

La Ley de IA: entre la salud y el Derecho en una nueva era. Un acompasamiento necesario*

The Law of IA: between health and law in a new era. A necessary reconciliation**

Cristina Gil Membrado

Catedrática de Derecho Civil
Universitat de les Illes Balears

<https://doi.org/10.36151/RJIB.2024.26.01>

SUMARIO: I. EVOLUCIÓN Y CONCEPTO DE LA INTELIGENCIA ARTIFICIAL (IA). II. IA Y SALUD. III. LA LEY DE IA. 1. Antecedentes. 2. El Reglamento (UE) 2024/1689 por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de IA). IV. TRANSPARENCIA Y EXPLICABILIDAD EN IA PARA LA SALUD. 1. El algoritmo y sus debilidades. A) Cómo funciona el algoritmo. B) Los sesgos del algoritmo en la toma de decisiones sanitarias. C) La opacidad algorítmica. 2. La necesaria transparencia. 3. La respuesta de la Ley de IA. V. DATOS PERSONALES E IA. 1. La toma de decisiones automatizadas y el perfilado de pacientes: un riesgo. 2. La anonimización de datos personales: una ilusión. 3. El Espacio Europeo de Datos Sanitarios. 4. La respuesta de la Ley de IA. VI. BIBLIOGRAFÍA.

Resumen: En el ámbito de la salud, la IA reviste un gran potencial tanto para el diagnóstico de enfermedades como para su detección precoz. La irrupción de la IA refuerza la implantación del modelo de medicina personalizada, que toma en consideración el genoma y que puede ser una valiosa herramienta de apoyo a la decisión médica. Además, la variedad de dispositivos digitales y las herramientas propias del internet de las cosas facilitan nuevos modelos de relación entre profesionales y pacientes que ofrecen la posibilidad de monitorizar a estos últimos y de asistirles a distancia.

El Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligen-

* Esta publicación es parte del Proyecto PID2022-136964NB-I00 El Derecho ante la Salud Digital, Personalizada y Robótica (SALUDPYR) financiado por MICIU/AEI/10.13039/501100011033/ y por FEDER, UE.

** *This publication is part of Project PID2022-136964NB-I00 The Law for Digital, Personalised and Robotic Health (SALUDPYR) funded by MICIU/AEI/10.13039/501100011033/ and by FEDER, EU.*

cia artificial y el Espacio Europeo de Datos Sanitarios son los instrumentos jurídicos para combatir las debilidades del algoritmo y, a su vez, para facilitar el despliegue de la IA en el ámbito de la salud.

Palabras clave: inteligencia artificial, salud, transparencia, datos personales, Espacio Europeo de Datos Sanitarios (EEDS).

Resum: En l'àmbit de la salut, la IA té un gran potencial tant per al diagnòstic de malalties com per a la seva detecció precoç. La irrupció de la IA reforça la implantació del model de medicina personalitzada, que pren en consideració el genoma i que pot ser una eina valuosa de suport a la decisió mèdica. A més, la varietat de dispositius digitals i les eines pròpies de l'internet de les coses faciliten nous models de relació entre professionals i pacients que ofereixen la possibilitat de monitoritzar aquests últims i assistir-los a distància.

El Reglament (UE) 2024/1689 del Parlament Europeu i del Consell, de 13 de juny de 2024, pel qual s'estableixen normes harmonitzades en matèria d'intel·ligència artificial i l'Espai Europeu de Dades Sanitàries són els instruments jurídics per combatre les debilitats de l'algoritme i, alhora, per facilitar el desplegament de la IA en l'àmbit de la salut.

Paraules clau: intel·ligència artificial, salut, transparència, dades personals, Espai Europeu de Dades Sanitàries (EEDS).

Abstract: In the field of health, AI has great potential for both disease diagnosis and early detection. The emergence of AI reinforces the implementation of the personalised medicine model, which takes the genome into account and can be a valuable tool for medical decision support. In addition, the variety of digital devices and the tools of the internet of things facilitate new models of relationship between professionals and patients that offer the possibility of monitoring and assisting patients remotely.

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and the European Health Data Space are the legal instruments to combat the weaknesses of the algorithm and, in turn, to facilitate the deployment of AI in healthcare.

Key words: Artificial Intelligence, Health, Transparency, Personal Data, European Health Data Space (EHDS).

I. EVOLUCIÓN Y CONCEPTO DE LA INTELIGENCIA ARTIFICIAL (IA)

Pese a que la eclosión de la IA parece reciente, ya en 1955 la programación hacía posible que el primer ordenador de IBM mejorara sus habilidades en el juego de damas¹ mediante el aprendizaje a través de su propia

¹ Fue Arthur L. Samuel (1901-1990) el creador de estos juegos que de modo incipiente apuntaban a lo que más tarde se conocería como inteligencia artificial [<https://darthmouthconference.wordpress.com/participantes-en-la-conferencia/arthur-samuel/>].

experiencia y fue en 1959 cuando se dio el nombre de inteligencia artificial a estos modelos que resolvían problemas complejos a partir de la información disponible.

Un paso más en su evolución lo constituyó el aprendizaje automático —*machine learning*—, de modo que la máquina, partiendo de información clasificada por el humano comenzó a generar patrones que le permitían, tras un proceso de aprendizaje nutrido por grandes cantidades de información, adoptar decisiones fundadas en su propia experiencia.

A modo de ejemplo, una de las áreas fundamentales en las que la IA se desenvuelve es en el análisis de imágenes médicas, como las radiografías, las tomografías computarizadas y las resonancias. Tras suministrar a la máquina muchas imágenes previamente analizadas por especialistas, esta identifica patrones y anomalías invisibles para el ojo humano, por lo que puede ayudar a una mayor precisión en el diagnóstico. Además, la IA se está utilizando como un valioso sistema de apoyo a la decisión médica, ya que a través del algoritmo pueden sugerirse diagnósticos o localizar áreas preocupantes en una imagen por lo que, siempre bajo el criterio último del profesional sanitario, facilita la tarea humana y mejora la decisión clínica.

Nos encontramos, no obstante, en este momento, en la superación de la fase anterior y en los albores de la siguiente. Ello implica que el algoritmo² ya está empezando a poder adoptar decisiones sin entrenamiento previo de decisiones humanas. Ello implica autonomía. Se trata de una inteligencia que parte de modelos de redes neuronales artificiales, lo que facilita un aprendizaje profundo, que trata de simular el cerebro humano y que recuerda a su anatomía, ya que implica capas de tratamiento de información organizadas desde un nivel de abstracción más bajo a uno más elevado, cuyas neuronas artificiales interactúan y proporcionan una decisión o información de salida no guiada por una programación previa.

La ventaja es que esta inteligencia permitirá diagnosticar, interpretar pruebas, detectar patologías en sus estadios iniciales, recabar información sobre enfermedades raras, desplegar la medicina personalizada y mejorar la gestión sanitaria, entre muchos otros hitos.

Los riesgos, no obstante, no son nada desdeñables y cabe referirse a las cuestiones que suscita esta tecnología disruptiva en pilares como la equidad, la igualdad, la transparencia y la rendición de cuentas, entre otros. Así, el uso de esta tecnología puede constituir una nueva brecha que acreciente

² «Conjunto ordenado y finito de operaciones que permite hallar la solución de un problema» [<https://dle.rae.es/algoritmo>].

la desigualdad y es inherente al algoritmo que opera bajo IA profunda —autónoma— que la mente humana no llegue a saber por qué un sistema de IA ha adoptado una determinada decisión. Aquí es donde el Derecho tiene que preguntarse si nuestro ordenamiento jurídico está preparado para hacer frente a los retos, que, cada vez de mayor calado, acompasados con una mayor autonomía de la IA, se van a plantear. No obstante, la regulación de este escenario resulta tan incipiente como el propio desarrollo algorítmico. De lo que se trata es que ambos evolucionen a la vez y el Derecho no quede rezagado, ya que es fundamental para paliar estos riesgos contar con un ordenamiento jurídico adecuado al fenómeno de la IA.

Hoy no podemos todavía hablar de una IA autónoma generalizada, pero ante la irrupción de este nuevo escenario —que se prevé distópico— la Comisión Europea, en abril de 2021, lanzó la primera regulación de IA, la Propuesta de Reglamento por el que se establecen normas armonizadoras en materia de inteligencia artificial³ y sus anexos⁴ que hoy se ha convertido en el Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial.⁵

El Reglamento define el sistema de IA como «un sistema basado en una máquina que está diseñado para funcionar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación tras el despliegue, y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales».⁶ Apuesta por una regulación transversal y por el equilibrio entre la minimización de riesgos y su promoción como tecnología

³ Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (ley de inteligencia artificial) y se modifican determinados actos legislativos de la unión Bruselas. 21.4.2021 COM(2021) 206 final 2021/0106 (COD) [https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-95851aa75e-d71a1.0008.02/DOC_2&format=PDF].

⁴ Anexos de la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la unión. Bruselas, 21.4.2021 COM(2021) 206 final [https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75e-d71a1.0008.02/DOC_2&format=PDF].

⁵ Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n° 300/2008, (UE) n° 167/2013, (UE) n° 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial). «DOUE» núm. 1689, de 12 de julio de 2024.

⁶ Art. 3.1.

beneficiosa para la humanidad, adoptando el reglamento que implicará una aplicación directa en los Estados miembros, lo que garantizará la uniformidad en la regulación.

Con todo, el consenso sobre una definición de IA es complejo, ya que se refiere a un cajón de sastre que abarca a conjunto de ciencias, teorías y técnicas destinadas a mejorar la habilidad de las máquinas en la realización de tareas que requieren inteligencia.⁷

II. IA Y SALUD

En el ámbito de la salud, la IA reviste un gran potencial tanto para el diagnóstico de enfermedades como para anticiparse a las mismas mediante la detección precoz que permitirá adoptar medidas de prevención. Ello es posible gracias a la capacidad sobresaliente de utilizar cantidades ingentes de información generada por la asistencia y por la investigación biomédica.⁸ A través de la tecnología *big data*⁹ la IA aprende y se nutre de grandes cantidades de información procedente de los historiales clínicos de pacientes y analiza a velocidad de vértigo las pruebas de imagen, liberando al humano de tareas tediosas y, a su vez, proporcionándole herramientas de gran valor para la toma de decisiones. En ocasiones, incluso son posibles diagnósticos impensables con las técnicas convencionales, como sucede con las enfermedades raras o poco prevalentes, ya que, a partir de análisis precisos y de los datos de registros que hasta el momento no se conectaban, se identifican matices que no eran perceptibles que posibilitan el conocimiento de la patología y de su tratamiento. Tanto es así, que podemos hablar de un nuevo estado de la ciencia, que es la decisión de la máquina y a la que el humano médico no puede permanecer ajeno, so pena de incurrir en infracción de la *lex artis*.¹⁰

⁷ BORGES BLÁZQUEZ, R. «El sesgo de la máquina en la toma de decisiones en el proceso penal». *Ius et Scientia*, vol. 6, núm. 2 (2020), pág. 55.

