Revista de Derecho y Genoma Humano, 61 – enero-diciembre 2025

Law and the Human Genome Review

e-ISSN 2605-1230

https://doi.org/10.1387/rdgh.28061

Informes institucionales y normativa

Sistemas de inteligencia artificial en la asistencia sanitaria: cómo garantizar la supervisión humana desde la normativa de protección de datos (v.2)

Guillermo Lazcoz Moratinos[1]

Centro de Investigación Biomédica en Red (CIBER*ER* - ISCIII)

Centro de Diagnóstico de Enfermedades Moleculares (CEDEM - UAM)

[1] Este trabajo es una versión actualizada del Premio de Investigación en Protección de Datos Personales “Emilio Aced” (2022) otorgado por la Agencia Española de Protección de Datos.
El texto no ha sido revisado por pares.
La primera versión fue realizada con financiación de la Convocatoria de becas y ayudas para la formación de doctores del programa nacional de formación de profesorado universitario (FPU) 2016, de acuerdo con las condiciones establecidas en Resolución de 22 de diciembre de 2016 (BOE de 17 de enero de 2017) -FPU16/06314-.
La segunda versión se ha realizado en el marco del proyecto GODAS: Gobernanza de los usos secundarios de datos de salud y genéticos en espacios compartidos (PID2022-137140OB-I00) de «PROYECTOS DE GENERACIÓN DE CONOCIMIENTO» en el marco del Programa Estatal para Impulsar la Investigación Científico-Técnica y su Transferencia, del Plan Estatal de Investigación Científica, Técnica y de Innovación 2021-2023.

Resumen: El objetivo de este trabajo es la realización de un análisis jurídico que permita evaluar la aptitud de la normativa de protección de datos (RGPD) para garantizar una supervisión humana durante todo el ciclo de vida de sistemas de IA que, a partir del tratamiento de datos personales, pueden utilizarse para fines de diagnóstico, prognosis o de elección terapéutica en la asistencia sanitaria. Para la consecución de este objetivo el análisis se complementa con un análisis de la normativa sectorial de productos sanitarios (MDR/IVDR) aplicable a dicho objeto de estudio, así como de la normativa de inteligencia artificial (RIA). Como se verá, el conjunto de obligaciones establecido desde el desarrollo del sistema hasta su implementación en el ámbito clínico, nos permiten establecer las bases normativas para entender cómo transitar desde la supervisión humana efectiva hasta la intervención humana significativa.

Palabras clave: Intervención humana, RGPD, supervisión humana, RIA, derecho sanitario

Abstract: The aim of this work is to carry out a legal analysis to assess the suitability of data protection law (GDPR) to provide human oversight throughout the life cycle of AI systems that, based on the processing of personal data, can be used for diagnostic, prognostic or therapeutic choice purposes in healthcare. To achieve this goal, the analysis is complemented by the analysis of the sectoral regulation of medical devices (MDR/IVDR), as well as the regulations on artificial intelligence (AIA). By examining the obligations that arise from the development of the AI system through to its clinical application, we can understand the regulatory foundations needed to move from effective human oversight to meaningful human intervention.

Keywords: Human intervention, GDPR, human oversight, AIA, health law

---

1. Inteligencia artificial en la asistencia sanitaria: una introducción

La irrupción en todos los ámbitos y sectores de nuestra sociedad de lo que venimos conociendo como tecnologías de inteligencia artificial (IA en adelante) puede ubicarse entre los fenómenos con mayor impacto y repercusión de nuestra actualidad. Si el devenir de esta irrupción terminará transformando nuestras vidas o en un letárgico nuevo invierno de la IA, está por ver. O por decidir, si asumimos que somos más sujetos activos que objetos pasivos en la evolución de este proceso socio-tecnológico.

La asistencia sanitaria es uno de los ámbitos de aplicación de estas tecnologías que más interés público y privado ha suscitado, tanto por su potencial para contribuir a la mejora de la salud y del bienestar general, como por los riesgos que su implementación conlleva para los derechos fundamentales.

La Comisión Europea ha otorgado un protagonismo especial al desarrollo de estas tecnologías en el ámbito de la salud en la Estrategia Europea de Datos[2]. Asimismo, como abordaremos con detalle más adelante, el Reglamento de Inteligencia Artificial (RIA o Reglamento de IA, en adelante) ha catalogado gran parte de los sistemas de IA en salud como de "alto riesgo"[3]. Un protagonismo similar vemos en el ámbito estatal. En la Estrategia Nacional de Inteligencia Artificial (ENIA), elaborada por el Grupo de Trabajo Interministerial en Inteligencia Artificial, coordinado por el Ministerio de Ciencia e Innovación[4], el ámbito de la salud se reconoce en repetidas ocasiones como uno de los sectores estratégicos a promover, colocándolo también como uno de los que mayor impacto pueden generar a corto/medio plazo. En la Carta de Derechos Digitales adoptada por el Gobierno de España[5], que carece de valor jurídico, el derecho a la protección de la salud en el entorno digital se reconoce como un entorno específico de protección y, además, se reconocieron dos derechos específicos en el ámbito de aplicación de los sistemas algorítmicos[6].

Hay quien asegura que prácticamente toda clase de profesional sanitario utilizará sistemas de IA en el futuro -en su desempeño laboral-[7]. En este sentido, el *software* 'tradicional' utilizado como sistemas de apoyo a la toma de decisiones clínicas (CDSS por sus siglas en inglés) en las últimas décadas, se irá sustituyendo progresivamente por tecnologías de IA basadas en algoritmos de aprendizaje automático[8]. Otras predicciones, más escépticas, aseguran que en medicina no hay sustituto para el conocimiento y la intuición humanos[9], y que por tanto los sistemas de IA no sustituirán a los profesionales sanitarios en el cuidado de los pacientes[10]. En realidad, estas predicciones no son incompatibles entre sí, puede que se extienda el uso de

---

[2] Así puede constatarse en los documentos publicados por la Comisión Europea en febrero de 2020. Vid. Comisión Europea, Comunicación de la Comisión «*Libro Blanco sobre la inteligencia artificial – un enfoque europeo orientado a la excelencia y la confianza*». Bruselas, 19.02.2020. COM (2020) 65 final. Disponible en: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_es.pdf; Comisión Europea Comunicación de la Comisión «*Una Estrategia Europea de Datos*». Bruselas, 19.02.2020. COM (2020) 66 final. Disponible en: https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:52020DC0066.

[3] 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.

[4] Gobierno de España, *Estrategia Nacional de Inteligencia Artificial (ENIA)*, Versión 1.0. Disponible en: https://portal.mineco.gob.es/RecursosArticulo/mineco/ministerio/ficheros/201202_ENIA_V1_0.pdf

[5] Nota de prensa: Sánchez presenta la Carta de Derechos Digitales, con la que "España se sitúa a la vanguardia internacional en la protección de derechos de la ciudadanía". La Moncloa, Madrid, miércoles 14 de julio de 2021. Disponible aquí: https://www.lamoncloa.gob.es/presidente/actividades/Paginas/2021/140721-derechos-digitales.aspx

[6] Título XXIII. Derecho a la protección de la salud en el entorno digital: *(…) 4. El empleo de sistemas digitales de asistencia al diagnóstico, y en particular de procesos basados en inteligencia artificial no limitará el derecho al libre criterio clínico del personal sanitario. 5. Los entornos digitales de salud garantizarán, conforme a la legislación sectorial, la autonomía del paciente, la seguridad de la información, la transparencia sobre el uso de algoritmos, la accesibilidad y el pleno respeto de los derechos fundamentales del paciente y en particular su derecho a ser informado o renunciar a la información y a consentir en el tratamiento de sus datos personales con fines de investigación y en la cesión a terceros de tales datos cuando tal consentimiento sea requerido.* Gobierno de España, *Carta de Derechos Digitales*, Plan de recuperación, transformación y resiliencia, 2021. Texto completo disponible aquí: https://www.lamoncloa.gob.es/presidente/actividades/Documents/2021/140721-Carta_Derechos_Digitales_RedEs.pdf

[7] Topol, *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*, 44.

[8] Davenport y Kalakota, «The Potential for Artificial Intelligence in Healthcare», 95.

[9] Garnacho-Montero y Martín-Loeches, «Clinical management of sepsis can be improved by artificial intelligence: no», 378.

[10] Sniderman, D'Agostino Sr, y Pencina, «The Role of Physicians in the Era of Predictive Analytics», 25.

estas tecnologías sin necesidad de sustituir a las personas y sus conocimientos. Aunque también puede ocurrir que acabemos siendo atendidos por máquinas que sustituyen a humanos, lo cual ya vemos en múltiples sectores y cadenas productivas[11].

Saliendo del paradigma de la sustitución para centrar la mirada en el futuro, hemos podido observar a lo largo de las últimas décadas que la mejora de cuidado clínico no es simplemente el resultado de un incremento del uso de tecnologías más complejas[12]. Además, la evidencia sugiere que la automatización tecnológica transforma la actividad humana más que sustituirla[13].

Como parte de este proceso de transformación, la tecnología, y en particular la IA, se ha propuesto a menudo como una solución a la incertidumbre que rodea muchos aspectos de la práctica sanitaria: cómo clasificar las condiciones de los pacientes (incertidumbre diagnóstica); por qué y cómo los pacientes desarrollan enfermedades (incertidumbre pato-fisiológica); qué tratamiento será más apropiado para ellos (incertidumbre terapéutica); si se recuperarán con o sin un tratamiento específico (incertidumbre pronóstica), y así sucesivamente[14]. En este sentido, esta investigación pone su atención en los sistemas de IA que, a partir del tratamiento de datos personales, arrojan inferencias sobre el (incierto) estado de salud del paciente.

El cuerpo humano es una pieza compleja de la maquinaria biológica que en la práctica clínica requiere recopilar, analizar y poner en valor información que proviene de una amplia gama de fuentes[15], tarea monumental incluso para los y las profesionales con más experiencia; aquí, la implementación de estos modelos algorítmicos se asocia a la promesa de llevar a cabo análisis más precisos y detallados a partir de los complejos datos sanitarios de sus pacientes -de datos personales, a fin de cuentas-, que faciliten la labor del personal sanitario.

Como veremos a continuación, las investigaciones realizadas hasta el momento nos invitan a pensar con optimismo en nuevas formas de acercarnos al cuidado clínico y de reorganizar los recursos sanitarios, reduciendo ese conjunto de incertidumbres que rodean a la práctica clínica. Ahora bien, en este acercamiento debemos ser conscientes de que el desarrollo de estas tecnologías está aún en un periodo de infancia[16].

No solo eso, Prainsack recalca que en la transformación estructural y sin precedentes que la digitalización ha traído al cuidado de la salud en los últimos 15 años, no podemos olvidar la intromisión de nuevos actores en este ámbito: las empresas tecnológicas que recolectan los datos para la monitorización cada vez más ubicua del conjunto de la sociedad -y más allá del contexto médico-paciente- y que desarrollan a su vez el *software* para extraer correlaciones y modular la toma de decisiones a partir del análisis de dichos datos[17].

Así las cosas, y admitiendo que las tecnologías de IA tienen el potencial de alterar las relaciones asistenciales y desplazar las responsabilidades que tradicionalmente desempeña el personal sanitario[18]; por el momento, no son los algoritmos los que proporcionan una asistencia sanitaria segura y eficaz, sino las personas y las instituciones que forman el conjunto del sistema sanitario[19].

