación del RIA.

El artículo ha de interpretarse en conjunción con los principios inspiradores expuestos en los considerandos del texto. Así, se establece en sentido amplio que los miembros de las autoridades competentes deben abstenerse de todo acto incompatible con el carácter de sus funciones y estarán sujetos a las normas de confidencialidad establecidas en el RIA (considerando 154), en conjunción con artículos que expresamente remiten al que ahora nos ocupa.

Traemos a colación, en este sentido: la obligación de tratar toda información obtenida por una autoridad nacional competente de conformidad con las obligaciones de confidencialidad (art. 21 *Cooperación con las autoridades competentes*), la obligación de las autoridades notificantes de preservar la confidencialidad de la información obtenida (art. 28 *Autoridades notificantes*), la obligación de los organismos notificados de contar con procedimientos documentados que garanticen que su personal, sus comités, sus filiales, sus subcontratistas y todos sus organismos asociados o personal de organismos externos mantengan la confidencialidad de la información (art. 31 *Requisitos relativos a los organismos notificados*), la obligación de los proveedores de modelos de IA de uso general de tratar toda información o documentación obtenida, incluidos los secretos comerciales, se tratarán de conformidad con las obligaciones de confidencialidad (art. 53 *Obligaciones de los proveedores de modelos de IA de uso general*), idéntica a la establecida en art. 55 (*Obligaciones de los proveedores de modelos de IA de uso general con riesgo sistémico*).

En cuanto a espacios controlados de pruebas para la IA (art. 57) el grupo de expertos científicos independientes creado por la Comisión desempeñará sus funciones con imparcialidad y objetividad y garantizarán la confidencialidad de la información y los datos obtenidos en el ejercicio de sus funciones y actividades.

En la misma sección tercera en la que se encuadra el artículo que nos ocupa el art. 74, relativo a *Vigilancia del mercado y control de los sistemas de IA en el mercado de la Unión*, hace expresa referencia al art. 78, estableciendo que cualesquiera información o documentación obtenidas por las autoridades de vigilancia del mercado se tratarán de conformidad con las obligaciones de confidencialidad. En el mismo sentido se establece, en el art. 75 para supuestos de *Asistencia mutua, vigilancia del mercado y control de sistemas de IA de uso general* y en el art. 77 en cuanto a *Poderes de las autoridades encargadas de proteger los derechos fundamentales*.

IV. Bibliografía

Barrio Andrés, M. (dir.) (2024). *El Reglamento Europeo de Inteligencia Artificial*. Editorial Tirant lo Blanch, Valencia.

Hernández Peña, J. C. (2022). *El marco jurídico de la inteligencia artificial. Principios, procedimientos y estructuras de gobernanza*. Editorial Thomson Reuters Aranzadi, Cizur Menor.

Artículo 93. ***Poderes para solicitar la adopción de medidas***

1. Cuando resulte necesario y conveniente, la Comisión podrá solicitar a los proveedores que:

a) adopten las medidas oportunas para cumplir las obligaciones establecidas en los artículos 53 y 54;

b) apliquen medidas de reducción de riesgos cuando la evaluación realizada de conformidad con el artículo 92 apunte a que existen motivos serios y fundados de preocupación por la existencia de un riesgo sistémico a escala de la Unión;

c) restrinjan la comercialización del modelo, lo retiren o lo recuperen.

2. Antes de solicitar que se adopten medidas, la Oficina de IA podrá entablar un diálogo estructurado con el proveedor del modelo de IA de uso general.

3. Si, durante el diálogo estructurado a que se refiere el apartado 2, el proveedor del modelo de IA de uso general con riesgo sistémico se compromete a adoptar medidas de reducción para hacer frente a un riesgo sistémico a escala de la Unión, la Comisión podrá, mediante una decisión, hacer dichos compromisos vinculantes y declarar que no hay ya motivos para actuar.