⁸ COTINO HUESO, L. «Inteligencia artificial, big data y aplicaciones contra la Covid-19: privacidad y protección de datos». *IDP: Revista de Internet, Derecho y Política*, núm. 31 (2020).

⁹ Big Data significa macrodatos, que a su vez significa «Conjunto de datos que, por su gran volumen, requieren técnicas especiales de procesamiento» [<https://dle.rae.es/macrodatos#TdGiPT3>]. Big data implica el procesamiento de un gran volumen de datos, variados en cuanto que —como hemos señalado— son estructurados y no estructurados, que se tratan a una gran velocidad por medio de tecnologías no convencionales, válidos y que aportan un valor. GIL MEMBRADO, C. *Riesgos del uso de algoritmos en el diagnóstico y en la investigación biomédica*. VIII Premio Nacional de Derecho Sanitario, Reus, Tarragona, 2023, pág. 68.

¹⁰ «Conjunto de reglas técnicas a que ha de ajustarse la actuación de un profesional en ejercicio de su arte u oficio. Su cumplimiento es determinante de la exoneración de responsabilidad en

La irrupción de la IA coadyuva al nuevo modelo de medicina personalizada, que toma en consideración la variabilidad individual del genoma, de modo que una de las grandes revoluciones en el ámbito oncológico es la atención a las características genómicas y moleculares del tumor específico del paciente para adoptar decisiones terapéuticas de modo personalizado.

Además de constituirse como un valioso decisor o una herramienta de apoyo a la decisión médica, la IA también puede llegar allá donde el humano no alcanza o su capacidad es inferior. Por ejemplo, es un poderoso instrumento para llevar a cabo procedimientos quirúrgicos donde se requiera una precisión más allá de las capacidades humanas. A modo de ejemplo, el robot cirujano es idóneo para intervenciones como la de implante coclear para horadar el conducto con una precisión que supera la del cirujano humano. Todavía es este el que dirige a la máquina, pero llegará un momento en que la máquina será capaz de hacerlo con plena autonomía. Lo mismo sucede con el robot *Da Vinci*¹¹ que implica un sistema de laparoscopia con brazo articulado robótico controlado a distancia por el cirujano. Todo ello, aunque no sea objeto de este trabajo, implica que, mientras la complejidad y autonomía de la máquina no alcance a no necesitar control por el humano, nuestros parámetros de responsabilidad civil serían trasladables con algunos ajustes si la máquina generara daños, pero, cuando un robot tome sus decisiones sin intervención humana, precisamente los rasgos de opacidad a los que nos referiremos motivarán un nuevo abordaje de la responsabilidad.¹²

caso de que el ejercicio profesional haya ocasionado daños a terceros: “[...] la actuación médica dispensada al recurrente se ajustó a las exigencias de la *lex artis*. [...] la asistencia sanitaria es una obligación de medios y no de resultados, de modo que no puede apreciarse responsabilidad cuando se han empleado las técnicas y los instrumentos de que se dispone” (STS 18.7.2016, rec. 4139/2014)». Diccionario Panhispánico del español jurídico [<https://dpej.rae.es/lema/lex-artis>].

¹¹ Desarrollado por la empresa norteamericana Intuitive Surgical y aprobado, en el año 2000, por la Administración de Alimentos y Medicamentos (FDA) de los Estados Unidos [<https://www.intuitive.com/en-us/products-and-services/da-vinci>].

¹² Después de no pocos bandazos, a consecuencia principalmente del difícil encaje de la responsabilidad culpabilística y del entramado de intereses que circundan la inteligencia artificial actualmente hay dos propuestas de Directiva relacionadas con esta materia: Propuesta de Directiva del Parlamento Europeo y del Consejo relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial (Directiva sobre responsabilidad en materia de IA) COM(2022) 496 final 2022/0303 (COD), Bruselas, 28.9.2022 [<https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52022PC0496>]; Propuesta de Directiva del Parlamento Europeo y del Consejo sobre responsabilidad por los daños causados por productos defectuosos Bruselas, 28.9.2022 COM(2022) 495 final 2022/0302(COD) [<https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:52022PC0495>]. Véase GIL MEMBRADO, C. «Daños producidos por la IA: la opacidad del algoritmo y el efecto de caja negra», en ÁLVAREZ LATA, N. (coord.). *Derecho*

Además, la variedad de dispositivos digitales y las herramientas propias del internet de las cosas (IoT)¹³ están forjando nuevos modelos de relación entre profesionales y pacientes que ofrecen la posibilidad de monitorizar a estos últimos y de asistir a distancia. Así, la salud móvil permite el control de la frecuencia cardíaca, el nivel de glucosa en sangre, la presión arterial, la temperatura corporal y la actividad cerebral, incluso el recordatorio de que hay que tomar un medicamento, o su administración.¹⁴

En lo que se refiere al desarrollo de nuevos fármacos, la IA está realizando grandes avances en el diseño de moléculas. En la optimización de los ensayos clínicos la IA tiene un gran papel y en cuanto a los estudios genéticos la IA facilita la obtención y la interpretación de la secuenciación genómica, así como la actuación sobre el genoma: ejemplos son las técnicas para obtener una imagen del genoma completo; encontrar la mejor terapia farmacológica personalizada en función del ADN de la célula;¹⁵ la mejora de la salud mediante la alteración genética de péptidos¹⁶ o el diseño de fármacos específicos según el perfil genético.¹⁷ Una segunda fase de la aplicación de la IA al tratamiento personalizado de cada tumor estriba en que los análisis genómicos y moleculares realizados compartidos por la investigación clínica repercutirá en el conocimiento oncológico global fomentando la generación de nuevos tratamientos.

de contratos, responsabilidad extracontractual e inteligencia artificial. Aranzadi, Navarra, 2024, págs. 501-547.

¹³ Esta terminología fue utilizada por primera vez por Kevin Ashton en 1999, un ingeniero del centro tecnológico de la Universidad Tecnológica de Massachusetts (MIT). El término intenta definir cómo una multitud de objetos serán capaces de comunicarse a través de Internet y facilitar la vida diaria [<https://mosaic.uoc.edu/2018/04/05/iot-el-internet-de-las-cosas/>].

¹⁴ NAVAS NAVARRO, S. «Aspectos jurídicos de las aplicaciones móviles de salud (apps sanitarias y wearables)». *Diario La Ley*, 14 de mayo de 2020, La Ley 5811/2020.

¹⁵ GIL MEMBRADO, C. *Riesgos del uso de algoritmos en el diagnóstico y en la investigación biomédica*. Reus, Tarragona, 2023.

¹⁶ Molécula que contiene dos o más aminoácidos (las moléculas que se unen entre sí para formar proteínas). Los péptidos pueden contener muchos aminoácidos que se llaman polipéptidos o proteínas [<https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/peptido>].

¹⁷ BENAVENT NÚÑEZ, D.; COLOMER MASCARÓ, J. [et al.]. *Inteligencia Artificial y decisiones clínicas: cómo está cambiando el comportamiento médico*. Fundación Gaspar Casal, Madrid, 2020, págs. 24 y ss.

III. LA LEY DE IA

GAMERO CASADO¹⁸ señalaba —antes de la aprobación de la Ley de IA— que no contar con una regulación específica de la IA no significa que su despliegue esté carente de regulación —habla en concreto de que no se trataría de una actividad sin ley, como se diría en el lejano oeste—. Apunta que existen principios y reglas generales del Derecho aplicables a la IA como lo son a cualquier actividad —igualdad, protección de datos, necesidad de motivación de decisiones algorítmicas—. No obstante, no podemos ser ajenos a las singularidades propias de la IA y a su impacto en los bienes y en los derechos de la ciudadanía. Por ello, el enfoque de la UE con relación a la IA se centra en la excelencia y en la confianza, con el objetivo de impulsar su uso garantizando, a la vez, la seguridad y los derechos fundamentales.¹⁹

1. ANTECEDENTES

En el espacio europeo tanto la Comisión Europea, como el Parlamento Europeo, como el Comité Económico y Social vienen buscando una armonización de las normas comunitarias para regular el nuevo escenario de la inteligencia artificial.

El Comité Económico y Social Europeo publicó en 2017 el Dictamen sobre «Inteligencia artificial: las consecuencias de la inteligencia artificial para el mercado único (digital), la producción, el consumo, el empleo y la sociedad»²⁰ con el propósito de unir en esta nueva era la economía, la industria y la sociedad, mediante la dotación de soluciones que sirvan como base del proyecto de integración europea.

La Comisión Europea en la «Comunicación sobre inteligencia artificial para Europa»²¹ hace mención con referencia a la IA a los sistemas que manifiestan un comportamiento inteligente, al ser capaces de analizar su entorno y pasar a la acción con cierto grado de autonomía. Uno de los hitos va a ser la variedad de sistemas que pueden subsumirse en esta tecnología, de ahí que abarque desde un mero programa informático, como puede ser un

¹⁸ GAMERO CASADO, E. «El enfoque europeo de inteligencia artificial». *Revista de Derecho Administrativo*, CDA 20 (2021), pág. 270.

¹⁹ Un enfoque europeo de la inteligencia artificial [<https://digital-strategy.ec.europa.eu/es/policies/european-approach-artificial-intelligence>].

²⁰ <https://eur-lex.europa.eu/legal-content/ES/ALL/?uri=CELEX%3A52016IE5369>

²¹ Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. Inteligencia artificial para Europa COM/2018/237 final [<https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=COM:2018:237:FIN>].

asistente de voz, un programa que analice imágenes, un motor de búsqueda o sistemas de reconocimiento facial. También la IA puede estar incorporada a dispositivos de hardware como robots, automóviles, drones o las diversas aplicaciones del internet de las cosas (IoT). Así, lejos de ser ciencia ficción, la IA se ha incorporado con plenitud a nuestro día a día.