---

[11] La práctica de la medicina no es ajena a la automatización que está desarrollándose en prácticamente la totalidad de los sectores productivos de la sociedad, por mucho que sea un sector menos vulnerable en este sentido por abarcar tareas poco rutinarias y una alta demanda de habilidades sociales. Vid. al respecto European Group on Ethics in Science and New Technologies, «Future of Work, Future of Society».

[12] Schoenbaum y Gottlieb, «Algorithm based improvement of clinical quality.», 1374.

[13] Vid. Parasuraman y Manzey, «Complacency and Bias in Human Use of Automation: An Attentional Integration».

[14] Cabitza, Ciucci, y Rasoini, «A Giant with Feet of Clay: On the Validity of the Data that Feed Machine Learning in Medicine», 122.

[15] Los datos relativos a la salud, en cuanto útiles para el diagnóstico, se integran en la actualidad no solo de fuentes más tradicionales del sector sanitario que han sido digitalizadas –la HC electrónica (HCe), datos de ensayos clínicos, o datos de secuenciación genética y microbiómica–, sino también por el internet de las cosas y aplicaciones que permiten la monitorización constante de nuestra vida diaria. Acerca de la consideración jurídica de estos datos en el contexto actual, vid. Romeo Casabona, «Revisión de las categorías jurídicas de la normativa europea ante la tecnología del big data aplicada a la salud».

[16] Bjerring y Busch, «Artificial Intelligence and Patient-Centered Decision-Making», 2.

[17] Prainsack, «The political economy of digital data: introduction to the special issue».

[18] Mittelstadt, «"The doctor will not see you now ": The algorithmic displacement of virtuous medicine», 3.

[19] Harvey y Cabitza, «Algorithms are the new drugs? Reflections for a culture of impact assessment and vigilance», 285.

Aunque su objeto son los sistemas de IA que, a partir del tratamiento de datos personales, pueden utilizarse para fines de diagnóstico, prognosis o de elección terapéutica, esta investigación centra su atención en las personas e instituciones que forman ese sistema sanitario partiendo de dos premisas que se desarrollan en esta introducción. Por un lado, el estado del arte muestra las complicaciones que se presentan en la actualidad para la implementación en el mundo real de sistemas de IA. Por otro, la importancia que adquiere, en este contexto y desde una perspectiva jurídico-política, el asegurar una supervisión humana efectiva para estos sistemas.

A partir de estas premisas, se desarrolla un estudio jurídico sobre la intervención humana para la toma de decisiones automatizada en el Reglamento General de Protección de Datos y, en particular, sobre la aplicación de esta normativa en el contexto del uso de sistemas de IA en la asistencia sanitaria. Este análisis pondrá de manifiesto, no obstante, que garantizar la supervisión humana de los sistemas de IA durante todo su ciclo de vida bajo la normativa de protección de datos, incluso considerando la aplicación de la normativa de productos sanitarios, resulta insuficiente. En consecuencia, la investigación finaliza con el análisis del Reglamento de IAy de las consecuencias que tendrá en este marco normativo: examinando los roles que esta intersección normativa -protección de datos, productos sanitarios y sistemas de IA de alto riesgo- otorga durante el conjunto del ciclo de vida de un sistema de IA a las distintas instituciones y personal clínico para garantizar la supervisión humana.

1.1. Las promesas de una tecnología por consolidarse

En este apartado se recoge una infinitesimal muestra del estado del arte del desarrollo de estas tecnologías en el ámbito sanitario[20]. No tiene una pretensión científica en sentido estricto, puesto que es un mero acercamiento -por parte de un jurista- al tipo de investigación que se está realizando en este ámbito, su eventual aplicación y también a las principales problemáticas que se están planteando.

Muchas de las herramientas basadas en IA que se están desarrollando para su aplicación en el ámbito biosanitario están dirigidas a su aplicación en la actividad asistencial, particularmente con fines de diagnóstico, de prognosis[21] o de elección terapéutica. Los sistemas con estos fines dentro de la actividad asistencial son los abarcados por esta investigación, aunque otros campos anexos de aplicación de IA en este ámbito puedan ser también de gran interés; por ejemplo, otras aplicaciones para el campo de la investigación biomédica, el desarrollo farmacológico, la gestión del sistema sanitario y estrategias de salud pública e incluso dirigidas a los pacientes en ámbitos como el del autocuidado[22]. También, por supuesto, es destacable la aplicación de estas tecnologías al ámbito de la robótica en el campo de la salud[23].

Volviendo a la aplicación en la actividad asistencial, destacan los sistemas construidos mediante el aprendizaje automático, y especialmente el aprendizaje profundo, para el diagnóstico basado en imagen médica. Especialidades como radiología, dermatología, oftalmología, cardiología u oncología podrían avanzar considerablemente gracias a esta clase de análisis basado en la imagen[24]. Se trabaja además por integrar en estas tecnologías una "memoria dinámica", que permita hacerlas sostenibles en un entorno clínico de continuos avances: en la tecnología de adquisición de imágenes, los protocolos o incluso en las opciones de tratamiento[25].

Por supuesto también se han desarrollado sistemas de IA alimentados por otros datos de la historia clínica –y otras fuentes de datos externas a la misma– que no son imagen médica o

[20] No puedo dejar de recomendar en este sentido varias publicaciones: Topol, *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*.; en lengua castellana Beunza Nuin, Puertas Sanz, y Condés Moreno, *Manual práctico de inteligencia artificial en entornos sanitarios*.; o el estudio Lekadir et al., «Artificial intelligence in healthcare: Applications, risks, and ethical and societal impacts».
[21] La diferencia entre diagnóstico y prognosis estriba en que los modelos de predicción diagnóstica hacen predicciones sobre el estado de salud actual de un paciente, mientras que los modelos de predicción pronóstica estiman la probabilidad de un resultado de salud en el futuro.
[22] Romeo Casabona et al., «Informe Anticipando, Inteligencia Artificial en salud: Retos éticos y legales», 10.
[23] Y en general todo el desarrollo de la sub-dimensión "física" de la IA, que incluye la intervención directa en pacientes a través de distintos dispositivos o máquinas. Vid. Fosch-Villaronga et al., «Accounting for diversity in AI for medicine».
[24] Rajpurkar et al., «AI in health and medicine», 32-33.
[25] Perkonigg et al., «Dynamic memory to alleviate catastrophic forgetting in continual learning with medical imaging».

cuyo objetivo no es el diagnóstico; la prognosis también ha destacado, por ejemplo, para la predicción de muertes prematuras por patologías crónicas o la predicción de obesidad a partir de datos tanto sociodemográficos y ambientales como genéticos, y en general, sistemas basados en la identificación de individuos con determinadas enfermedades o afecciones y su clasificación según el estadio, la gravedad y otras características[26]. El aprovechamiento de esta capacidad "predictiva" podría ser especialmente útil a la hora de optimizar la actuación y asignación de recursos en servicios de emergencia o a la hora de detectar posibles imprevistos en el transcurso de operaciones quirúrgicas[27].

Mayor ambición y atención mediática han reunido plataformas diagnósticas como *IBM Watson*[28] o *Babylon*[29], aunque su funcionamiento también ha despertado críticas y las expectativas iniciales de ambos sistemas de IA se han visto mermadas de forma considerable[30]. También la pandemia de la COVID-19 ha traído ejemplos de estas tecnologías, en especial en relación con el diagnóstico de esta enfermedad a partir de radiografías torácicas con algoritmos de aprendizaje profundo[31].

Este conjunto de tecnologías destinadas a utilizarse en el ámbito clínico y enfocadas, por lo general, a la consecución de mejoras en la salud de la ciudadanía y a la mejora de los servicios y la atención sanitaria, cuentan con un apoyo institucional y social considerablemente alto. Sin embargo, es muy diferente -más bien contraria- el grado de aceptabilidad que vemos para sistemas que, igualmente basados en la evaluación del estado de salud de las personas, pretenden utilizarse para la contratación en el entorno laboral o en el ámbito del seguro[32]. En cualquier caso, el estudio de estas tecnologías no se aborda en esta investigación.

El mayor problema de estas investigaciones, no obstante, está en que la gran mayoría se encuentran en fases experimentales tal y como se está evidenciando en la literatura. En un estudio de revisión se analizaron 516 trabajos publicados en los primeros seis meses de 2018 reportando el desempeño de los algoritmos de IA para el análisis diagnóstico de imágenes médicas, entre los cuales halló que solo el seis por ciento –31 estudios– validaron externamente sus algoritmos[33]. Otros estudios han corroborado esta carencia en la validación externa posteriormente[34], y también se ha puesto de manifiesto que pocos tienen carácter prospectivo o realizan ensayos aleatorizados (RCT)[35]. Y cuando se realizan estos ensayos, la adherencia a estándares de calidad como CONSORT-AI resulta inexistente[36]. Con carácter general, la IA médica continúa en una fase temprana de validación e implementación[37].

Estas circunstancias están provocando el actual desajuste entre las expectativas y los resultados reales de los enfoques de IA para realizar predicciones y diagnósticos precisos en la atención sanitaria[38]. Todas estas conclusiones apuntan también al marco jurídico, una revisión de las aplicaciones basadas en algoritmos de aprendizaje automático para evaluar el riesgo de

[26] Lekadir et al., «Artificial intelligence in healthcare: Applications, risks, and ethical and societal impacts», 23.
[27] Lekadir et al., 22-23.
[28] Una descripción del producto IBM Watson puede encontrarse aquí: https://www.ibm.com/es-es/watson-health
[29] Una descripción del producto Babylon puede encontrarse aquí: https://www.babylonhealth.com/us/ai/learn-more
[30] Sobre qué pudo ir mal con IBM Watson: https://www.linkedin.com/pulse/what-went-wrong-ibm-watson-health-william-chan/; también una crónica sobre el fatal devenir de Babylon puede leerse aquí: https://twitter.com/DrMurphy11/status/1038889635152363522?s=20
[31] En España hubo varias iniciativas en este sentido, vid. Diario ABC. «Coronavirus: un nuevo sistema diagnostica Covid-19 por radiografía con un 97% de éxito». Publicado el 4 de mayo de 2020. Disponible en: https://www.abc.es/espana/comunidad-valenciana/abci-coronavirus-nuevo-sistema-diagnostica-covid-19-radiografia-97-por-ciento-exito-202005031142_noticia.html?ref=https%3A%2F%2Fwww.google.com%2F
[32] Sartor y Lagioia, «The impact of the General Data Protection Regulation (GDPR) on artificial intelligence (PE 641.530)», 28.
[33] Kim et al., «Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: Results from recently published papers».
[34] Liu et al., «A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis»; Martinez-Millana et al., «Artificial intelligence and its impact on the domains of universal health coverage, health emergencies and health promotion: An overview of systematic reviews».
[35] Nagendran et al., «Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies».
[36] Plana et al., «Randomized Clinical Trials of Machine Learning Interventions in Health Care: A Systematic Review».
[37] Rajpurkar et al., «AI in health and medicine», 34.
[38] Dick et al., «Accuracy of Computer-Aided Diagnosis of Melanoma: A Meta-analysis».

cáncer de piel reveló metodologías deficientes y peores resultados, y concluyó que el actual proceso de regulación del "marcado CE" no proporciona una protección adecuada[39].

En consecuencia, asistimos en la actualidad a una brecha entre las expectativas y los resultados reales de la IA, en la literatura, varias publicaciones se han referido a esta problemática como la última milla de la implementación de la IA[40]. Esta "última milla" representa la dificultad para validar en el entorno real clínico los resultados de algoritmos con un altísimo rendimiento estadístico o de "laboratorio"[41].