Margarita Castilla Barea
Catedrática de Derecho Civil en la Universidad de Cádiz

I. Consideraciones previas

Como consecuencia de las situaciones y actuaciones que contemplan los artículos 89 a 92 precedentes, la Comisión puede llegar a la conclusión de que no es preciso adoptar ninguna medida específica en relación con los modelos de IA o que, por el contrario, es necesario que los proveedores emprendan distintas acciones, ya sea para adecuarlos a las exigencias que les imponen los arts. 53 a 55 RIA, ya sea para mitigar los riesgos que se ha constatado que tales modelos comportan y que incluso pueden requerir que se restrinja su comercialización, o se los retire o recupere del mercado. Por consiguiente, el art. 93 RIA no contempla acciones que directamente pone en marcha la Comisión, sino la panoplia de las que puede exigir que implementen los proveedores de modelos de IA para adaptarlos a la legalidad y al nivel de riesgos que, a tenor del propio Reglamento, resultan asumibles para este tipo de inteligencias artificiales.

II. Concordancias

– Art. 14 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. Poderes de las autoridades de vigilancia del mercado.

III. Comentario

Como resultado de cualquiera de las actuaciones posibles que contemplan los artículos precedentes, la Comisión puede concluir la necesidad de que los proveedores de los modelos de IA interesados adopten determinadas medidas y el art. 93 RIA la faculta para conminarles a ello. En particular, según el art. 93.1 RIA la Comisión puede requerirles para:

a) Adoptar las medidas oportunas para cumplir las obligaciones establecidas en los arts. 53 y 54. En nuestra opinión, la remisión correcta debe entenderse hecha a los arts. 53 y 55, ya que es este último el que se refiere a los requisitos exigibles a los modelos de IA de uso general con riesgo sistémico. También nos parecería factible una remisión a las obligaciones establecidas en los arts. 53 a 55 RIA, que incluiría las cargas que en el art. 54 RIA se imponen a los representantes autorizados que los proveedores de terceros países deben designar antes de introducir en el mercado de la Unión sus modelos de IA de uso general.

b) Aplicar medidas de reducción de riesgos cuando, tras una evaluación efectuada de acuerdo con lo previsto en el art. 92 RIA anterior, se haya apreciado que existen motivos serios y fundados de preocupación por la existencia de un riesgo sistémico a escala de la Unión. Sobre la noción de estos últimos, parece oportuno remitirse a lo ya expuesto en el comentario del art. 90 RIA, sobre los distintos supuestos de riesgos sistémicos acerca de los cuáles pueden versar las alertas cualificadas del grupo de expertos científicos independientes.

Finalmente, también puede la Comisión requerir a los proveedores de modelos de IA de uso general que adopten medidas para:

c) Restringir la comercialización del modelo, retirarlo del mercado o recuperarlo. Respecto a estos términos, conviene tener en cuenta las nociones contenidas en el art. 3 RIA. La restricción de la comercialización (art. 3.10) supone controlar y limitar el suministro del modelo de IA lo que, lógicamente, supone admitir la viabilidad de que se mantenga en el mercado, aunque sea con cautelas y exigencias adicionales respecto, por ejemplo, a las cualidades que deben reunir los responsables del despliegue a quienes se pueda suministrar el modelo en cuestión. Por eso restricción de la comercialización y retirada del mercado son medidas antitéticas, puesto que esta última (art. 3.17) consiste, justamente, en impedir que se comercialice un sistema de IA —un modelo de IA de uso general en este caso— que se encuentra ya en la cadena de suministro. Por su parte, la

Este libro analiza, artículo por artículo, todos los temas y aspectos críticos suscitados por el nuevo Reglamento Europeo de Inteligencia Artificial, RIA o AI Act.

En cada comentario se presenta **el sentido, la finalidad y la función del precepto, para permitir a los operadores jurídicos y técnicos interpretar adecuadamente su contenido**, a través de un estudio integral que ofrece tanto una visión de conjunto, como sus concretas interpretaciones específicas, en un singular aporte de valor añadido fundamentalmente práctico.

Los autores son destacados expertos en el Derecho digital, reconocidos por su capacidad de análisis estructurado, exhaustivo y pragmático de las cuestiones relacionadas con la inteligencia artificial.

Con todo ello, se pretende ofrecer a todos los operadores jurídicos concernidos con el desarrollo y despliegue de sistemas de IA **un primer conjunto de soluciones y argumentos sólidos para la aplicación de esta nueva norma.**

ISBN: 978-84-18662-88-1

LA LEY