Por ello, la Comisión Europea en la citada Comunicación ya considera necesario un planteamiento coordinado que aborde los retos que plantea la IA. En lo que nos interesa, resulta fundamental que esta nueva tecnología respete los valores y derechos de la Unión.

Posteriormente, la propia Comisión lanza la Comunicación «Generar confianza en la inteligencia artificial centrada en el ser humano».²² En esta ocasión se afirma que la IA no es un fin en sí mismo, sino un medio para servir a la persona e incrementar su bienestar, por lo que la dignidad humana y los derechos de las personas deben situarse por encima de la capacidad de la máquina de aprender, de tomar decisiones y de ejecutarlas sin intervención humana. Para ello, además de ajustarse a la ley, el despliegue de esta tecnología debe respetar unos principios éticos en cada una de las fases de su desarrollo de modo que su objetivo debe ser mejorar la capacidad de la persona, no sustituirla.

Un hito en esta evolución lo constituye el «Libro Blanco de la UE sobre inteligencia artificial»,²³ que fija los pilares para promover el uso de la IA y para el abordaje de los riesgos vinculados a los usos de esta tecnología. Precisamente en su texto propone que la IA, al utilizarse en escenarios de riesgo, como sucede con la salud, tiene que ser transparente con el objeto de comprobar la trazabilidad y para hacer posible la intervención humana en el proceso decisorio. Además, incide en la necesidad de contar con datos no sesgados con el objeto de que el funcionamiento de estas tecnologías garantice la efectividad de los derechos fundamentales.

En el Parlamento Europeo también se han impulsado iniciativas como, entre otras, el «Marco de los aspectos éticos de la inteligencia artificial, la

²² Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. Generar confianza en la inteligencia artificial centrada en el ser humano. COM/2019/168 final [<https://eur-lex.europa.eu/legal-content/es/ALL/?uri=CELEX-%3A52019DC0168>].

²³ Libro Blanco sobre la inteligencia artificial - un enfoque europeo orientado a la excelencia y la confianza. Bruselas, 19.2.2020 COM(2020) 65 final [<https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52020DC0065>].

robótica y las tecnologías conexas»²⁴ en favor de una IA antropocéntrica y antropogénica y de la necesidad de contar con un marco regulador armonizado basado en el Derecho de la Unión, la Carta y el Derecho internacional en materia de derechos humanos, en particular en lo referente a la tecnología de alto riesgo. Entre otros principios hace referencia a «la dignidad humana, la autonomía y la autodeterminación de la persona, la prevención de daños, la promoción de la equidad, la inclusión y la transparencia, la eliminación de los sesgos y la discriminación, también en lo que respecta a los grupos minoritarios, así como respetar los principios de limitación de las externalidades negativas de la tecnología empleada, de explicabilidad de las tecnologías y de garantía de que las tecnologías sirvan a las personas y no las sustituyan ni decidan por ellas, con el objetivo último de elevar el bienestar de todos los seres humanos». Para ello solicita dar solución a riesgos como «los que atañen a la recopilación y el uso ético de los macrodatos, la cuestión de la transparencia algorítmica y el sesgo algorítmico».

2. EL REGLAMENTO (UE) 2024/1689 POR EL QUE SE ESTABLECEN NORMAS ARMONIZADAS EN MATERIA DE INTELIGENCIA ARTIFICIAL (LEY DE IA)

El texto más ambicioso en materia de IA hasta el momento resultó propuesto por la Comisión en abril de 2021²⁵ con el objetivo de abordar el uso de la IA en función del riesgo que revista —inaceptable, alto, limitado y mínimo— y con el objeto de dotar un marco jurídico uniforme para el desarrollo, la comercialización y el uso de la IA en la Unión conciliándolo con la salvaguarda de los derechos frente a los riesgos inherentes a esta tecnología.

Con relación a la definición de la IA y considerando que no ha sido una cuestión pacífica por la variedad de tecnologías que abarca, se opta por acoger la propuesta por la Organización para la Cooperación y el Desarrollo Económico (OCDE).²⁶ Esta armonización permite una mayor adecuación a

²⁴ 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)).

²⁵ En su momento, Propuesta de reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (ley de inteligencia artificial) y se modifican determinados actos legislativos de la Unión COM(2021) 206 final 2021/0106 (COD), Bruselas, 21.4.2021.

²⁶ *An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.*

la interacción con otras normas que también se adhieran a las recomendaciones de la OCDE.

En lo que se refiere al ámbito de aplicación, con vocación de aplicación extraterritorial, rige para agentes públicos y privados, en el interior y fuera de la UE, bien porque se introduzca en el mercado de la UE o bien porque afecte a personas establecidas en la Unión.

No afecta a las competencias de los Estados miembros en materia de seguridad nacional, ni se aplica a los sistemas de IA utilizados con fines exclusivamente militares o de defensa, ni a los sistemas de IA usados únicamente con fines de investigación e innovación, ni a quien utilice estos sistemas con fines no profesionales. En lo referente a la investigación y a la innovación, parece que ha pesado la comunidad *open source* preocupada por las barreras que esta regulación pudiera desplegar frente a la investigación y a la innovación. Con el propósito de dar una respuesta la previsión es el establecimiento de una excepción, aunque limitada a algunos casos concretos para modelos gratuitos de IA fundados en código abierto, si bien con el respeto a la transparencia y a los derechos de propiedad intelectual.²⁷

En relación con el enfoque basado en el riesgo implica la observancia de mayores cautelas cuanto mayor sea el riesgo. Se parte de que la gran mayoría de los sistemas de IA revisten un mínimo riesgo, como sucede, por ejemplo, con los filtros de spam. Para estos no se establecen obligaciones más allá de que las empresas, de modo voluntario, suscriban códigos de conducta.

Para los sistemas de alto riesgo, los requisitos son mayores e implican la mitigación de riesgos, los datos de alta calidad, el registro de actividad, documentación detallada, información a los usuarios de que interactúan con IA, supervisión humana, además de muestra de solidez, precisión y medidas de ciberseguridad. Como ejemplos de sistemas de IA de alto riesgo son infraestructuras críticas como aquellas relacionadas con el agua, el gas y la electricidad; productos sanitarios; sistemas que determinen el acceso a instituciones educativas o a contrataciones; determinados sistemas usados por la policía, para el control fronterizo, la administración de justicia y los procesos democráticos. Además, aquellos sistemas de IA para identificación biométrica, categorización y reconocimiento de emociones. En el Ane-

²⁷ ZORRAQUINO A.; MATAS BRANCÓS, A.; VILANOVA JOU, R. «El Reglamento de IA, cada vez más cerca: acuerdo preliminar del Consejo y el Parlamento Europeo». *Newsletter de PwC Tax & Legal*, PWC, diciembre 2023 [https://periscopiofiscalylegal.pwc.es/wp-content/uploads/2023/12/El-Reglamento-de-IA-cada-vez-mas-cerca.-Acuerdo-preliminar-del-Consejo-y-el-Parlamento-Europeo-v_5.pdf].

xo III se recogen los casos considerados de alto riesgo y es obligación de la Comisión velar por que la lista se vaya actualizando.

En lo que se refiere a la identificación biométrica remota en tiempo real en espacios de acceso público, se permite su uso excepcionalmente para casos de víctimas de determinados delitos, de prevención de amenazas reales, presentes o previsibles como los atentados terroristas y la búsqueda de personas sospechosas de delitos graves, y para ello se establecen obligaciones adicionales.

Más allá hay otros sistemas cuyo riesgo se considera inadmisibile, de lo que deriva su prohibición. Se trata de aquellos que amenacen de modo claro los derechos fundamentales de la persona. En esta categoría encontraríamos a los sistemas que manipulen el comportamiento humano para eludir su voluntad —como sucedería con juguetes que por medio de asistencia vocal inciten a comportamientos peligrosos a los menores o sistemas que permitan la puntuación social por gobiernos o empresas, así como determinadas aplicaciones predictivas policiales—. Además, están prohibidos algunos usos de sistemas biométricos como los de reconocimiento de emociones en el lugar de trabajo.

Especial atención se presta al riesgo específico que se genera al utilizar robots conversacionales, frente a los que se impone obligaciones de transparencia con el objeto de que los usuarios sean conscientes de su interacción con una máquina. También la obligación específica de transparencia se establece respecto de falsificaciones y otros contenidos generados por IA. La transparencia tiene una gran importancia para paliar los riesgos del contenido del tipo *deepfakes*²⁸ que se pueden crear mediante herramientas de IA generativa. Los usuarios también deberán ser informados cuando se utilice la categorización biométrica o el reconocimiento de emociones. Se impone la obligación a los proveedores para que en el diseño de los sistemas se facilite el reconocimiento de la generación o manipulación artificial de vídeos, audios, textos o imágenes.

Todo ello aborda la IA de uso general, que es aquella que puede utilizarse para fines muy distintos, y los modelos fundacionales, que pueden realizar

²⁸ «Los *deepfakes* son vídeos manipulados para 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». INCIBE. Instituto Nacional de Ciberseguridad [<https://www.incibe.es/aprendeciberseguridad/deepfakes>].

una gran cantidad de tareas como sucede con el uso de IA generativa capaz de crear texto o vídeos.

En estos supuestos, si la IA es tan potente que puede provocar riesgos sistémicos,²⁹ pudiendo causar graves accidentes o ser utilizados para ciberataques, se establecen obligaciones suplementarias sobre gestión de riesgos, seguimiento de incidentes graves, evaluación de modelos y emulación de adversarios, entre otras obligaciones, que se pondrán en práctica mediante códigos de conducta.

En cuanto a la gobernanza, se crea una nueva Oficina Europea de Inteligencia Artificial —Oficina de IA— dentro de la Comisión Europea, que, junto con las autoridades nacionales competentes en materia de vigilancia del mercado, se convierte en el primer órgano mundial que aplicará normas vinculantes en la materia.

IV. TRANSPARENCIA Y EXPLICABILIDAD EN IA PARA LA SALUD

Hemos hecho alusión a los riesgos que en materia de transparencia puede infligir el uso de la IA. Ya en el Libro Blanco,³⁰ la Comisión apuntaba la conveniencia de mejorar el marco normativo para el abordaje de la falta de transparencia (opacidad de la IA).