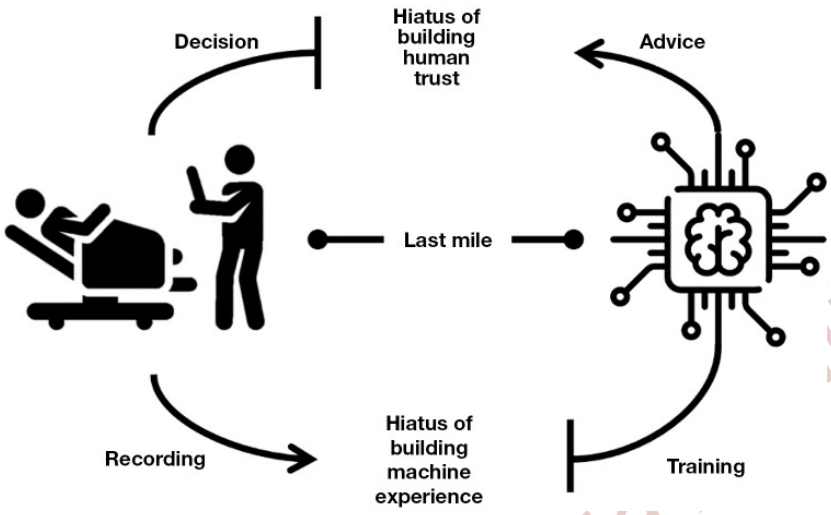

Ilustración 1. Cabitza et al. 2020 (licencia CC)

Dentro de esta última milla, Cabitza et al. definen dos hiatos, en su sentido anatómico, dos hendiduras abiertas[42]. La primera representa el fracaso de la IA médica para tener un impacto positivo en las decisiones de los y las profesionales, independientemente de la precisión intrínseca de la herramienta. Esto puede deberse, por ejemplo, a que la interfaz es inadecuada, o a la aparición de sesgos de automatización o de complacencia del ser humano frente a la automatización de sus tareas. Esta hendidura que aborda la confianza humana en el resultado algorítmico está siendo objeto de estudio y preocupación en la literatura, muy especialmente en lo que se refiere a los efectos en la interpretabilidad de los resultados del denominado como efecto “caja negra”.

Ciertamente, la casi imposibilidad de comprender el modo de funcionamiento matemático de los sistemas de IA más complejos generará complejas respuestas en los profesionales[43]. Las preocupaciones por este fenómeno en el ámbito clínico incluyen problemas de rendición de cuentas y transparencia, de discriminación oculta y sesgos emergentes que comprometan la seguridad y calidad de la asistencia, o el surgimiento de resultados inciertos que potencialmente socaven la autoridad epistémica de los expertos clínicos que utilicen las

[39] Freeman et al., «Algorithm based smartphone apps to assess risk of skin cancer in adults: systematic review of diagnostic accuracy studies»; También con similares resultados en relación a la predicción del cáncer de mama con sistemas de IA, vid. Freeman et al., «Use of artificial intelligence for image analysis in breast cancer screening programmes: systematic review of test accuracy».
[40] Coiera, «The Last Mile: Where Artificial Intelligence Meets Reality»; Cabitza, Campagner, y Balsano, «Bridging the “Last Mile” Gap between AI Implementation and Operation: “Data Awareness” That Matters».
[41] Esta dificultad puede no ser tal, y es que en ocasiones no hay ni siquiera un esfuerzo por realizarse. Cabitza et al. y concluyen con varias recomendaciones entre las cuales destacaría la primera y la última de ellas: *(1) Haga una validación externa; (…) y (4) Y entonces, de nuevo: haga una validación externa, aunque no le gusten los resultados.* Vid. Cabitza et al., «The importance of being external. methodological insights for the external validation of machine learning models in medicine».
[42] Vid. Cabitza, Campagner, y Balsano, «Bridging the “Last Mile” Gap between AI Implementation and Operation: “Data Awareness” That Matters».
[43] García Garmendia, «Spanish influenza score: Predictive power without giving up the classic», 67-68.

herramientas, entre otras[44]. Esta opacidad epistemológica de determinados sistemas de IA y las problemáticas que acarrea, ha sido objeto de amplios debates en la literatura acerca de su compatibilidad con la práctica médica[45].

Ahora bien, hay una segunda hendidura directamente relacionada con la producción y la disponibilidad de una cantidad suficiente de datos clínicos, fiables y exactos, que sean adecuados para ser la "experiencia" con la que se puede entrenar una máquina. Cabitza et al. señalan que científicos de datos y desarrolladores de sistemas de IA tienden a suponer que los conjuntos de datos con los que se entrenan sus modelos de predicción son (I) veraces, (II) fiables y (III) representativos de la población[46]. Sin embargo, esta triple hipótesis rara vez es defendible y a menudo está mal fundamentada, al menos en cierta medida[47].

Por supuesto, y aunque no pueden ser olvidadas otras cuestiones relevantes como la interoperabilidad de los sistemas[48] o la multiplicidad de los datos[49], en esta hendidura los sesgos cobran un protagonismo esencial que puede tener repercusión jurídica en distintas fases: entrenamiento de los sistemas de IA, procesos de validación, interpretación de los datos de salida o impugnación de los mismos por supervisores humanos. Parece evidente que las inexactitudes derivadas de los sesgos en este contexto pueden dar lugar a eventos adversos que perjudiquen la salud de los pacientes, como mínimo.

En lo que respecta al desarrollo de esta investigación, es importante analizar cómo se interrelaciona la supervisión humana con estas problemáticas.

Aunque la mayoría de los estudios se han centrado en la comparación directa -y en cierto sentido falaz, como se verá en el desarrollo de esta investigación- de los sistemas de IA con los seres humanos, en la práctica médica los seres humanos tendrán una colaboración activa con los sistemas de IA[50]. Para Fosch-Villaronga et al. en estos procesos de automatización no se produce una disminución de la supervisión humana como tal, y la posición del personal clínico seguirá siendo integral y crucial para muchas funciones[51]. Es decir, el rol humano se está redefiniendo con el uso de estos sistemas, y las personas proporcionarán una supervisión de los mismos que, obligada o no por el ordenamiento jurídico -no entremos por ahora en esto-, tiene una relación directa con su implementación, rendimiento y seguridad: ¿es posible mitigar los sesgos algorítmicos a partir de la supervisión humana de la IA? ¿cómo afecta la opacidad de estos sistemas a la posibilidad de supervisar su funcionamiento? ¿cómo se verá afectado el juicio humano ante el uso prolongado de estos sistemas? ¿cómo pueden diseñarse procesos de toma de decisiones para garantizar una supervisión humana efectiva en la implementación de los sistemas de IA?

1.2. La importancia de la supervisión humana en este contexto

Desde un punto de vista político-jurídico, pocas cuestiones han suscitado tanta atención de las instituciones europeas como la regulación de los sistemas de IA -no solo en el ámbito de la

[44] Durán y Jongsma, «Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI», 329.
[45] Por ejemplo, Núñez Reiz y Sánchez García, «En respuesta a “Big Data Analysis y Machine Learning en medicina intensiva: identificando nuevos retos ético-jurídicos”».
[46] Cabitza, Campagner, y Balsano, «Bridging the “Last Mile” Gap between AI Implementation and Operation: “Data Awareness” That Matters», 3.
[47] Así lo constata el informe de la Agencia de los Derechos Fundamentales de la Unión Europea (FRA): *Studies where patients were shown their medical files and asked about their accuracy found that up to 50 % of information was incomplete or erroneous. A lot of important data in EMR/EHR is unstructured in the form of free text, which further reduces data quality. Low levels of accuracy, completeness and overall data quality increases the risk of medical error.* Agencia de los Derechos Fundamentales de la Unión Europea (FRA), «Getting the future right - Artificial intelligence and fundamental rights», 39.
[48] Vid. Recomendación (UE) 2019/243 de la Comisión, de 6 de febrero de 2019, sobre un formato de intercambio de historiales médicos electrónicos de ámbito europeo. Considerando 18: *Digitalizar los historiales médicos y propiciar su intercambio podrían contribuir también a la creación de grandes estructuras de datos sanitarios que, junto con el uso de nuevas tecnologías, como la analítica de macrodatos y la inteligencia artificial, pueden contribuir a nuevos descubrimientos científicos*.
[49] Sobre este tema, vid. Cabitza et al., «The elephant in the record: On the multiplicity of data recording work».
[50] Rajpurkar et al., «AI in health and medicine», 33.
[51] Fosch-Villaronga et al., «A human in the loop in surgery automation».

asistencia sanitaria- para afrontar, entre otros, las dificultades y retos planteados en el anterior apartado.

En 2018, la Comisión hizo pública la Estrategia europea sobre la IA[52] con el triple objetivo de potenciar la capacidad tecnológica e industrial de la Unión, prepararse para las transformaciones socioeconómicas que origina la IA y garantizar el establecimiento de un marco ético y jurídico apropiado para la misma. En paralelo, y como resultado del compromiso conjunto de los EEMM, la Comisión aprobaba también un Plan coordinado sobre la inteligencia artificial para la generación de una IA "*made in Europe*"[53], que fue renovado en 2021[54]. El grupo de expertos de alto nivel sobre inteligencia artificial (AI HLEG) fue designado y constituido para asesorar sobre la Estrategia europea en materia de IA, y sus resultados sirvieron de recursos para nuevas iniciativas políticas[55]. Entre otras, el 19 de febrero de 2020 se publica el Libro Blanco sobre la IA por la Comisión[56], coincidiendo a su vez con la publicación de la Estrategia europea de datos[57]. El Libro Blanco definió un ecosistema de confianza en el que debe promoverse un marco normativo para la AI que aborde las oportunidades y los riesgos de estas tecnologías. Más recientemente, el Parlamento Europeo aprobaba una ambiciosa Resolución de 20 de octubre de 2020[58], en la que se incluía un anexo con una propuesta legislativa para la tramitación de un Reglamento sobre los principios éticos para el desarrollo, el despliegue y el uso de la inteligencia artificial, la robótica y las tecnologías conexas.

En todas estas iniciativas es necesario destacar la importancia que ha adoptado la supervisión humana como requisito indispensable para la regulación de los sistemas de IA de alto riesgo, aunque los acercamientos normativos a esta cuestión hayan variado en distintas iniciativas.

En el Libro Blanco sobre IA, la Comisión consideró que sólo se puede lograr una IA fiable mediante una participación adecuada de los seres humanos en relación con las aplicaciones de IA de alto riesgo[59]. Para ello, puede requerirse un tipo y grado diferente de participación humana dependiendo del uso previsto de la IA y sus potenciales efectos. Fuera del alcance de esta clase de participación quedarían las manifestaciones indirectas de intervención humana[60]. Por consiguiente, el Libro Blanco sobre IA entendió que el requisito de la supervisión humana se logra mediante mecanismos de gobernanza que requieren diferentes tipos de participación o intervención humana en el proceso de adopción de decisiones, y ofreció una lista no exhaustiva de esos mecanismos de gobernanza[61].