COTINO HUESO concibe la transparencia algorítmica como un concepto general inclusivo que abarca «trazabilidad, explicabilidad, interpretabilidad, comprensibilidad, inteligibilidad, legibilidad, auditabilidad, testabilidad, comprobabilidad, simulabilidad, descomponibilidad, verificabilidad, replicabilidad, comunicación, código abierto, así como referencias a la transparencia técnica en sentido estricto».³¹

²⁹ Se entiende que los modelos de IA de uso general entrenados con una potencia informática superior a 10^{25} operaciones de coma flotante por segundo (FLOPS) revisten riesgos sistémicos. La Oficina Europea de Inteligencia Artificial irá revisando este umbral para adecuarlo al estado de la tecnología. Por el momento, se fija por encima del umbral de 10^{25} porque este es el propio de modelos de sistemas de IA de uso general más avanzados en la actualidad (GPT-4 de OpenAI y Gemini de Google DeepMind). Capacidades superiores todavía son incomprensibles en grado suficiente por lo que podrían plantear riesgos sistémicos y, por ello, se somete a sus proveedores a obligaciones adicionales. Inteligencia artificial: preguntas y respuestas. 12 de diciembre de 2023 [https://ec.europa.eu/commission/presscorner/detail/es/QANDA_21_1683].

³⁰ Libro Blanco sobre la inteligencia artificial - un enfoque europeo orientado a la excelencia y la confianza. Bruselas, 19.2.2020 COM(2020) 65 final, pág. 17 [<https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52020DC0065>].

³¹ COTINO HUESO, L. «Transparencia y explicabilidad de la inteligencia artificial y “compañía” (comunicación, interpretabilidad, inteligibilidad, auditabilidad, testabilidad, comprobabilidad, simu-

DE LA NUEZ SÁNCHEZ-CASCADO³² se refiere a casos concretos planteados en este sentido como el descubrimiento de que el algoritmo usado para analizar los riesgos de salud de unos doscientos millones de pacientes en Estados Unidos discriminaba sistemáticamente a la población de color en el sistema sanitario. Con el objetivo de asignar recursos sanitarios, el algoritmo segmentaba a los pacientes en atención al riesgo médico asociado a sus características de modo que los menos saludables podían acceder a un programa específico sanitario. A consecuencia de los sesgos algorítmicos miles de personas de color quedaron fuera del programa a pesar de que su salud era peor que la de muchas personas de raza blanca que sí que accedieron. Otro caso similar, se produjo en 2015 cuando se descubrió que la aplicación *Mountain View* de reconocimiento de imágenes etiquetaba a las personas de raza negra como gorilas a consecuencia de que para el reconocimiento de personas frente a animales a la máquina se le había suministrado imágenes de personas de raza blanca.

Por ello, resulta oportuna la previsión de la Ley de IA, que ya en su exposición de motivos apunta que uno de sus objetivos es que los sistemas de IA sean seguros y respeten la normativa en materia de derechos fundamentales. Resulta necesario para contar con una IA segura, confiable y ética que garantice la explicabilidad del algoritmo, o, lo que es lo mismo, la trazabilidad de cada uno de los pasos seguidos por la IA en el proceso decisorio que afecta a la persona y a su derecho no solo a controlar su información personal, sino a no ser objeto de decisiones sesgadas, erróneas o discriminatorias.

1. EL ALGORITMO Y SUS DEBILIDADES

La irrupción de la IA ha provocado una gran admiración por las expectativas que genera, pero una de las primeras cuestiones que plantea es si la decisión del algoritmo es mejor que la humana, ya que, la ventaja básica del algoritmo sobre la mente humana radica en su capacidad para procesar datos y aprender de ellos.

La tecnología *big data* tiene una capacidad exponencial para procesar datos recabados de una gran cantidad de fuentes con el objeto de realizar análisis predictivos y de proporcionar decisiones automatizadas que, si bien

labilidad. ...) Para qué, para quién y cuánta», en COTINO HUESO, L.; CASTELLANOS J. (coords.). *Transparencia y explicabilidad de la inteligencia artificial*. Tirant lo Blanch, Valencia, 2022.

³² DE LA NUEZ SÁNCHEZ-CASCADO, E. «Inteligencia artificial y transparencia». *La Ley Derecho Digital e Innovación*, 6786, núm. 5 (2020).

son sumamente valiosas, a modo de ejemplo, para la gestión sanitaria, no están exentas de sesgos que pueden dar lugar a decisiones erróneas o discriminatorias y de cierta imprevisibilidad provocada por el comportamiento autónomo y por su complejidad desde la perspectiva técnica.

BARRIO ANDRÉS³³ se remite a un ejemplo que da muestra de la torpeza y de la rigidez del algoritmo. Narra que en 2016 en el Departamento de Servicios Sociales de Arkansas se sustituyó a una enfermera que realizaba visitas domiciliarias a los pacientes discapacitados para evaluar sus necesidades de cuidado y de asistencia por un software que utilizaba complejos cálculos mediante un algoritmo. Los resultados, en algunos casos, fueron ilógicos y arbitrarios. Así, las necesidades de una persona que había sufrido la amputación de un pie eran inexistentes porque, al no tener pie, el sistema arrojaba el resultado de que la persona no tenía ningún problema en el pie, cuando la realidad es que sus necesidades eran muy superiores a las de otra persona con una lesión en el pie —que sí se consideraba como necesitado de cuidado y de asistencia—.

AÑÓN ROIG señala que la IA introduce un grado de incertidumbre provocado por el riesgo algorítmico basado en modelos matemáticos capaces de procesar sujetos, reconocer patrones, predecir conductas y adoptar decisiones por medio de la estadística computacional. Estos modelos han producido un desplazamiento de la probabilidad clásica que se fundamenta en el principio de causalidad. En su lugar, los razonamientos tienen lugar por medio de correlaciones múltiples y no evidentes, lo que provoca un grado de opacidad en mayor o menor medida. Es el precio de incrementar la eficiencia: la eventual introducción de sesgos, en ocasiones discriminatorios, en el razonamiento artificial.³⁴

A) *Cómo funciona el algoritmo*

El algoritmo no es más que una serie de instrucciones para resolver problemas que ya lleva mucho tiempo formando parte de las decisiones cotidianas al margen de la IA —en una simple hoja de *Excel* ya está presente—.

Para cumplir su función, es decir, hacer recomendaciones, predicciones o adoptar decisiones hace uso de información proporcionada por humanos o por otras máquinas y a partir de esta, mediante un procedimiento de

³³ BARRIO ANDRÉS, M. «Luces y sombras del Estado algorítmico de Derecho». *Derecho Digital e Innovación*, núm. 5 (segundo trimestre 2020), La Ley 5962/2020, pág. 4.

³⁴ AÑÓN ROIG, M. J. «Desigualdades algorítmicas: conductas de alto riesgo para los derechos humanos». *Derechos y libertades*, núm. 47, Época II, junio (2022), págs. 22 y 23.

abstracción, genera patrones respecto de los que extrae recomendaciones, predicciones o decisiones.

Cuanto más avanzado es el algoritmo, se alimenta de más datos y se vuelve más autónomo, a la vez que realiza mejores predicciones, pero cuando llega a un determinado nivel de abstracción, su funcionamiento resulta incomprensible para la mente humana. Conforme se va volviendo más complejo, el algoritmo bebe de más variables por lo que es imposible reconocer los patrones en los que la máquina se ha basado para tomar una determinada decisión.³⁵

B) Los sesgos del algoritmo en la toma de decisiones sanitarias

Al referirnos al Libro Blanco, se ha adelantado la posibilidad de que una información sesgada opere en detrimento de los derechos fundamentales. Si la IA se alimenta de datos sesgados generará correlaciones falsas y errores que podrían generar efectos discriminatorios, ya que el resultado del algoritmo puede constituir prejuicios y patrones excluyentes.

El sesgo implica un «error sistemático en el que se puede incurrir cuando al hacer muestreos o ensayos se seleccionan o favorecen unas respuestas frente a otras».³⁶ Cuando una decisión se adopta por una máquina, igual que cuando se adopta por un humano, el sesgo es inevitable. Los macrodatos —como el cerebro humano— no son neutrales. Sin embargo, existe una —en ocasiones— fe ciega en el algoritmo bajo la creencia de que su objetividad es muy superior a la humana.

Con anterioridad se ha hecho referencia a que el algoritmo se entrena con grandes cantidades de datos y puesto que nos encontramos en un escenario concreto haremos referencia a los tipos de sesgos cognitivos en el diagnóstico médico descritos por PHUA y TAN. Los autores hacen referencia al sesgo de disponibilidad que implica que la elección de un diagnóstico puede verse condicionado por la mayor presencia en la mente del profesional, por ejemplo, al haberse encontrado recientemente con un paciente con esa patología; el de anclaje, que implica fijar o anclar el diagnóstico al inicio del proceso sin revisarlo con posterioridad al contar con nueva información; el de la representatividad, que supone atender a síntomas muy representativos de un determinado diagnóstico y no atender a otros que podrían ser indicativos de otro diagnóstico; la falacia del jugador implica apreciar

³⁵ BORGES BLÁZQUEZ, R. «El sesgo de la máquina...», *op. cit.*, págs. 58 y 59.

³⁶ <https://dle.rae.es/sesgo>

de modo erróneo la probabilidad de concurrencia de dos sucesos que, sin embargo, son independientes —así sucede con el jugador que piensa que si no ha tenido ocasión en una partida es más probable que sí lo tenga en la siguiente, igual que el médico piensa que si un paciente se ha presentado en la consulta con una enfermedad grave es poco probable que el mismo día se presente otro aquejado de la misma dolencia—; detener la búsqueda cuando se ha realizado un diagnóstico aunque aparezca una situación fuera de lo normal en el paciente; la influencia del etiquetado y decisiones previas que supone atender sin cuestionar la opinión de otras personas; el coste de la inversión que supone haber invertido recursos conforme al diagnóstico ya realizado; maximizar y minimizar probabilidades en función de los resultados, que implica subestimar la probabilidad de que concurren eventos negativos y sobreestimar la probabilidad de que concurren los positivos —pensar que una complicación resulta menos probable de lo que es en realidad—; el sesgo retrospectivo implica que una vez conocido el desenlace —por ejemplo, la muerte del paciente— se sobreestima la posibilidad de llegar al diagnóstico idóneo olvidando que antes de conocer dicho desenlace, la información no era tan clara y, por lo tanto, el diagnóstico no podía ser concluyente; finalmente, se apunta que el profesional tiende a sobreestimar su pericia. Los referenciados autores también se refieren a sesgos de carácter emocional o afectivo como la incomodidad de los pacientes o de la familia, que puede abocarle a acortar la exploración, o, al contrario, la insistencia del paciente o de los familiares en la realización de más pruebas diagnósticas de las indicadas o las propias emociones del profesional, a lo que no son ajenas las decisiones.³⁷

En atención a todo ello, se puede concluir que, puesto que la IA se alimenta de historias clínicas y de otra documentación clínica, por lo menos, adoptará y amplificará en su información de salida todos los citados sesgos propios de humanos.

En una IA débil —no profunda y, por lo tanto, con un grado de opacidad asumible— puesto que la lógica explicacional del algoritmo es inteligible cabrá analizar de dónde procede una información de salida sesgada, sin embargo, no sucederá así a medida que el algoritmo vaya siendo más complejo, más avanzado y, por lo tanto, más opaco. Nos referiremos a ello en lo sucesivo.

³⁷ PHUA, D.H.; TAN N.C. «Cognitive aspect of diagnostic errors». *Annals of de Academy of Medicine*, Singapore, núm. 1, vol. 42 (2013), págs. 33-41.

C) La opacidad algorítmica

Es inherente al proceso de decisión del algoritmo, en mayor medida cuanto más complejo, el conocimiento de la información de entrenamiento y la decisión de salida, pero no lo que sucede a lo largo del razonamiento. La opacidad del algoritmo puede generar riesgos para los derechos fundamentales, para lo que resulta necesaria una acción legislativa común para que este proceso sea transparente.

En la práctica clínica, el profesional de la salud contará con una potente herramienta, que, sin duda, gracias a la velocidad y a la eficacia en el tratamiento de información, atenderá a matices que pasan desapercibidos al ojo clínico y al cerebro del profesional humano, que, por contrapartida, no siempre va a ser capaz de conocer el proceso seguido por la máquina, de modo que, en este caso, se antoja una confianza ciega en la máquina.

Para minimizar este riesgo el principio de transparencia se alza como básico para que, en la utilización de la IA profunda, el humano pueda conocer el proceso seguido por la máquina desde la información de entrada (*inputs*) hasta la de salida (*outputs*).

De lo contrario, el profesional sanitario tendrá en sus manos una potente herramienta, que, sin embargo, no será más que una caja negra respecto de la que ni el propio programador sabrá lo que sucede en su interior. Así las cosas, el profesional no podrá descartar o validar un diagnóstico emitido por IA puesto que no será conocedor de la lógica decisional algorítmica. En atención a todo lo apuntado, difícilmente el profesional actuaría conforme a la *lex artis* si acepta ciegamente los postulados de la IA. En este caso, nos veríamos inmersos en una especie de dictadura algorítmica partiendo del dogma de fe de que la programación y el aprendizaje por parte del algoritmo se ha realizado correctamente.

Con todo, debemos ser conscientes de que, aunque partamos de la transparencia técnica que conlleva utilizar códigos fuente abiertos,³⁸ los algoritmos de aprendizaje profundo, que, precisamente, son aquellos más valiosos en cuanto a su potencial, son opacos incluso para sus desarrolladores, por lo que la ininteligibilidad de la caja negra, por el momento, está garantizada al margen de los esfuerzos que se puedan realizar para intentar garantizar, en la medida de lo posible, la explicabilidad con el objeto de contar con una IA fiable.

³⁸ GIL MEMBRADO, C. «La apertura del código fuente y la transparencia en la contratación de servicios», en COBAS COBIELLA, M. E.; GUILLÉN CATALÁN, R. (dirs.). *Equidad y transparencia en la prestación de servicios*. Dykinson, Madrid, 2023.

Nos encontramos en un ámbito, el de salud, que, como sucede en muchos otros de la vida, la imprevisibilidad, la incertidumbre y la inseguridad no son aceptables, por lo que todas las medidas tendentes a promover y comprobar las garantías en el diseño y en el entrenamiento de los algoritmos serán bienvenidas, si bien, debemos ser conscientes de los riesgos a los que estamos expuestos. De todos modos, la transparencia a exigir deberá ser mayor cuanto mayor sea el impacto y el riesgo generado por la IA, así cuanto menor sea la intervención humana.

2. LA NECESARIA TRANSPARENCIA

MAGDALENA LAYOS³⁹ contrapone el «modelo transparente» en el que la abstracción matemática es comprensible, legible e interpretable por un humano, frente al «modelo de caja negra» que no es capaz de describir el porqué de una respuesta para un humano.

Ello es comprensible si, como antes se ha expuesto, la IA fuerte, avanzada o profunda trata de emular al cerebro humano cuyos razonamientos también son incomprensibles.

Sin embargo, en el ámbito en el que nos encontramos, si la IA se alza como decisora o como herramienta de apoyo a la decisión médica, si el humano no es capaz de comprender la lógica que el algoritmo ha seguido en la decisión, no podrá validar o descartar una propuesta diagnóstica o terapéutica de la máquina, siendo que esta puede ser errónea, sesgada, inequitativa o discriminatoria.

Un ejemplo de discriminación algorítmica lo constituye COMPASS,⁴⁰ un software usado por los Juzgados de Estados Unidos para establecer la probabilidad de reincidencia de un condenado con el objeto de determinar su condena. Después de su uso durante más de veinte años, un recurso en la Corte Suprema motivó el análisis del algoritmo y el resultado fue que la proporción de falsos positivos en la población de raza negra duplicaba a los asignados a la raza blanca, mientras que en los falsos negativos la propor-

³⁹ MAGDALENA LAYOS, L. «¿Por qué debería confiar en ti (máquina)?», en LLEDÓ YAGÜE, F.; BENÍTEZ ORTÚZAR, I.; MONJE BALMASEDA, O. (dirs.). *La robótica y la inteligencia artificial en la nueva era de la revolución industrial 4.0*. Dykinson, Madrid, 2021, pág. 633.

⁴⁰ Sobre COMPASS véase MORALES MORENO, A. M. «Algoritmos en el estrado, ¿realmente los aceptamos? Percepciones del uso de la inteligencia artificial en la toma de decisiones jurídico-penales». *IUS ET SCIENTIA*, vol. 7, núm. 2 (2021); ROMEO CASABONA, C. M. «Riesgo, procedimientos actuariales basados en inteligencia artificial y medidas de seguridad». *REDS*, núm. 13, julio-diciembre (2018); BORGES BLÁZQUEZ, R. «El sesgo de la máquina...», *op. cit.*

ción se invertía. Ello significa que si bien el algoritmo, en conjunto, acertaba, al analizar los resultados en función de la raza los errores se cometían en perjuicio de la raza negra, si atendíamos a la reincidencia real posteriormente registrada. ¿A qué o a quién se puede imputar este resultado discriminatorio? La respuesta a esta pregunta se desconoce porque COMPASS es un algoritmo de caja negra, es decir, opaco. Conocemos que entre los datos de entrada no figuraba la raza, si bien, el algoritmo era capaz, de algún modo, de extraerla bien a partir de los datos de entrenamiento o bien a partir de los inducidos por el propio algoritmo.

En la materia que nos ocupa, poco podría hacer el profesional salvo aplicar ciegamente el resultado proporcionado por la IA si esta es de caja negra. ROMEO CASABONA y LAZCOZ MORATINOS⁴¹ nos brindan un ejemplo a través de un ensayo en el que se pretendía determinar si la IA podía utilizarse para predecir el riesgo de muerte tras una neumonía. El resultado fue sorprendente, ya que el algoritmo consideró que el asma era un indicador de bajo riesgo, es decir, no influía en la predicción. A todas luces, esta decisión no es correcta puesto que el asma es una patología que afecta a los pulmones. En este caso, el modelo algorítmico era sencillo e inteligible por lo que se pudo apreciar cuál era el motivo de esta decisión algorítmica, que no era otra que el hecho de que los pacientes asmáticos cuando presentaban una neumonía eran ingresados en cuidados intensivos, y, por lo tanto, sometidos de modo temprano a una vigilancia y monitorización estricta que reducía el riesgo de fallecimiento respecto a lo que sucedía en el resto de la población. Si este algoritmo hubiera sido de caja negra y el profesional se hubiera limitado a aplicarlo las decisiones adoptadas hubieran sido dramáticas. De ahí, que en general y especialmente en determinados ámbitos la transparencia algorítmica sea una necesidad.

3. LA RESPUESTA DE LA LEY DE IA

La Ley de IA debe poner los medios para que los sistemas de IA no creen ni reproduzcan sesgos.

Con todo, SIMÓ SOLER y ROSSO⁴² también se preguntan —con acierto— si en la toma de decisiones «analógicas» se comprenden con exactitud todas las decisiones, es decir, si se desvela cuál ha sido el proceso cognitivo

⁴¹ ROMEO CASABONA, C.; LAZCOZ MORATINOS, G. *Inteligencia artificial aplicada a la salud: ¿Qué marco jurídico?* Fundación Merck Salud, Madrid, 2020, pág. 81.

⁴² SIMÓ SOLER, E.; ROSSO, P. «Inteligencia artificial y derecho: entre el mito y la realidad». *La Ley*, núm. 13447 (2021).

por el cual se opta por una decisión y no por otra teniendo en cuenta que en el proceso de toma de decisiones humanas intervienen sesgos cognitivos.

BONET NAVARRO⁴³ señala que, si bien se alerta respecto al peligro de la caja negra habría que interrogarse sobre los caminos —a veces, atajos— mentales que se emplean para adoptar una decisión —el autor se refiere, en concreto, a la judicial—. Para el autor no resulta adecuado negar el potencial de la IA por el mero hecho de no ser —ni llegará a ser nunca— un sistema perfecto, ya que siempre podrá ser sesgada. El autor pone de manifiesto la poca exigencia que mostramos con las decisiones humanas, al admitir la imperfección como algo natural, por lo que como mínimo la IA merecería ser tratada con los mismos parámetros con los que se trata al humano.

Por ello y en esta línea, la IA, si se entrena con datos de calidad y está diseñada con pleno respeto a los derechos fundamentales, debería contribuir a reducir los sesgos y la discriminación adoptando decisiones más equitativas. En la materia apuntada —las decisiones judiciales— el Considerando 61 deja constancia de que con el objeto de evitar posibles sesgos, errores y opacidades, procede la consideración como de alto riesgo de aquellos sistemas de IA cuyo objetivo es ayudar a las autoridades judiciales a investigar e interpretar los hechos y el Derecho, así como a aplicar la Ley a hechos concretos.