Aunque con menor desarrollo, 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

[52] Comisión Europea, Comunicación de la Comisión «Inteligencia artificial para Europa». Bruselas, 25.4.2018. COM(2018) 237 final. Disponible en: https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=COM%3A2018%3A237%3AFIN

[53] Comisión Europea, Comunicación de la Comisión «Plan coordinado sobre la inteligencia artificial». Bruselas, 7.12.2018. COM(2018) 795 final. Disponible en: https://eur-lex.europa.eu/legal-content/ES/ALL/?uri=COM:2018:795:FIN

[54] Comisión Europea, Comunicación de la Comisión «Fomentar un planteamiento europeo en materia de inteligencia artificial». Bruselas, 21.4.2021. COM(2021) 205 final. Disponible en: https://eur-lex.europa.eu/legal-content/ES/ALL/?uri=COM:2021:205:FIN

[55] Comisión Europea, Comunicación de la Comisión «Generar confianza en la inteligencia artificial centrada en el ser humano». Bruselas, 8.4.2019. COM(2019) 168 final. Disponible en: https://eur-lex.europa.eu/legal-content/ES/ALL/?uri=CELEX:52019DC0168

[56] Comisión Europea, Comunicación de la Comisión «Libro Blanco sobre la inteligencia artificial – un enfoque europeo orientado a la excelencia y la confianza». Bruselas, 19.02.2020. COM (2020) 65 final.

[57] Comisión Europea, Comunicación de la Comisión «Una Estrategia Europea de Datos». Bruselas, 19.02.2020. COM (2020) 66 final.

[58] Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas (2020/2012(INL)). Disponible en: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_ES.html

[59] Comisión Europea, Libro Blanco sobre la inteligencia artificial, 21.

[60] Tales como los procedimientos de prueba, inspección o certificación necesarios para verificar el cumplimiento de los requisitos obligatorios. En un sentido amplio podrían considerarse formas de intervención humana, pero estos mecanismos de gobernanza no están integrados activamente en el proceso de adopción de decisiones.

[61] Por ejemplo, y entre otros: *El resultado del sistema de IA no es efectivo hasta que un humano no lo haya revisado y validado (por ejemplo, la decisión de denegar una solicitud de prestaciones de seguridad social solo podrá adoptarla un ser humano)*. Esta clase de mecanismo, como veremxos más adelante, es el adoptado por el RGPD en su artículo 22 con la prohibición de la toma de decisiones, incluida la elaboración de perfiles, basada únicamente en el tratamiento automatizado que obliga al responsable del tratamiento a introducir en el proceso decisorio a un interventor humano, salvo que acuda a alguna de las bases de legitimación excepcionales de los apartados 22(2) y (4) RGPD.

conexas también recalcó la importancia de la supervisión humana como mecanismo de gobernanza para estos sistemas, estableciendo como un principio ético de obligado cumplimiento el desarrollo de una IA antropocéntrica, antropogénica y controlada por seres humanos en su artículo séptimo, cuyo contenido puede resumirse en la garantía de una supervisión humana integral (apartado primero) que, en todo caso, permita el restablecimiento del control humano cuando sea necesario (apartado segundo). Dicho apartado primero establece que debe de garantizarse en todo momento, haciendo referencia a las fases de desarrollo, despliegue y uso, una supervisión humana integral. No opta, en definitiva, por ningún mecanismo de intervención humana en particular, sino que se limita a regular que la supervisión humana debe darse en todo el ciclo de vida de la IA.

Finalmente, se produjo la llegada del Reglamento de IA, tras su entrada en vigor el primero de agosto de 2024. Se trata de una normativa con un enfoque basado en el riesgo que diferencia entre prácticas de IA prohibidas (Cap. 2 RIA), sistemas de IA de alto riesgo (Cap. 3 RIA), obligaciones de transparencia para determinados sistemas de IA (Cap. 4 RIA) y modelos de IA de uso general (Cap. 5 RIA). Así, el grueso del Reglamento se centra en los sistemas de alto riesgo, y dando un especial protagonismo a las obligaciones y requisitos de obligado cumplimiento para el diseño y desarrollo de los sistemas de IA antes de su comercialización por parte de los proveedores. Como era de esperar, la UE destacó nuevamente el peso de la supervisión humana, estableciendo en su artículo 14 la misma como requisito obligatorio de los sistemas de IA de alto riesgo. Así, estos sistemas deben ser diseñados y desarrollados de forma que se garantice que puedan ser supervisados eficazmente por agentes humanos durante el uso y despliegue de los mismos. El modelo regulatorio por el que ha optado la Comisión responsabiliza a los proveedores de los sistemas del cumplimiento de los requisitos obligatorios como la supervisión humana -aunque como veremos establece también algunas obligaciones en relación con los responsables del despliegue-, poniendo definitivamente el enfoque normativo sobre las fases de diseño y desarrollo de estas tecnologías que habían permanecido fuera del análisis doctrinal jurídico y de las iniciativas políticas anteriores[62].

Resulta oportuno subrayar dos características fundamentales sobre las que se construye la supervisión humana como requisito obligatorio para los sistemas de IA de alto riesgo. Por un lado, el conjunto de obligaciones está principalmente dirigido al proveedor en la fase de diseño y desarrollo de los modelos, reforzando un modelo de supervisión durante todo el ciclo de vida de los sistemas. Por otro lado, no se opta por un mecanismo de gobernanza concreto, sino por cualificar dicha supervisión, requiriendo de medidas que puedan procurar una supervisión *efectiva* por personas físicas, que consiste en que las personas a quienes se encomiende la supervisión del sistema deben ser capaces de, entre otros, entender por completo las capacidades y limitaciones del sistema, ser conscientes del sesgo de automatización, interpretar correctamente la información de salida del sistema, desestimar, invalidar o revertir dicha información o interrumpir el sistema accionando un botón específicamente destinado a tal fin (art. 14(4) RIA).

El desafío normativo actual consiste en determinar cuál es la participación humana adecuada para las numerosas aplicaciones de los modelos algorítmicos y sus distintos ámbitos de aplicación para la toma de decisiones automatizada (¿qué forma debe adoptar la participación?; ¿cuál es el fundamento de esta participación?; ¿para qué sistemas debe ser obligatoria?)[63]; y en cómo determinar en el ámbito normativo la cualificación que se exija para dicha participación (¿qué es adecuado, efectivo o significativo?). El RGPD ya incluyó mecanismos de gobernanza basados en la intervención humana para la toma de decisiones automatizada, incluida la elaboración de perfiles, veamos de qué forma.

[62] Lehr y Ohm, «Playing with the data: what legal scholars should learn about machine learning», 655 y ss.
[63] Para el AI-HLEG los mecanismos de supervisión deberán apoyar otras medidas de seguridad y control, y añade: (...) Si el resto de las circunstancias no cambian, cuanto menor sea el nivel de supervisión que pueda ejercer una persona sobre un sistema de IA, mayores y más exigentes serán las verificaciones y la gobernanza necesarias. Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial de la Comisión Europea (HLEG-AI), *Directrices éticas para una IA fiable*, 20.

2. La intervención humana para la toma de decisiones automatizada basada en la elaboración de perfiles en el RGPD

En su desarrollo del derecho de protección de datos, el RGPD trata de afrontar los nuevos retos que la rápida evolución tecnológica presenta, en un mundo globalizado en el que la magnitud de la recogida y del intercambio de datos personales ha aumentado de manera significativa, permitiendo que empresas privadas y autoridades públicas utilicen datos personales en una escala sin precedentes -considerando 6 RGPD-. Entre estos retos está, sin duda, el uso de sistemas de IA que tratan datos personales para elaborar inferencias sobre el estado de salud de las personas objeto de esta investigación. Si nos centramos en la regulación de los resultados de estos sistemas son dos las figuras fundamentales que encontramos en el RGPD: la elaboración de perfiles y la toma de decisiones automatizada.

Ferretti et al. consideran que los sistemas de IA con fines relacionados con la salud han de ser considerados como elaboración de perfiles conforme a la definición del RGPD[64]. Ello será así siempre que el sistema de IA cumpla con determinados requisitos. El tratamiento de datos personales (datos de entrada) por el sistema es el primer requisito necesario, considerando que se trata del requisito básico para que el Reglamento sea aplicable[65]. A su vez, el sistema debe realizar inferencias (datos de salida) que evalúen aspectos personales de la salud de una persona física para ser considerado elaboración de perfiles[66]; si el sistema evaluase un aspecto personal distinto, no querría decir que no se considera elaboración de perfiles, sino que no estaríamos ante un sistema de IA con fines de salud en sentido estricto[67].

Una vez establecido en qué medida podemos considerar que la implementación de sistemas de IA con fines de salud debe considerarse elaboración de perfiles a efectos del RGPD, conviene aclarar que la regulación de la toma de decisiones automatizada basada en la elaboración de perfiles en el RGPD: ni se limita exclusivamente a las disposiciones contenidas en el artículo 22 RGPD, ni se aplica exclusivamente a decisiones basadas únicamente en el tratamiento automatizado.

Esta investigación parte de la distinción en el RGPD de tres clases de decisiones automatizadas -en sentido amplio- que pueden producirse en la implementación de sistemas de IA con fines de salud.

Por un lado, tenemos decisiones que pueden o no producir un efecto jurídico o de afectación significativa similar (en adelante efectos jurídicos o significativos). Por otro lado, las decisiones pueden o no estar basadas únicamente en el tratamiento automatizado. Sin embargo, si una decisión no produce ese efecto jurídico o significativo, es decir, si se coloca por debajo del umbral de ese enfoque basado en el riesgo, será indiferente que la misma esté o no basada únicamente en el tratamiento automatizado.

Ello quiere decir que la primera clase de decisiones que podemos definir son [1] las decisiones que no producen efectos jurídicos o significativos estén o no basadas únicamente en el tratamiento automatizado. Ahora bien, entre las decisiones que sí producen dichos efectos, debemos distinguir necesariamente dos clases de decisiones[68]; [2] las decisiones que producen efectos jurídicos o significativos que no están basadas únicamente en el tratamiento

[64] Ferretti, Schneider, y Blasimme, «Machine learning in medicine: Opening the New Data Protection Black Box», 323. 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 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*.

[65] Por supuesto, no es necesario que todos los datos de entrada que el sistema requiere sean de carácter personal, en principio es suficiente con que alguno de los datos tratados tenga carácter personal para la aplicación de la normativa de protección de datos.

[66] Por ejemplo, no se consideraría elaboración de perfiles el sistema de IA que, a partir de los datos de distintos pacientes, predice la sobrecarga que determinada especialidad clínica de un centro hospitalario va a sufrir en las próximas semanas.

[67] Pongamos, por ejemplo, un sistema de IA que evalúa para un hospital privado el riesgo de impago por parte de los usuarios. Esta clase de sistemas no son objeto de esta investigación.

[68] Art. 22(1) RGPD: *Todo interesado tendrá derecho a no ser objeto de una decisión basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzca efectos jurídicos en él o le afecte significativamente de modo similar*.

automatizado –legítimas conforme al apartado 22(1)–, esto es, sistemas de apoyo a la toma de decisiones; y [3] las decisiones que producen efectos jurídicos o significativos y que están basadas únicamente en el tratamiento automatizado –prohibidas con carácter general por el apartado 22(1) y solo excepcionalmente legitimadas[69] siempre que se adopten medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado[70]–.

Dado el objeto de esta investigación, debe matizarse que las decisiones adoptadas a partir de sistemas de IA con fines de salud están basadas en categorías especiales de datos del artículo 9(1) RGPD[71]. Por tanto, es aplicable la prohibición general contenida en el apartado 22(4)[72]. Ello se traduce en que no cabe la adopción de decisiones basadas únicamente en el tratamiento de datos que produzcan efectos jurídicos o significativos sobre las excepciones contenidas en el apartado 22(2), sino que únicamente pueden adoptarse sobre las excepciones del apartado 22(4).