En atención a ello se establecen requisitos de obligatorio cumplimiento para los sistemas de IA de alto riesgo que se materializan en la exigencia de ser técnicamente sólidos para que los falsos resultados —bien sean positivos o negativos— no afecten desproporcionadamente a grupos protegidos.

Para ello estos sistemas deberán ser entrenados con datos suficientemente representativos en aras a reducir al mínimo los sesgos y dotar herramientas para detectarlos y corregirlos. Además, se deberá permitir el rastreo y auditorías, para lo que será necesario conservar la documentación que acredite qué datos se han utilizado para entrenar al algoritmo.

De este modo, el artículo 10 de la Ley de IA establece que los sistemas de IA de alto riesgo que utilicen técnicas que impliquen entrenamiento de modelos con datos deberán desarrollarse a partir de datos de entrenamiento, validación y prueba que cumplan con estrictos criterios de calidad que la propia norma recoge con el objeto de atender a los posibles sesgos.

⁴³ Se refieren a dicho autor SIMÓ SOLER y ROSSO en BONET NAVARRO, J. «¿Puede ser más justa, en materia penal, la inteligencia artificial?». *Revista Inmexius, Revista de Derecho Penal y Procesal Penal*, núm. 40 (2020), pág. 20.

Además, en su artículo 15.4 se recogen previsiones para los sistemas de IA de alto riesgo que continúan aprendiendo tras haber sido introducidos en el mercado o ponerse en servicio a los efectos de la subsanación de posibles sesgos en la información de salida que se deban a datos de entrada en futuras operaciones —bucle de retroalimentación—.

Para paliar estos riesgos resulta obligatorio la realización de pruebas *ex ante* con el objeto de reducir al mínimo la toma de decisiones asistidas por IA erróneas o sesgadas, y más en un ámbito como el de la salud, ya que no basta con contar con sistemas comprensibles, predecibles y controlables *ex post*. Desde el mismo diseño algorítmico y durante todo el proceso —con especial atención a las actualizaciones— resulta necesario corregir las vulnerabilidades.⁴⁴ Ello debe implicar la realización de constantes revisiones y evaluaciones con el objeto de minimizar los efectos negativos que esta tecnología, con capacidad de análisis y de predicción excepcional, supone.

Además, las obligaciones relativas a la realización de pruebas *ex ante*, la gestión de riesgos y la vigilancia humana facilitarán el respeto de otros derechos fundamentales, ya que contribuirán a reducir al mínimo el riesgo de adoptar decisiones asistidas por IA erróneas o sesgadas en esferas críticas como la educación y la formación, el empleo, servicios importantes, la aplicación de la ley y el poder judicial. En caso de que se sigan produciendo violaciones de los derechos fundamentales, la transparencia y la trazabilidad garantizadas de los sistemas de IA, unidas a unos controles *ex post* sólidos, permitirán ofrecer a las personas afectadas una compensación efectiva.

La transparencia en los sistemas de IA es fundamental y es necesario avanzar en su fiabilidad de modo que el interesado pueda obtener la explicación con un lenguaje accesible y comprensible que alcance a cómo actúa la IA en el concreto contexto.

V. DATOS PERSONALES E IA

BLÁZQUEZ RUIZ señala que históricamente los datos médicos se presentaban de modo fraccionado, descoordinado e incluso incompleto. A partir de técnicas como el reconocimiento óptico de caracteres, el dictado y el escaneo automatizado basado en el procesamiento del lenguaje natural, se

⁴⁴ En este sentido, MARTÍN JIMÉNEZ, F. J. «Inteligencia artificial y ética: hacia una aplicación de los principios éticos en el ámbito de la UE». *Cuadernos Europeos Deusto*, núm. 68 (2023), pág. 109. Cita el autor a Colcelli y Burzagli.

hace posible que la IA unifique bases de datos que abarcan incluso los datos no estructurados⁴⁵ como las notas clínicas.⁴⁶

La protección de datos personales es un derecho fundamental, que se recoge en el artículo 16 del Tratado de Funcionamiento de la UE,⁴⁷ en el artículo 8 de la Carta de los Derechos Fundamentales de la UE⁴⁸ y en el artículo 18 de la CE. Este derecho ha sido desarrollado por el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo⁴⁹ y precisado en nuestro ordenamiento interno por la Ley Orgánica 3/2018 de Protección de Datos y de Garantía de los Derechos Digitales.⁵⁰

1. LA TOMA DE DECISIONES AUTOMATIZADAS Y EL PERFILADO DE PACIENTES: UN RIESGO

Esta tecnología tiende a la generalización de la toma de decisiones automatizadas, y, para ello, uno de los métodos utilizados es el perfilado de pacientes⁵¹ por su gran capacidad predictiva y por la ayuda que puede suponer

⁴⁵ Un dato estructurado es un archivo de tipo texto constituido por información ordenada y organizada, en filas y columnas perfectamente identificadas con títulos, tal y como encontramos en la mayoría de Bases de Datos (BD) convencionales, como, por ejemplo: nombre, apellidos, DNI, fecha de nacimiento, edad, sexo..., datos que pueden ser almacenados, consultados y modificados, en forma de tabla y que además pueden ser procesados fácilmente.

Un dato no estructurado, es todo lo contrario a lo anterior, son datos cuyo manejo está más limitado. Por lo general son datos binarios sin estructura interna, que no aportan utilidad hasta ser identificados y ordenados de forma correcta. Una vez ordenados pueden ser categorizados para realizar búsquedas con el fin de obtener información.

MENÉNDEZ GARCÍA, P.; MOLINA CUEVAS, I. [*et al.*]. «La transformación digital del sistema sanitario: la revolución del bigdata, todo por mejorar la calidad asistencial» [<https://revistamedica.com/transformacion-digital-sistema-sanitario-bigdata-calidad-asistencial/#Datos-estructurados-y-no-estructurados>].

⁴⁶ BLÁZQUEZ RUIZ, F. J. «Riesgos para la privacidad en la aplicación de la inteligencia artificial al ámbito biosanitario. Implicaciones éticas y legales». *Anales de la Cátedra Francisco Suárez*, núm. 56 (2022), pág. 250.

⁴⁷ Versión consolidada del Tratado de Funcionamiento de la Unión Europea, DOUE C 326/47, 26.10.2012 [<https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:12012E/TXT>].

⁴⁸ Carta de los Derechos Fundamentales de la Unión Europea (2010/C 83/02), DOUE C 83/389. 30.3.2010 [<https://www.boe.es/doue/2010/083/Z00389-00403.pdf>].

⁴⁹ Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos), DOUE L 119/1, 4.5.2016 [<https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32016R0679>].

⁵⁰ Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales (BOE núm. 294, de 6.12.2018).

⁵¹ Art. 4.4 RGPD: «“elaboración de perfiles”: toda forma de tratamiento automatizado de datos personales consistente en utilizar datos personales para evaluar determinados aspectos personales

para la mejora de la planificación, de la toma de decisiones, de la elaboración de protocolos clínicos y de un sinnúmero de aplicaciones beneficiosas. No obstante, el perfilado puede implicar también una grave intrusión en la vida de la persona, privarle de determinadas terapias o implicar restricciones a la autonomía tanto del paciente como del profesional. Así podría suceder que al paciente encasillado en un determinado estereotipo se le privara de un determinado tratamiento por razones de eficacia y de eficiencia del servicio sanitario sin tener en cuenta las variables personalizadas y en contra de la medicina del futuro, que, como con anterioridad se ha apuntado, es la personalizada.

La limitación a la adopción de este tipo de decisiones basadas en perfiles se establece en el artículo 22 del RGPD, que contempla el derecho del interesado a no ser objeto de decisiones basadas únicamente en el tratamiento automatizado, incluyendo la elaboración de perfiles, que produzca efectos jurídicos para la persona o le afecte significativamente. No cabe duda de que las decisiones sobre su salud afectan de modo significativo a la persona —administrarle o no un determinado tratamiento o someterle o no a un determinado procedimiento diagnóstico o intervención quirúrgica—.

Al margen de esta disposición general, en el punto 2 del mismo artículo se recogen una serie de excepciones entre las que se encuentra el consentimiento explícito del interesado y, en el punto 3, una serie de salvaguardas para los casos en que esta excepción se aplique. Estas medidas, en beneficio de los derechos, de las libertades y de los intereses legítimos del interesado implicarían como mínimo el derecho a obtener una intervención humana del profesional y el derecho del paciente a expresar su punto de vista y a impugnar la decisión.

SARASÍBAR IRIARTE señala que «no se trata de sustituir el criterio del juez, o también aplicable por ejemplo a un médico, por un programa, sino de ampliar y reforzar su capacidad de diagnóstico, de que constituya un instrumento de apoyo y de mejora, en todo caso».⁵² En relación con lo anterior, en estos casos, la intervención humana significativa resulta fundamental.

de una persona física, en particular para analizar o predecir aspectos relativos al rendimiento profesional, situación económica, salud, preferencias personales, intereses, fiabilidad, comportamiento, ubicación o movimientos de dicha persona física».

⁵² SARASÍBAR IRIARTE, M. «La cuarta revolución industrial: el derecho administrativo ante la inteligencia artificial». *Revista Vasca de Administración Pública*, núm. 115 (2019), pág. 389.

2. LA ANONIMIZACIÓN DE DATOS PERSONALES: UNA ILUSIÓN

IA y su motor —*big data*— implican tratamientos masivos de datos procedentes de las más diversas fuentes.

En aras a proteger la intimidad, el uso de información no personal es el modelo más tuitivo de los derechos de la persona. Sin embargo, en el ámbito médico poco aportan los datos que no pueden vincularse de ningún modo a persona física identificada o identificable.⁵³ Por si este obstáculo no fuera suficiente, además, el proceso de anonimización que durante un tiempo ha servido para tratar información al margen de la normativa de protección de datos, es insuficiente actualmente para que la privacidad del interesado resulte, sin lugar a dudas, protegida. El despliegue de tecnologías *big data* ha supuesto la máxima de que hoy ya no existen datos anonimizados. De ahí que esto sea ya una de las debilidades de la joven normativa en materia de protección de datos. Si bien, durante un tiempo, la anonimización, incluso la seudonimización⁵⁴ —en este caso se sigue tratando información personal— fueron la panacea para excluir o flexibilizar el rigor de las normas tuitivas de los datos de la persona, hoy la reidentificación es posible en cualquier caso si atendemos al desarrollo de la tecnología *big data* que hace posible la identificación gracias a las ingentes cantidades de datos que maneja, por lo que la reidentificación de la persona, una vez desvinculada la información, cada vez es más sencilla.

Además, determinada información que, en principio no es de salud, a partir del cruce de datos procedentes de otras fuentes puede darnos información de salud. Por ejemplo, pensemos en una persona cuyo domicilio está ubicado en un área con altos niveles de contaminación y padece una enfermedad pulmonar susceptible de agravarse por el citado factor. En este caso el dato del domicilio, pese a no ser en principio de salud, se convierte en un dato sensible de salud puesto que nos proporciona información a ella relativa. De modo similar, la lista de la compra puede proporcionar una

⁵³ Art. 4.1 RGPD: «“datos personales”: toda información sobre una persona física identificada o identificable («el interesado»); se considerará persona física identificable toda persona cuya identidad pueda determinarse, directa o indirectamente, en particular mediante un identificador, como por ejemplo un nombre, un número de identificación, datos de localización, un identificador en línea o uno o varios elementos propios de la identidad física, fisiológica, genética, psíquica, económica, cultural o social de dicha persona».

⁵⁴ Art. 4.5 RGPD: «“seudonimización”: el tratamiento de datos personales de manera tal que ya no puedan atribuirse a un interesado sin utilizar información adicional, siempre que dicha información adicional figure por separado y esté sujeta a medidas técnicas y organizativas destinadas a garantizar que los datos personales no se atribuyan a una persona física identificada o identificable».

información valiosa sobre el estado de salud, de modo que unos datos que de modo aislado no son de salud, lo serán al ponerlos en relación con otra información.⁵⁵

Por ello, sería necesario un nuevo régimen de protección de datos en el que se prescindiera del término personales. Hoy cualquier dato puede ser personal y ese matiz reviste un gran alcance en el ámbito sanitario en el que, o bien, todos los datos de salud serán personales, o bien, indirectamente, sin tratar información personal, se podrán deducir datos de salud, que, además, revisten una sensibilidad superior a cualquier otro, por lo que su mal uso, repercutiría, sin duda, en un ataque a los derechos fundamentales de la persona. El riesgo y la posible vulneración de derechos se incrementa si atendemos al uso secundario⁵⁶ de datos, o lo que es lo mismo, la utilización de los datos para fines distintos a los previstos inicialmente.

Además, la extraordinaria capacidad de cruzar datos puede dar lugar a correlaciones con resultados inesperados para el propio paciente, que, si bien inicialmente no eran parte de lo esperable, puede ponerse de manifiesto información que ni siquiera el propio paciente conoce o quiera conocer. Ello implicará cautelas especiales en el proceso de consentimiento informado previo y unas medidas técnicas y organizativas muy estrictas y restrictivas para evitar los hipotéticos daños a los derechos de la persona que se ocasionarían si los datos caen en manos de terceros para su uso ilegítimo.

3. EL ESPACIO EUROPEO DE DATOS SANITARIOS

La Directiva (UE) 2019/1024 relativa a los datos abiertos y reutilización de la información del sector público⁵⁷ impulsa a los Estados miembros a promover datos abiertos bajo el cumplimiento del principio de datos abiertos desde el diseño y por defecto con respeto al principio de transparencia y el control de datos abiertos a través de un marco jurídico que respete los datos personales. Fruto de la propuesta de 2020 de Reglamento del Parlamento Europeo y del Consejo sobre gobernanza de los datos en el ámbito

⁵⁵ GIL MEMBRADO, C. *Riesgos del uso del algoritmo...*, *op. cit.*

⁵⁶ Sobre el uso de datos secundarios véase DE MONTALVO JÄÄSKELÄINEN, F. «El uso secundario de los datos de salud en el marco del desarrollo de la e-Health ¿es necesario el cambio del paradigma asentado en el consentimiento informado?». *E-salud, autonomía y datos clínicos. Un nuevo paradigma*. Dykinson, Madrid, 2021.

⁵⁷ Directiva (UE) 2019/1024 del Parlamento Europeo y del Consejo de 20 de junio de 2019 relativa a los datos abiertos y la reutilización de la información del sector público (versión refundida), DOUE L 172/56, 26.6.2019 [<https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32019L1024>].

europeo contamos con el Reglamento (UE) 2022/868 de Gobernanza de Datos⁵⁸ para lograr una mayor disponibilidad de datos e impulsar su reutilización.

En el Considerando 45 de la Ley de IA se deja constancia de que, para el desarrollo de sistemas de IA de alto riesgo es imprescindible que los intervinientes cuenten con conjuntos de datos de alta calidad. Para ello, los espacios comunes europeos de datos y el intercambio regulado de información son esenciales para «brindar un acceso fiable, responsable y no discriminatorio a datos de alta calidad con los que entrenar, validar y probar los sistemas de IA», precisamente para minimizar los sesgos a los que anteriormente nos hemos referido.

En este ámbito, el de la salud, resulta fundamental la Propuesta de Reglamento para crear el espacio europeo de datos sanitarios⁵⁹ que tiene como finalidad liberar el potencial que reviste el manejo de datos sanitarios y resultará de ayuda para que la persona tome el control de los datos sanitarios propios; para la mejora de la prestación de asistencia sanitaria, de la investigación, de la innovación y para elaborar políticas sanitarias. Este espacio permitirá a la UE aprovechar de modo pleno el potencial del intercambio, el uso y la reutilización segura y protegida de los datos sanitarios.

El Espacio Europeo de datos sanitarios se alza como un ecosistema previsto para la promoción de la salud que coadyuvará al horizonte apuntado en el artículo 45 de la Ley de IA, de modo que facilitará el acceso no discriminatorio a datos sanitarios y el entrenamiento, a partir de esos conjuntos de datos, de algoritmos de IA de una manera segura, oportuna, transparente y fiable que respete la privacidad, y contando con la debida gobernanza institucional.

Para ello, el Espacio Europeo de Datos Sanitarios contendrá reglas, normas y prácticas comunes, infraestructuras y un marco de gobernanza con el objetivo de que la persona tenga un mayor control y acceso digital a sus datos sanitarios electrónicos, así como a promover la libre circulación de los historiales médicos electrónicos para su uso tanto primario como se-

⁵⁸ Reglamento (UE) 2022/868 del Parlamento Europeo y del Consejo de 30 de mayo de 2022 relativo a la gobernanza europea de datos y por el que se modifica el Reglamento (UE) 2018/1724 (Reglamento de Gobernanza de Datos), DOUE L 152/1, 3.6.2022 [<https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32022R0868>].

⁵⁹ Propuesta de Reglamento del Parlamento Europeo y del Consejo sobre el Espacio Europeo de Datos Sanitarios; COM/2022/197 final [<https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52022PC0197>].

cundarios tales como la investigación, la innovación, y la formulación de políticas sanitarias.

4. LA RESPUESTA DE LA LEY DE IA

En estrecha relación con la minimización de sesgos en los datos de salida de la IA, el artículo 10.5 de la Ley de IA establece que si fuera estrictamente necesario para la garantía de la vigilancia, la detección y la corrección de sesgos en la IA de alto riesgo, los proveedores de los sistemas podrán tratar categorías especiales de datos personales mencionados en el artículo 9.1 del RGPD —datos personales que revelen el origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas, o la afiliación sindical, y el tratamiento de datos genéticos, datos biométricos dirigidos a identificar de manera unívoca a una persona física, datos relativos a la salud o datos relativos a la vida sexual o la orientación sexual de una persona física—. Para ello tendrán que establecerse las salvaguardas adecuadas para los derechos y las libertades fundamentales de la persona, lo que incluye el establecimiento de limitaciones técnicas a la reutilización y la utilización de medidas de seguridad y protección de la privacidad tales como la seudonimización o el cifrado, cuando no pueda hacerse uso de la anonimización.

Para que la privacidad sea efectiva resulta necesario que la misma se considere desde el mismo momento en que los algoritmos que sustentan la IA se diseñan. Ello está en estrecha relación con lo dispuesto por el artículo 25 del RGPD, que se refiere a la protección de datos desde el diseño y por defecto. Ello implica que, conforme al estado de la técnica y teniendo en cuenta el coste, la naturaleza, el ámbito, el contexto y los fines del tratamiento, así como los riesgos que entrañe el tratamiento de datos para los derechos y las libertades de las personas físicas, el responsable del tratamiento⁶⁰ debe aplicar tanto en el momento de determinar los medios del tratamiento como durante el mismo las medidas técnicas y organizativas apropiadas. Se refiere la propia norma a la seudonimización, aunque ya hemos apuntado la fragilidad de este método para proteger el derecho a la privacidad de la persona. Sin duda, el estado de la técnica ha venido a superar los métodos que en el tiempo de aprobación del RGPD eran pertinentes y

⁶⁰ «[...] la persona física o jurídica, autoridad pública, servicio u otro organismo que, solo o junto con otros, determine los fines y medios del tratamiento; si el Derecho de la Unión o de los Estados miembros determina los fines y medios del tratamiento, el responsable del tratamiento o los criterios específicos para su nombramiento podrá establecerlos el Derecho de la Unión o de los Estados miembros» (art. 4.7 RGPD).

adecuados, y que ahora quedan estrechos si atendemos a las posibilidades que el *big data* ofrece con relación al cruce de bases de datos.

Además, en el citado precepto se establecen especiales cautelas en lo que se refiere al tratamiento de los datos personales que sean necesarios para los fines específicos del tratamiento. Esta cautela también queda pequeña si atendemos a las posibilidades de tratamiento de datos con fines secundarios, cuyo paradigma, en el escenario donde nos movemos, son los fines de investigación biomédica. Las medidas tendentes a proteger la intimidad del individuo, como la minimización de datos⁶¹ y la protección de datos por defecto⁶² pueden limitar la potencialidad de estas tecnologías para mejorar los resultados que repercutirán en la salud de la humanidad.

Con relación a la intervención humana, la Ley de IA es constante en incidir en la vigilancia humana que debe acompañar a todo el proceso decisonal de la máquina. En su artículo 14 establece que los sistemas de IA de alto riesgo deben diseñarse y desarrollarse para que puedan ser vigilados de modo efectivo por personas físicas, de modo que el objetivo de la vigilancia humana será, entre otros, prevenir o reducir al mínimo los riesgos para la salud. Para ello se establecen como medidas, en función de las circunstancias, el entendimiento completo de las capacidades y de las limitaciones del sistema de IA de alto riesgo y el control de su funcionamiento. Además, la norma llama la atención sobre la necesidad de ser conscientes de la posible tendencia a confiar de modo automático o en exceso en la información de salida generada por la IA, por lo que a todos los señalados, se añade un sesgo más, el de automatización, en particular respecto de aquellos sistemas que se utilicen para aportar información o recomendaciones para que la persona física —en el caso que nos ocupa, el profesional médico— adopte una decisión.