Las decisiones que no superan el umbral del riesgo [1], no entran en el ámbito de aplicación del artículo 22, es decir, el artículo 22 no se aplica a decisiones de bajo impacto[73]. En cambio, sí entran en su ámbito de aplicación las decisiones que alcanzan ese umbral de riesgo [2] y [3] y, para cada una, el RGPD establece un mecanismo obligatorio de intervención humana distinto[74].

Además, el RGPD no es un mero instrumento normativo para el reconocimiento y ejercicio de derechos que permite el control individual de los datos personales, sino que representa el control colectivo sobre la justificación de los procesos que recopilan, procesan y usan los datos. Y ello se traduce especialmente en el establecimiento de una sólida responsabilidad sobre el cumplimiento normativo y su demostración como piedra angular del RGPD, también en la regulación de la toma de decisiones automatizada.

Es decir, la ubicación entre los derechos de la persona interesada de la prohibición de la toma de decisiones basada únicamente en el tratamiento automatizado o de los derechos de información y acceso, no debe inducirnos a pensar que el RGPD regula fundamentalmente la toma de decisiones automatizada a partir de dicho reconocimiento y ejercicio de derechos que permite el control individual de los datos personales. Esta visión reduccionista se trata de refutar posteriormente, poniendo de manifiesto que desde la perspectiva de la gestión de riesgos exigida por el principio de responsabilidad del RGPD, la intervención humana constituye una medida organizativa que debe adoptarse necesariamente por el responsable del tratamiento al establecer procesos de toma de decisiones basados en la elaboración de perfiles. Y para la implementación de sistemas de IA con fines de salud en particular, se analiza la intervención humana como medida organizativa de la evaluación de impacto de protección de datos (EIPD, en adelante).

---

[69] Art. 22(2) RGPD: *El apartado 1 no se aplicará si la decisión: a) es necesaria para la celebración o la ejecución de un contrato entre el interesado y un responsable del tratamiento; b) está autorizada por el Derecho de la Unión o de los Estados miembros que se aplique al responsable del tratamiento y que establezca asimismo medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado, o c) se basa en el consentimiento explícito del interesado.*

[70] Entre estas medidas, se contempla con carácter general en el apartado 22(3) RGPD el derecho a obtener intervención humana por parte del responsable con posterioridad a la adopción de la decisión basada únicamente en el tratamiento automatizado y que produce efectos jurídicos o significativos. En los términos del considerando 71 RGPD: *(…) En cualquier caso, dicho tratamiento debe estar sujeto a las garantías apropiadas, entre las que se deben incluir la información específica al interesado y el derecho a obtener intervención humana, a expresar su punto de vista, a recibir una explicación de la decisión tomada después de tal evaluación y a impugnar la decisión.*

[71] Art. 9(1) 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 las orientación sexuales de una persona física.*

[72] Art. 22(4) RGPD: *Las decisiones a que se refiere el apartado 2 no se basarán en las categorías especiales de datos personales contempladas en el artículo 9, apartado 1, salvo que se aplique el artículo 9, apartado 2, letra a) o g), y se hayan tomado medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado.*

[73] Brkan, «Do Algorithms Rule the World? Algorithmic Decision-Making in the Framework of the GDPR and Beyond», 93. Solo a efectos de *graves consecuencias* según el Grupo de Trabajo sobre Protección de Datos del Artículo 29, «Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679», 23.

[74] Distinción sobre la que se vuelve en el siguiente apartado.

2.1. Intervención humana como mecanismo de gobernanza en el artículo 22 RGPD: dos clases distintas de intervención

La intervención humana como mecanismo de gobernanza de la toma de decisiones automatizada en el RGPD no ha sido protagonista de los debates doctrinales sobre esta figura.

Por lo general, el análisis doctrinal ha centrado el interés en los derechos de información y acceso y, en particular, la discusión relativa a la existencia o no de un derecho a una explicación para las decisiones automatizadas y su extensión. Esto es, en la aplicación de los derechos y obligaciones derivadas del principio de transparencia por lo general. Una perspectiva ya destacada en la tramitación legislativa del RGPD por el Supervisor Europeo de Protección de Datos (SEPD), que al referirse a la elaboración de perfiles y la toma de decisiones a partir de éstos, recomendaba un mayor grado de transparencia por parte de los responsables del tratamiento en los siguientes términos: «*El problema no es (…) la práctica de la elaboración de perfiles, sino, más bien, la falta de información adecuada sobre la lógica algorítmica a partir de la que se desarrollan tales perfiles y que repercute en el interesado*»[75].

No obstante, esta perspectiva ha tenido en cierto sentido un efecto disfuncional sobre el análisis jurídico de la toma de decisiones automatizada en el RGPD, eclipsando el resto de mecanismos de gobernanza que se establecen para la regulación de esta clase de tratamiento de datos personales[76]. Entre otros, la intervención humana.

Volviendo sobre las tres clases de decisiones automatizadas distinguidas anteriormente, se ha destacado que para las decisiones que producen efectos jurídicos o significativos se establecen distintas formas de intervención humana en función de si están basadas únicamente en el tratamiento automatizado [3], o no [2]. Siempre que se produce un efecto jurídico o significativo, la prohibición del apartado primero introduce la intervención humana como componente esencial de la toma de decisiones automatizada [2], en forma de *human in the loop*. Así, el responsable del tratamiento puede esquivar la prohibición siempre que no base la decisión únicamente en el tratamiento automatizado, es decir, introduciendo un agente humano en el proceso de toma de decisiones.

Por otro lado, también cabe la toma de decisiones basada únicamente en el tratamiento automatizado cuando se acuda a alguna de las excepciones contenidas en el apartado segundo, es decir, a las excepciones para la prohibición general contenida en el apartado primero[77]. En este caso, el RGPD legitima la toma de decisiones basada únicamente en el tratamiento automatizado que produce efectos jurídicos o significativos [3], ahora bien, condiciona dicha legitimación al cumplimiento de las medidas de salvaguarda contenidas en el apartado tercero[78], entre las cuales, el RGPD introduce la intervención humana como medida de salvaguarda bajo requerimiento del interesado, en forma de *human out of the loop*.

Ello se traduce en que los distintos mecanismos de intervención humana se integran en tipos de decisiones diferentes y, por ende, en distintas fases de la toma de decisiones; en el primer caso, la intervención humana como componente esencial de la toma de decisiones automatizada tiene lugar antes de la producción de efectos jurídicos o significativos para la persona interesada; mientras que la intervención humana como medida de salvaguarda bajo requerimiento del interesado tiene lugar tras la producción de dichos efectos, tal y como puede observarse en la siguiente tabla[79]:

---

[75] Supervisor Europeo de Protección de Datos (SEPD), «Dictamen 3/2015 -La gran oportunidad de Europa. Recomendaciones del SEPD sobre las opciones de la UE en cuanto a la reforma de la protección de datos-», 10.

[76] Sobre este efecto, vid. más ampliamente Hert y Lazcoz, «When GDPR-principles blind each other. Accountability, not transparency, at the heart of algorithmic governance».

[77] Una vez más, si la decisión se adopta sobre categorías especiales de datos a las que se refiere el artículo 9 RGPD, las excepciones de dicha prohibición están limitadas a las establecidas por el apartado 4 del artículo 22 RGPD.

[78] Salvo para el caso de que la decisión automatizada se encuentre autorizada por el Derecho de la Unión o de los EEMM, en cuyo caso las medidas de salvaguarda no son necesariamente las contenidas en el apartado tercero; en cualquier caso, la norma debe establecer medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado.

[79] Cuando el tratamiento está basado en categorías especiales de datos personales se aplica la prohibición agravada del apartado 22(4) RGPD y, por ende, han de aplicarse las excepciones y medidas de salvaguarda contenidas en dicho apartado.

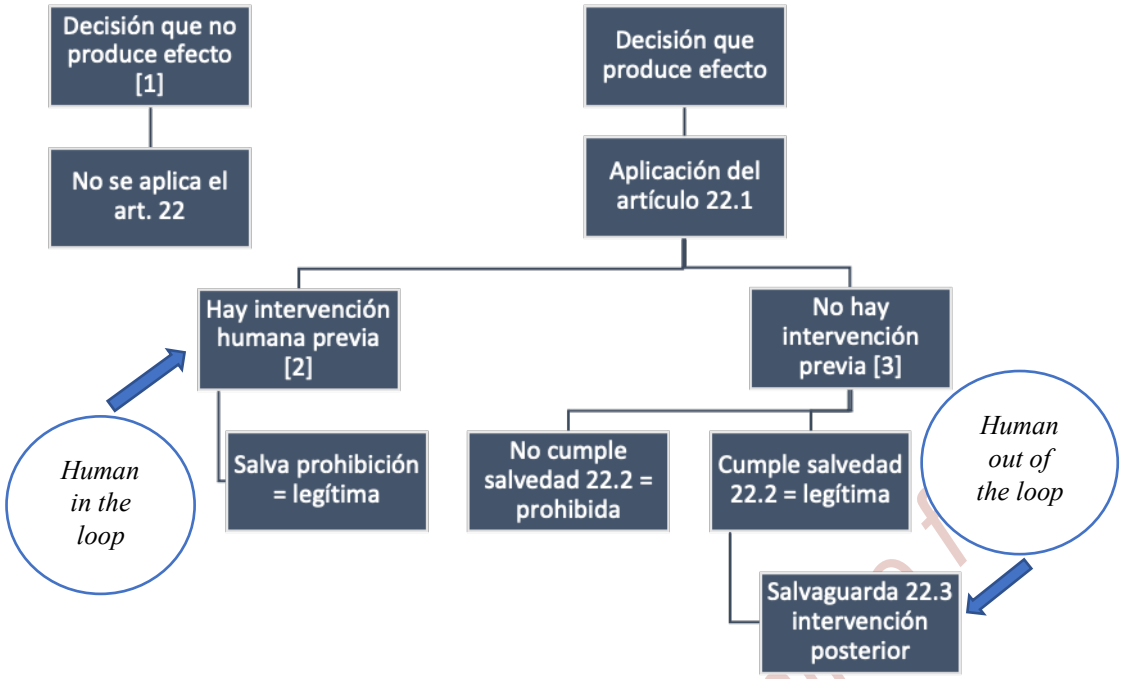

Tabla 1. Elaborada por el autor

Esta investigación se centra exclusivamente en el primero de los mecanismos, o lo que es lo mismo, en la aplicación de la intervención humana para la implementación de sistemas de IA con fines de salud para el apoyo a la toma de decisiones clínica. Ello responde a varios motivos. Por un lado, las disposiciones del artículo 22 RGPD cuando la toma de decisiones se adopta sobre la base de categorías restringen notablemente las posibilidades de implementar procesos decisorios basados únicamente en el tratamiento de datos. Las únicas excepciones son la adopción de la decisión por razones de un interés público esencial, sobre la base del Derecho de la Unión o de los Estados miembros[80], y el consentimiento explícito de la persona interesada[81]. Por otro lado, considerando el estado del arte tecnológico, parece razonable entender que las decisiones basadas únicamente en el tratamiento automatizado o procesos "plenamente" automatizados no son el escenario más probable a corto y medio plazo[82], y tampoco parecen ser un escenario atractivo desde una perspectiva político-jurídica[83].

2.1.1. Intervención humana como componente esencial de las decisiones no basadas únicamente en el tratamiento automatizado

El artículo 22 RGPD prohíbe con carácter general las decisiones basadas únicamente en el tratamiento automatizado que produzcan un efecto jurídico o significativo, lo cual significa que serán legítimas aquellas decisiones que incorporen intervención humana a la toma de decisiones basada en el tratamiento automatizado de datos, incluida la elaboración de perfiles, incorporando dicha intervención de forma previa a la producción de efectos jurídicos o

[80] No contando en la actualidad con ningún desarrollo legislativo en el ordenamiento jurídico español en este sentido.
[81] Parece poco recomendable en este sentido diseñar por parte de los centros clínicos procesos para la implementación de sistemas de IA con fines de salud cuya aplicación dependa de la voluntad individual de los interesados. En este sentido, parece más garantista -también desde la perspectiva de la seguridad jurídica del responsable del tratamiento- establecer directamente procesos con intervención humana previa a la producción de efectos significativos en el entorno clínico.
[82] Vid. Topol, «High-performance medicine: the convergence of human and artificial intelligence».
[83] Al respecto, vid. más adelante la disposición establecida en la Carta de Derechos Digitales adoptada por el Gobierno de España que prioriza el derecho al libre criterio clínico del personal sanitario sobre el criterio algorítmico en el uso de esta clase de sistemas.

significativos para la persona interesada[84]. Esto es, tal y como señala Jones, la prohibición general se traduce en un derecho al *human in the loop*[85].

Por un lado, la prohibición pretende que los interesados no sean objeto de decisiones que produzcan determinados efectos sin intervención humana alguna; por otro, se constituye la introducción de la intervención humana como una garantía normativa para el derecho a no ser objeto de esta clase de decisiones prohibidas -intervención humana como componente esencial de las decisiones no basadas únicamente en el tratamiento automatizado-. De este modo, el RGPD asegura que la toma de decisiones incluye también un razonamiento humano y, por ende, el tratamiento automatizado no es el fundamento exclusivo -basadas únicamente- de la toma de decisiones[86].

Ahora bien, ¿qué clase de intervención humana es requerida por el RGPD?

Esta tarea requiere determinar el umbral mínimo de intervención humana requerido por el RGPD para considerar que una decisión no se basa únicamente en el tratamiento automatizado[87], si bien, la vaguedad del término en sí dificulta este análisis de la intervención humana[88].

2.1.1.1. Intervención humana significativa y no meramente formal

Una parte de la doctrina, al analizar la aplicación de los derechos de información y acceso para las decisiones automatizadas, ha criticado que la mera inclusión formal de intervención humana es suficiente para considerar que una decisión no está basada únicamente en el tratamiento automatizado[89]. Este argumento, no obstante, es más que rebatible.

Las directrices del Grupo de Trabajo del Artículo 29 (GT29) son claras al establecer que la intervención humana no puede ser únicamente un gesto simbólico[90]. Siguiendo esta misma postura encontramos voces autorizadas en la doctrina. Para Brkan una interpretación formalista que involucre al ser humano como parte necesaria del proceso de toma de decisiones, pero que en última instancia deje el poder de decisión a la máquina, no garantizaría un nivel suficientemente alto de protección de datos exigido por el RGPD[91]. Malgieri y Comandé apuntan a que cualquier decisión humana de carácter pasivo entra en el ámbito de la prohibición del 22(1) como decisión basada únicamente en el tratamiento automatizado[92]. Para Mendoza y Bygrave, aunque una decisión se atribuya formalmente a una persona, se

[84] Esta idea parece cristalizarse en la redacción del Considerando 71 y su primera mención a la intervención humana, cuando dice: *El interesado debe tener derecho a no ser objeto de una decisión, que puede incluir una medida, que evalúe aspectos personales relativos a él, y que se base únicamente en el tratamiento automatizado y produzca efectos jurídicos en él o le afecte significativamente de modo similar, como la denegación automática de una solicitud de crédito en línea o los servicios de contratación en red en los que no medie intervención humana alguna*.

[85] Jones, «The right to a human in the loop: Political constructions of computer automation and personhood», 224.

[86] Wagner, «Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision-Making Systems», 108.

[87] Malgieri y Comandé, «Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation», 244.

[88] Gil González y Hert, «Understanding the legal provisions that allow processing and profiling of personal data—an analysis of GDPR provisions and principles», 617.

[89] Wachter, Mittelstadt, y Floridi, «Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation», 88. Esta interpretación se apoya sobre la sentencia de 28 de enero de 2014 del Tribunal Federal de Justicia de Alemania o *Bundesgerichtshof* (BGH) para el caso SCHUFA sobre valoraciones crediticias en aplicación del artículo 15 DPD - *Scoring und Datenschutz*. 28. 1. 2014-VI ZR 156/13-.

[90] Grupo de Trabajo sobre Protección de Datos del Artículo 29, «Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679», 23. Dichas directrices fueron refrendadas por el Comité Europeo de Protección de Datos (CEPD en adelante) en su primera sesión plenaria: Endorsement 1/2018, Bruselas, 25 de mayo de 2018. Disponible aquí: https://edpb.europa.eu/news/news/2018/endorsement-gdpr-wp29-guidelines-edpb_es

[91] Brkan, «Do Algorithms Rule the World? Algorithmic Decision-Making in the Framework of the GDPR and Beyond», 101.

[92] Malgieri y Comandé, «Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation», 252.

considerará que se basa únicamente en un tratamiento automatizado si una persona no evalúa activamente el resultado del tratamiento antes de su formalización como decisión[93].

Sobre este consenso, con base en el cual una intervención humana formal o pasiva debe considerarse equivalente a una decisión basada únicamente en el tratamiento automatizado; *a contrario sensu*, podría llegarse a la conclusión de que un proceso de toma de decisiones en el que un humano participa de forma activa y tiene una influencia real sobre las decisiones finales, no sería considerado como una decisión basada únicamente en el tratamiento automatizado. Ahora bien, ¿cómo puede definirse esta clase de intervención humana en el marco del RGPD?

Aquí, las directrices refrendadas por el CEPD establecen un concepto clave: intervención humana *significativa*.

Dicho término se introduce, en primer lugar, al definir cómo las decisiones automatizadas pueden solaparse parcialmente con la elaboración de perfiles, se aporta el siguiente ejemplo en el cual una decisión no basada únicamente en el tratamiento automatizado incluye una elaboración de perfiles: «*(…), antes de conceder una hipoteca, un banco puede tener en cuenta la calificación crediticia del prestatario, y pueden producirse otras intervenciones humanas significativas adicionales antes de que se tome ninguna decisión sobre la persona*»[94]. Es decir, las decisiones no basadas únicamente en el tratamiento se relacionan con la introducción de intervenciones humanas significativas.

De nuevo, entre las formas de utilizar la elaboración de perfiles, elabora la distinción entre la toma de decisiones basada en la elaboración de perfiles y la toma de decisiones basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles[95], según las directrices esta última se aprueba y traslada a la persona interesada sin ninguna evaluación previa y significativa por parte de un ser humano[96]. De forma que las directrices reafirman que dicha intervención significativa debe tener lugar antes de la producción de efectos jurídicos o significativos, en forma de *human in the loop*.

Por último, a la hora de analizar qué clase de participación humana evita la toma de decisiones basadas únicamente en el tratamiento automatizado, las directrices ponen el foco sobre las obligaciones del responsable. El responsable del tratamiento no puede eludir la prohibición "fabricando" una intervención humana y, por lo tanto, los responsables del tratamiento deben garantizar que cualquier intervención humana sea significativa para el proceso de toma de decisiones del apartado 22(1)[97]. Este concepto ha sido igualmente recogido por la Agencia Española de Protección de Datos (en adelante, AEPD) en su Guía sobre la adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial: «*Para que pueda considerarse que existe participación humana, la supervisión de la decisión ha de ser realizada por una persona*

[93] Mendoza y Bygrave, «The Right Not to be Subject to Automated Decisions Based on Profiling BT -», 4. La agencia de protección de datos de Reino Unido, Information Commissioner's Office (ICO), recoge en sus directrices sobre decisiones automatizadas y elaboración de perfiles que la participación humana debe tener un carácter activo y no meramente simbólico: *The question is whether a human reviews the decision before it is applied and has discretion to alter it, or whether they are simply applying the decision taken by the automated system*. ICO, «Guide to the UK General Data Protection Regulation (UK GDPR)», 8.

[94] Grupo de Trabajo sobre Protección de Datos del Artículo 29, «Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679», 9.

[95] En los términos expuestos anteriormente, pero poniendo el foco sobre la elaboración de perfiles, esta sería la misma diferenciación realizada para las [2] las decisiones que producen efectos jurídicos o significativos que no están basadas únicamente en el tratamiento automatizado – legítimas conforme al apartado 22(1) –; y [3] las decisiones que producen efectos jurídicos o significativos y que están basadas únicamente en el tratamiento automatizado – prohibidas con carácter general por el apartado 22(1) –.

[96] Grupo de Trabajo sobre Protección de Datos del Artículo 29, «Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679», 9.

[97] Grupo de Trabajo sobre Protección de Datos del Artículo 29, 23. Es cierto que el responsable del tratamiento puede eludir a través de la intervención humana los derechos de información y acceso establecidos para las decisiones basadas únicamente en el tratamiento automatizado. Ahora, el RGPD no da una carta blanca al responsable del tratamiento, transformar una decisión basada únicamente en el tratamiento en una decisión basada en la elaboración de perfiles requiere aumentar *significativamente el nivel de intervención humana de forma que el modelo ya no sea un proceso de toma de decisiones totalmente automatizadas*. Grupo de Trabajo sobre Protección de Datos del Artículo 29, 33. Las dimensiones de esta problemática se abordarán más adelante.

competente y autorizada para modificar la decisión, y para ello ha de realizar una acción significativa y no simbólica»[98].

Ahora bien, no se trata únicamente de superar un concepto formal de intervención humana, sino que resulta indispensable para la seguridad jurídica del responsable del tratamiento poder determinar qué ha de entender por intervención humana significativa.

2.1.1.2. Criterios para entender la intervención humana significativa

En las directrices del GT29 encontramos un primer esfuerzo por delimitar este concepto, aportando algunos elementos comunes para entender la intervención humana significativa.

En primer lugar, enlazando con esa pasividad a la que ya hemos hecho referencia, la aplicación rutinaria de los resultados algorítmicos generados automáticamente sin influencia real por el agente humano que las supervisa, no puede considerarse intervención significativa[99]. Para Veale y Edwards, ello implica considerar la frecuencia con la que el operador está en desacuerdo con el sistema y modifica o mejora los resultados[100]. Es decir, el responsable del tratamiento debe evaluar si las personas que, bajo su autoridad, han sido encomendadas para intervenir en el proceso decisorio, incurren o no en ese sesgo de automatización o, en definitiva, aplican de forma rutinaria dichos resultados sin influencia real en la toma de decisiones.

En segundo lugar, para que la participación humana sea significativa, ésta debe llevarse a cabo por parte de una persona autorizada y competente para modificar la decisión, cuyo análisis debe tener en cuenta todos los datos pertinentes[101]. Brkan considera que tener en cuenta todos los datos pertinentes presenta una gran complejidad en la práctica, exacerbada en contextos en que las decisiones se toman a partir de correlaciones automáticas extraídas en contextos de *Big Data*[102].