Esta intervención humana debe ser significativa,⁶³ lo cual indica que un humano autorizado y competente deberá ser capaz, a lo largo de todo el proceso, de modificar las decisiones de la máquina. Así, en una decisión diagnóstica adoptada por la máquina constituirá un apoyo a la decisión médica, que el profesional deberá en última instancia aceptar o rechazar. La

⁶¹ Entre los principios aplicables al tratamiento de datos personales el de minimización incluye el tratamiento únicamente de aquellos datos «adecuados, pertinentes y limitados a lo necesario en relación con los fines para los que son tratados» (art. 5.1.c RGPD).

⁶² La protección de datos personales desde el diseño y por defecto se recoge en el artículo 25 RGPD.

⁶³ 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)) (2021/C 404/04), DOUE C 404/63, 6.10.2021.

garantía de que esta intervención no es meramente simbólica vendrá dada por la evaluación de impacto⁶⁴ que de conformidad al RGPD⁶⁵ el responsable tiene que realizar cuando el tratamiento de datos entrañe un alto riesgo para los derechos y libertades de la persona.

Para que la intervención humana del profesional sea significativa, y también para que el paciente sea capaz de expresar su punto de vista y de oponerse a la decisión es necesario —y en este punto retomamos la opacidad y la explicabilidad de la lógica decisional del algoritmo— no solo saber cómo funciona el sistema de IA, que en palabras de MAGDALENA LAYOS⁶⁶ sería la interpretabilidad o una aproximación a la explicación del algoritmo *ex ante*, es decir, antes de tomar la decisión. La explicabilidad del algoritmo es algo bien distinto, que implica el análisis *ex post* de la concreta información de salida del sistema de IA.

El gran inconveniente, todavía no resuelto, es que la propia normativa que pretende velar por la transparencia algorítmica, en realidad lo que permite es el acceso a la lógica aplicada o a la interpretabilidad a la que nos referíamos. De este modo, si atendemos al RGPD, en su artículo 15.1.h, en relación con el derecho de acceso del interesado, manifiesta que abarca «la existencia de decisiones automatizadas, incluida la elaboración de perfiles» y la «información significativa sobre la lógica aplicada».

Si la normativa sobre protección de datos no da una respuesta a la necesaria transparencia que debe materializarse en el derecho a la explicación del algoritmo, tampoco este se satisface a través del derecho del paciente a la información clínica previsto en el artículo 4 de la Ley 41/2002, en cuanto recoge el derecho del paciente a conocer, con motivo de cualquier actuación sobre su salud, toda la información disponible. En los casos en los que el algoritmo es de caja negra, esta información no resulta disponible, por lo que ni el profesional ni el paciente podrán acceder a ella.

La tensión entre el avance de la máquina y la salvaguarda de los derechos de la persona se hace patente, ya que la exigencia de transparencia sin ambages podría frenar el progreso tecnológico. Sin embargo, en ámbitos como el sanitario, no resulta concebible una IA opaca salvo que se abandone como principio ético capital la IA antropocéntrica y antropogénica, lo que implicaría colocar a la máquina al mismo nivel que el humano.

⁶⁴ PALMA ORTIGOSA, A. «Decisiones automatizadas en el RGPD. El uso de algoritmos en el contexto de la protección de datos». *Revista General de Derecho Administrativo*, núm. 50 (2019).

⁶⁵ Art. 35 RGPD.

⁶⁶ MAGDALENA LAYOS, L. «¿Por qué debería confiar en ti (máquina)?», cit., pág. 636.

VI. BIBLIOGRAFÍA

- AÑÓN ROIG, M. J. «Desigualdades algorítmicas: conductas de alto riesgo para los derechos humanos». *Derechos y libertades*, núm. 47, Época II, junio (2022).
- BARRIO ANDRÉS, M. «Luces y sombras del Estado algorítmico de Derecho». *Derecho Digital e Innovación*, núm. 5 (segundo trimestre 2020), La Ley 5962/2020.
- BENAVENT NÚÑEZ, D.; COLOMER MASCARÓ, J. [et al.]. *Inteligencia Artificial y decisiones clínicas: cómo está cambiando el comportamiento médico*. Fundación Gaspar Casal, Madrid, 2020.
- BLÁZQUEZ RUIZ, F. J. «Riesgos para la privacidad en la aplicación de la inteligencia artificial al ámbito biosanitario. Implicaciones éticas y legales». *Anales de la Cátedra Francisco Suárez*, núm. 56 (2022).
- BONET NAVARRO, J. «¿Puede ser más justa, en materia penal, la inteligencia artificial?». *Revista Inmexius, Revista de Derecho Penal y Procesal Penal*, núm. 40 (2020).
- BORGES BLÁZQUEZ, R. «El sesgo de la máquina en la toma de decisiones en el proceso penal». *Ius et Scientia*, vol. 6, núm. 2 (2020).
- COTINO HUESO, L. «Inteligencia artificial, big data y aplicaciones contra la Covid-19: privacidad y protección de datos». *IDP: Revista de Internet, Derecho y Política*, núm. 31 (2020).
- «Transparencia y explicabilidad de la inteligencia artificial y compañía (comunicación, interpretabilidad, inteligibilidad, auditabilidad, testabilidad, comprobabilidad, simulabilidad...). Para qué, para quién y cuánta», en COTINO HUESO, L.; CASTELLANOS J. (coords.). *Transparencia y explicabilidad de la inteligencia artificial*. Tirant lo Blanch, Valencia, 2022.
- DE LA NUEZ SÁNCHEZ-CASCADO, E. «Inteligencia artificial y transparencia». *La Ley Derecho Digital e Innovación*, 6786, núm. 5 (2020).
- DE MONTALVO JÄÄSKELÄINEN, F. «El uso secundario de los datos de salud en el marco del desarrollo de la e-Health ¿es necesario el cambio del paradigma asentado en el consentimiento informado?». *E-salud, autonomía y datos clínicos. Un nuevo paradigma*, Dykinson, Madrid, 2021.
- GAMERO CASADO, E. «El enfoque europeo de inteligencia artificial». *Revista de Derecho Administrativo*, CDA 20 (2021).
- GIL MEMBRADO, C. *Riesgos del uso de algoritmos en el diagnóstico y en la investigación biomédica*. VIII Premio Nacional de Derecho Sanitario. Reus, Tarragona, 2023.
- «La apertura del código fuente y la transparencia en la contratación de servicios», en COBAS COBIELLA, M. E.; GUILLÉN CATALÁN, R. (dirs.). *Equidad y transparencia en la prestación de servicios*. Dykinson, Madrid, 2023.
 - «Daños producidos por la IA: la opacidad del algoritmo y el efecto de caja negra», en ÁLVAREZ LATA, N. (coord.). *Derecho de contratos, responsabilidad extracontractual e inteligencia artificial*. Aranzadi, Madrid, 2024.

- «Diabetes y Espacio Europeo de Datos Sanitarios», en GIL MEMBRADO, C.; LUQUIN BERGARECHE, R. (dirs.). *Salud Digital. Aplicaciones móviles, telemedicina y chatbots*. Dykinson, Madrid, 2024.
- MAGDALENA LAYOS, L. «¿Por qué debería confiar en ti (máquina)?», en LLEDÓ YAGÜE, F.; BENÍTEZ ORTÚZAR, I.; MONJE BALMASEDA, O. (dirs.). *La robótica y la inteligencia artificial en la nueva era de la revolución industrial 4.0*. Dykinson, Madrid, 2021.
- MARTÍN JIMÉNEZ, F. J. «Inteligencia artificial y ética: hacia una aplicación de los principios éticos en el ámbito de la UE». *Cuadernos Europeos Deusto*, núm. 68 (2023).
- MENÉNDEZ GARCÍA, P.; MOLINA CUEVAS, I. [et al.]. «La transformación digital del sistema sanitario: la revolución del big data, todo por mejorar la calidad asistencial» [<https://revistamedica.com/transformacion-digital-sistema-sanitario-bigdata-calidad-asistencial/#Datos-estructurados-y-no-estructurados>].
- MORALES MORENO, A. M. «Algoritmos en el estrado, ¿realmente los aceptamos? Percepciones del uso de la inteligencia artificial en la toma de decisiones jurídico-penales». *IUS ET SCIENTIA*, vol. 7, núm. 2 (2021).
- NAVAS NAVARRO, S. «Aspectos jurídicos de las aplicaciones móviles de salud (apps sanitarias y wearables)». *Diario La Ley* (14 de mayo de 2020), La Ley 5811/2020.
- PALMA ORTIGOSA, A. «Decisiones automatizadas en el RGPD. El uso de algoritmos en el contexto de la protección de datos». *Revista General de Derecho Administrativo*, núm. 50, (2019).
- PHUA, D.H.; TAN N.C. «Cognitive aspect of diagnostic errors». *Annals of de Academy of Medicine, Singapore*, núm. 1, vol. 42 (2013).
- ROMEO CASABONA, C. M. «Riesgo, procedimientos actuariales basados en inteligencia artificial y medidas de seguridad». *REDS*, núm. 13, julio-diciembre (2018).
- ROMEO CASABONA, C.; LAZCOZ MORATINOS, G. *Inteligencia artificial aplicada a la salud: ¿Qué marco jurídico?* Fundación Merck Salud, Madrid, 2020.
- SARASÍBAR IRIARTE, M. «La cuarta revolución industrial: el derecho administrativo ante la inteligencia artificial». *Revista Vasca de Administración Pública*, núm. 115 (2019).
- SIMÓ SOLER, E.; ROSSO, P. «Inteligencia artificial y derecho: entre el mito y la realidad». *La Ley*, núm. 13447 (2021).
- ZORRAQUINO A.; MATAS BRANCÓS, A.; VILANOVA JOU, R. «El Reglamento de IA, cada vez más cerca: acuerdo preliminar del Consejo y el Parlamento Europeo». *Newsletter de PwC Tax & Legal*, PwC, diciembre 2023 [https://periscopiofiscalylegal.pwc.es/wp-content/uploads/2023/12/El-Reglamento-de-IA-cada-vez-mas-cerca.-Acuerdo-preliminar-del-Consejo-y-el-Parlamento-Europeo-v_5.pdf].