Por otro lado, las directrices resaltan una diferencia para la intervención humana como medida de salvaguarda para decisiones basadas únicamente en el tratamiento automatizado; en este caso, el agente humano es calificado como revisor y su análisis debe tener en cuenta cualquier información adicional facilitada por el interesado[103]. Ello pone de manifiesto una vinculación instrumental entre la intervención humana como medida de salvaguarda con los derechos del interesado a una explicación, a expresar su punto de vista e impugnar la decisión[104].

En resumen, las directrices refrendadas por el CEPD definen que la intervención humana significativa debe [a] llevarse a cabo por persona autorizada y competente para modificar la

[98] Agencia Española de Protección de Datos (AEPD), «Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción», 10.
[99] Grupo de Trabajo sobre Protección de Datos del Artículo 29, «Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679», 23. Parece implícita una alusión al sesgo de automatización, sobre la que profundizaremos al hablar del fundamento de la intervención humana en el RGPD. Vid. Apartado 1.2.2. Fundamento para la intervención humana en el artículo 22 RGPD, en este mismo capítulo.
[100] Veale y Edwards, «Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling», 400.
[101] Grupo de Trabajo sobre Protección de Datos del Artículo 29, «Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679», 23 y 30.
[102] En sus propias palabras: *In particular, it remains unclear how a human with limited capacities of data analysis will be able to justify that the final decision needs to be different from an algorithmic one, given that the automated system might not only have taken into account the data relating to the data subject affected by the decision, but a multitude of other complex datasets. If the automated decision was a simple sum of data appertaining to a particular data subject, an in-depth human review of automated decision would be much more feasible. If, however, the decision is based on complex relations between data in a Big Data environment, the human will have a much more difficult task in reviewing such a decision*. Brkan, «Do Algorithms Rule the World? Algorithmic Decision-Making in the Framework of the GDPR and Beyond», 108.
[103] Grupo de Trabajo sobre Protección de Datos del Artículo 29, «Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679», 30.
[104] Almada, «Human Intervention in Automated Decision-making: Toward the Construction of Contestable Systems», 1; Brkan, «Do Algorithms Rule the World? Algorithmic Decision-Making in the Framework of the GDPR and Beyond», 87; Malgieri, «Automated decision-making in the EU Member States: The right to explanation and other "suitable safeguards" in the national legislations», 22.

decisión, [b] realizarse sobre un análisis que tenga en cuenta todos los datos disponibles y [c] no conllevar aplicación rutinaria de los resultados algorítmicos. Cuando esta participación humana se introduzca para el uso sistemas de apoyo a la toma de decisiones [2], dicha intervención significativa debe tener lugar necesariamente antes de la producción de efectos jurídicos o significativos para la persona interesada.

Sin desmerecer esta labor interpretativa realizada por el GT29, cuyos elementos pueden considerarse una digna primera tentativa teórica para abordar el derecho a la intervención humana en el RGPD, no son, desde luego, suficientes para paliar la incertidumbre derivada de la escasa aplicación jurisprudencial[105], así como de la poca atención que la doctrina ha prestado a este elemento en particular[106].

En particular, ha de considerarse un criterio que puede ya advertirse como fundamento de los trabajos de la Comisión Europea en los años noventa para la regulación de la toma de decisiones automatizada, y que dieron lugar a la Directiva 95/46/CE del Parlamento Europeo y del Consejo, de 24 de octubre de 1995, relativa a la protección de las personas físicas en lo que respeta al tratamiento de datos personales y a la libre circulación de estos datos. En dichos trabajos, se manifiesta un temor a la pérdida del control humano en las decisiones automatizadas que uno mismo adopta, como responsable del tratamiento: «*La utilización abusiva de la informática en la toma de decisiones constituye uno de los riesgos esenciales que se plantean en el futuro, ya que el resultado que proporciona la máquina, que utiliza programas cada día más refinados, e incluso sistemas expertos, tiene un carácter aparentemente objetivo e incontestable, al que el encargado de tomar las decisiones puede conceder una importancia excesiva, en dejación de su propia responsabilidad. (…) sólo se prohíbe la aplicación por parte del usuario de los resultados producidos por el sistema. La informática puede servir de ayuda a la decisión, pero no constituir en ningún caso su única base, debiendo dejarse a la apreciación humana el lugar que le corresponde*»[107].

A mi juicio, la relación entre esa pérdida de control y la dejación de la responsabilidad hace referencia inevitablemente al principio de responsabilidad contenido en la normativa europea de protección de datos. Conforme al artículo 5 RGPD, en el tratamiento de datos personales el responsable del tratamiento debe cumplir y demostrar el cumplimiento con el Reglamento y sus principios. Así, el Reglamento introduce la intervención humana como un componente esencial de la toma de decisiones automatizada, con un mecanismo de gobernanza basado en la supervisión humana que impide la dejación de responsabilidad del responsable del tratamiento sobre las decisiones que adopta y afectan significativamente a las personas interesadas.

De hecho, la doctrina ha vinculado la intervención humana con el cumplimiento de varios de los principios del Reglamento, especialmente en lo que se refiere a la licitud y lealtad (art. 5(1)(a)) y a la exactitud (art. 5(1)(d)) en la toma de decisiones. Los humanos se consideran cruciales para evitar correlaciones indebidas y, por tanto, para garantizar la equidad en el tratamiento de datos[108], y no solo para eliminar la discriminación, sino también para reducir los falsos positivos[109]. Por su parte, Brkan sostiene que el artículo 22 del GDPR refleja el escepticismo europeo hacia los sesgos algorítmicos y las decisiones eventualmente erróneas que pueden tomar las máquinas no verificadas por los humanos[110]. En este sentido, para Hoofnagle, van

[105] En aplicación del artículo 22 RGPD, la interpretación más relevante es la Sentencia del TJUE (Sala Primera) de 7 de diciembre de 2023, en el asunto C-634/21 SCHUFA Holding (Scoring) [ECLI:EU:C:2023:957]. No obstante, en dicho caso no se entró a valorar la entidad de la intervención humana exigida por el RGPD, sino simplemente su concurrencia o no de forma previa a la concurrencia de un efecto jurídico o significativo por la realización de un pronóstico sobre la probabilidad de un comportamiento futuro de una persona (score o calificación) en el ámbito bancario.

[106] Huq, «A Right to a Human Decision», 624.

[107] Comisión de las Comunidades Europeas, Propuesta modificada de Directiva « relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos». Bruselas 15.10.1992. COM(92) 422 final- SYN 287, p. 27. Disponible aquí: https://eur-lex.europa.eu/procedure/EN/1990_287

[108] Favaretto, De Clercq, y Elger, «Big Data and discrimination: perils, promises and solutions. A systematic review», 21.

[109] Roig, «Safeguards for the right not to be subject to a decision based solely on automated processing (Article 22 GDPR)», 6.

[110] Brkan, «Do Algorithms Rule the World? Algorithmic Decision-Making in the Framework of the GDPR and Beyond», 97. También vemos esta idea en relación con la regulación de la IA y la inclusión de la supervisión humana como requisito obligatorio. En el Libro Blanco sobre IA se entiende que entiende que la supervisión humana contribuye a garantizar que un sistema de IA no cause efectos adversos, vid. Comisión Europea, Libro Blanco sobre la inteligencia

der Sloot y Zuiderveen Borgesius, el considerando 71 es un mandato explícito para minimizar el riesgo de errores, tanto en términos de exactitud como de los posibles efectos discriminatorios[111]. Y las directrices del GT29 resaltan que los responsables del tratamiento deben tener en cuenta la exactitud en todas las fases del proceso de elaboración de perfiles, inclusive al aplicarlo para tomar una decisión que afecta a una persona[112].

2.1.2.Intervención humana significativa para la toma de decisiones clínica

Como resumen del apartado anterior, podemos considerar que la intervención humana como componente esencial de la toma de decisiones será significativa (1) en tanto se adopte por persona autorizada y competente para modificar el resultado algorítmico; (2) siempre y cuando dicho análisis no conlleve la mera aplicación rutinaria de los resultados algorítmicos, a partir del análisis de todos los datos disponible por el interventor humano; y (3) en la medida en que contribuya a un tratamiento automatizado lícito, leal y exacto en aplicación del principio de responsabilidad.

Ahora bien, la aplicación de este análisis al contexto de la toma de decisiones automatizada en el ámbito clínico precisa de una necesaria profundización sobre estas características o criterios generales de la intervención humana significativa.

Una vez más, debe rebatirse la idea de que los sistemas de apoyo a la toma de decisiones sean un simple peldaño hacia la automatización plena. Quienes implementan sistemas de IA con vistas a proporcionar apoyo a la toma de decisiones (frente a la automatización "plena") deben tener clara la naturaleza del apoyo y cómo se integra en otras tareas, cómo se establece la confianza en ese apoyo y cómo puede afectar al trabajo[113]. Ha de tenerse en cuenta que el personal médico ya asume tareas que implican una compleja toma de decisiones en entornos también complejos y cada vez más dependientes de sistemas tecnológicos[114]. Y que una integración pobre de las TIC en este entorno, no digamos de sistemas de IA, puede tener un impacto negativo sobre el cuidado de la salud y, en particular, sobre la atención que provee el personal sanitario. En este sentido, es clave poder determinar qué ha de considerarse como intervención humana significativa en el contexto particular de la toma de decisiones clínica.

2.1.2.1. Intervención humana con autoridad y competencia para modificar los resultados de los sistemas de IA

En cuanto a los requisitos de autoridad y competencia, esto es, que la intervención humana se lleve a cabo por parte de una persona autorizada y competente para modificar la decisión, su cumplimiento en el ámbito clínico no parece ofrecer grandes dificultades -al menos desde un punto de vista formal-.

Nuestro ordenamiento reconoce en el artículo 3 de la Ley de Autonomía del Paciente la figura del médico responsable: *el profesional que tiene a su cargo coordinar la información y la asistencia sanitaria del paciente o del usuario, con el carácter de interlocutor principal del mismo en todo lo referente a su atención e información durante el proceso asistencial, sin perjuicio de las obligaciones de otros profesionales que participan en las actuaciones asistenciales* (art. 3 de la Ley 41/2002). Entre otros, el médico o médica responsable debe garantizar el cumplimiento del derecho de información de los pacientes (art. 4.3), adoptar decisiones conforme a su criterio cuando el paciente no esté en disposición de hacerlo (art. 9.3.a), ponderar la necesidad de consentimiento escrito por el paciente (art. 10.2), proponer el alta forzosa del paciente (art. 21.1) o emitir el informe de alta médica (art. 3). Es decir, la

artificial, 21. Mientras que en el RIA, el artículo 14(2) establece que la supervisión humana tendrá como objetivo prevenir o minimizar los riesgos para la salud, la seguridad o los derechos fundamentales.

[111] Hoofnagle, van der Sloot, y Borgesius, «The European Union general data protection regulation: what it is and what it means», 92.

[112] Grupo de Trabajo sobre Protección de Datos del Artículo 29, «Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679», 13.

[113] Jamieson y Goldfarb, «Clinical considerations when applying machine learning to decision-support tasks versus automation», 779.

[114] Vid. Sittig y Singh, «A New Sociotechnical Model for Studying Health Information Technology in Complex Adaptive Healthcare Systems».

autoridad del médico o médica responsable sobre el proceso asistencial es muy amplia y la introducción de sistemas de apoyo a la toma de decisiones no pone la misma en entredicho.

En cuanto a la competencia del personal clínico, el cumplimiento de la Ley 44/2003, de 21 de noviembre, de ordenación de las profesiones sanitarias y de las distintas normas que emanan de la misma es, a mi juicio, una garantía suficiente para asegurar -como digo, al menos en abstracto- la competencia de las personas que intervienen en los procesos clínicos para modificar los resultados arrojados por los sistemas automatizados[115].

Aun careciendo de valor jurídico, en la Carta de Derechos Digitales adoptada por el Gobierno de España encontramos una disposición que claramente busca reforzar la autoridad y competencia de los profesionales en el uso de sistemas de IA. El título XXIII sobre el derecho a la protección de la salud en el entorno digital, establece que: *4. El empleo de sistemas digitales de asistencia al diagnóstico, y en particular de procesos basados en inteligencia artificial no limitará el derecho al libre criterio clínico del personal sanitario*. Se trata de un reconocimiento -pionero, a mi juicio- de la libertad de decisión de quien asume el rol de supervisor/interventor humano en la implementación de sistemas de IA, frente a la mayoría de las disposiciones que reconocen esta libertad de decisión más o menos limitada para la persona receptora o afectada por el proceso de toma de decisiones.

En cualquier caso, si observamos la normativa que regula el ejercicio de las profesiones sanitarias y de la que emana, precisamente, la competencia y autoridad a la que hacía referencia arriba, esta disposición puede leerse como un refuerzo del derecho al libre ejercicio de las profesiones sanitarias ante la automatización de los procesos clínicos. El libre ejercicio de las profesiones sanitarias se reconoce primero en la Ley 14/1986, de 25 de abril, General de Sanidad[116], y se desarrolla en la Ley 44/2003, de 21 de noviembre, de ordenación de las profesiones sanitarias, que garantiza un ejercicio con plena autonomía técnica y científica de la profesión bajo determinados principios y valores jurídicos y deontológicos[117].

En definitiva, a pesar de que la mencionada Carta de Derechos Digitales no tenga fuerza jurídica, la libertad de criterio por los profesionales sanitarios en el ámbito de sus competencias ejercida conforme a los principios jurídicos y deontológicos aplicables y únicamente bajo las limitaciones establecidas por ley, sí está reconocida por el ordenamiento y debe garantizarse en el marco de la automatización. En este caso, en relación con el alcance de la intervención humana significativa introducida por el artículo 22 RGPD. Así, el reconocimiento de este derecho de libertad de ejercicio de las profesiones sanitarias es, en primer lugar, un reconocimiento de la autoridad y competencia de los profesionales sanitarios para modificar -y para seguir- el criterio de un sistema de IA.

[115] Cuestión distinta es que para el desarrollo de una competencia efectiva en el uso de esta clase de sistemas de IA sea una formación específica para el personal sanitario. Sobre esta cuestión se volverá más adelante.

[116] Art. 88 LGS: *Se reconoce el derecho al ejercicio libre de las profesiones sanitarias, de acuerdo con lo establecido en los artículos 35 y 36 de la Constitución.*

[117] Art. 4(7) Ley 44/2003, de 21 de noviembre: *El ejercicio de las profesiones sanitarias se llevará a cabo con plena autonomía técnica y científica, sin más limitaciones que las establecidas en esta ley y por los demás principios y valores contenidos en el ordenamiento jurídico y deontológico, y de acuerdo con los siguientes principios: a) Existirá formalización escrita de su trabajo reflejada en una historia clínica que deberá ser común para cada centro y única para cada paciente atendido en él.; La historia clínica tenderá a ser soportada en medios electrónicos y a ser compartida entre profesionales, centros y niveles asistenciales.; b) Se tenderá a la unificación de los criterios de actuación, que estarán basados en la evidencia científica y en los medios disponibles y soportados en guías y protocolos de práctica clínica y asistencial. Los protocolos deberán ser utilizados de forma orientativa, como guía de decisión para todos los profesionales de un equipo, y serán regularmente actualizados con la participación de aquellos que los deben aplicar.; c) La eficacia organizativa de los servicios, secciones y equipos, o unidades asistenciales equivalentes sea cual sea su denominación, requerirá la existencia escrita de normas de funcionamiento interno y la definición de objetivos y funciones tanto generales como específicas para cada miembro del mismo, así como la cumplimentación por parte de los profesionales de la documentación asistencial, informativa o estadística que determine el centro.; d) La continuidad asistencial de los pacientes, tanto la de aquellos que sean atendidos por distintos profesionales y especialistas dentro del mismo centro como la de quienes lo sean en diferentes niveles, requerirá en cada ámbito asistencial la existencia de procedimientos, protocolos de elaboración conjunta e indicadores para asegurar esta finalidad.; e) La progresiva consideración de la interdisciplinariedad y multidisciplinariedad de los equipos profesionales en la atención sanitaria.*

2.1.2.2. Intervención humana que no conlleve la aplicación rutinaria de los resultados de los sistemas de IA

La introducción de un sistema de IA de apoyo a la toma de decisiones puede producir un impacto sobre los procesos de decisión clínicos que se traduzcan en una aplicación rutinaria de los resultados automatizados, lo cual equivaldría a la utilización de un sistema plenamente automatizado de facto.

Las directrices refrendadas por el CEPD indican que la *aplicación rutinaria de perfiles generados automáticamente a personas sin que ello tenga influencia real alguna en el resultado*[118]. Aunque no encontramos desarrollo alguno que nos permita conocer qué podemos interpretar por 'aplicación rutinaria' o 'influencia real', dado el contexto parece evidente que hace referencia a la influencia del sesgo de automatización en el supervisor o interventor humano.

El sesgo de automatización consiste en que la personas o personas que utilizan un sistema algorítmico como apoyo a la toma de decisiones, otorgan más peso del que corresponde a las predicciones o resultados del sistema automatizado[119]. Se trata de un sesgo cognitivo que afecta a la toma de decisiones humana de forma muy similares a otra clase de sesgos, como puede ser el sesgo de confirmación, pero al que se otorga una categoría particular por derivar de forma específica de la interacción de la persona con un sistema automatizado[120]. Sus efectos serán inocuos cuando la recomendación algorítmica sea correcta, pero podrán derivar en un perjuicio o daño cuando sea incorrecta. La evidencia que recogen Parasuraman y Manzey señala que hay tres razones fundamentales para que las personas sobreestimen las capacidades del sistema en el proceso decisorio: (1) la tendencia humana a elegir el camino de menor esfuerzo cognitivo en la toma de decisiones; (2) la percepción del sistema algorítmico como un ente objetivo dotado de autoridad; y (3) la difusión del sentido de la responsabilidad que afecta a los seres humanos cuando trabajan en grupo o conjuntamente[121].

Ahora bien, si aspiramos al desarrollo de sistemas fiables y que tengan una alta capacidad para adoptar las decisiones correctas en el contexto dado, ¿sería compatible un alto grado de coincidencia con el criterio algorítmico con una aplicación no rutinaria del mismo? A mi modo de ver, la respuesta es necesariamente afirmativa. La problemática, cuando el propio ordenamiento exige que estos sistemas sean de alta calidad[122], está en cómo determinar si un operador humano se está viendo afectado por un sesgo de automatización que conlleve una aplicación rutinaria y sin influencia real del sistema.

A mi modo de ver, optar por tratar de identificar los sesgos de automatización observando únicamente los resultados en el proceso decisorio, y en particular evaluando si los perjuicios producidos son significativos, puede comprometer la toma de decisiones futura. De este modo, el sesgo de automatización que en un contexto particular no produce un resultado nocivo -puesto que lleva a la confirmación de un resultado algorítmico acertado-, puede transformarse a medio plazo en un problema agravado si dicho contexto varía y la persona no es capaz de realizar sus funciones de intervención, supervisión o revisión correctamente. Por esa misma razón, entiendo que la influencia de ese sesgo de automatización -es decir, la determinación de si la intervención humana conlleva una *aplicación rutinaria sin influencia real*- debe evaluarse de forma continua en el proceso de toma de decisiones.

Así, podemos distinguir varios factores que pueden influir en la capacidad humana para influir sobre las decisiones de los sistemas de IA en el ámbito clínico. Si centramos nuestra atención en cómo es el proceso decisorio, el tiempo es un factor esencial en las posibilidades de deliberación y de reflexión para el operador humano en la toma de decisiones, y así, la falta de tiempo en el proceso de toma de decisiones puede reforzar los sesgos de automatización y

[118] Grupo de Trabajo sobre Protección de Datos del Artículo 29, «Directrices sobre la evaluación de impacto relativa a la protección de datos (EIPD) y para determinar si el tratamiento "entraña probablemente un alto riesgo" a efectos del Reglamento (UE) 2016/679», 23.
[119] Challen et al., «Artificial intelligence, bias and clinical safety», 234.
[120] Parasuraman y Manzey, «Complacency and Bias in Human Use of Automation: An Attentional Integration», 391.
[121] Parasuraman y Manzey, 391-92.
[122] Aplicación del principio de exactitud al tratamiento de datos en la aplicación del RGPD y de los criterios de calidad y seguridad en la aplicación de la normativa europea de productos sanitarios, en este caso.

limitar la capacidad de influencia de la persona sobre el sistema automatizado. Partamos de la constricción temporal del uso de estas tecnologías en los procesos deliberativos clínicos. Por el momento, es bien sabido que las tecnologías *eHealth* consumen mucho tiempo y recursos[123]. Estudios muestran que en atención primaria se dedica casi un 40% del tiempo de la consulta al sistema informático[124], o que el conjunto de tareas realizadas frente al ordenador supone para el personal médico más tiempo que las dedicadas a atender pacientes[125] y, lo que resulta aún más preocupante, la aparente relación entre el estrés y agotamiento de los y las profesionales con el uso de las tecnologías de la información sanitarias[126].

Por todo ello, será necesario que el responsable del tratamiento evalúe no solo el tiempo del que dispone el personal médico en el proceso clínico para la intervención en la toma de decisiones, sino también el tiempo que consume el uso adecuado del sistema automatizado.

Otro de los aspectos que no puede perderse de vista es la información disponible para la persona a la hora de intervenir en el proceso decisorio. Fritz expone que los datos de los pacientes suelen ser el centro del debate sobre la digitalización de la atención sanitaria, sin embargo, este enfoque en los datos tiende a pasar por alto que los datos sólo representan una parte -importante, eso sí- del contexto clínico[127]. Precisamente, uno de los efectos del sesgo de automatización es que puede llevar a decisiones que no se basan en un análisis exhaustivo de toda la información disponible para el operador humano. En este sentido, entiendo que el concepto de *rutinario* en la práctica clínica -y en otros contextos- puede vincularse con otro de los mandatos de las directrices del GT29 para la intervención humana: el análisis debe *tener en cuenta todos los datos pertinentes*[128]. Si hacemos una interpretación amplia de lo que el GT29 quiere decir al hablarnos de considerar todos los *datos disponibles*, puede entenderse que esos datos no son exclusivamente los que sirven como datos de entrada para el sistema de IA, sino el conjunto de datos e información presente en el contexto decisorio y que trasciende la *datificación* llevada a cabo por el sistema[129].

Es decir, el tiempo efectivo disponible en el proceso de toma de decisiones y la información considerada en el mismo son factores determinantes -aunque no los únicos[130]- a la hora de evaluar si se produce, o no, una aplicación rutinaria del sistema automatizado. A mi modo de ver, una respuesta razonable para el control de la influencia de estos factores y su demostración es la protocolización de estos procesos de toma de decisiones y la evaluación periódica de dichos protocolos. Por ello, más adelante, propongo seguir esta vía utilizando la EIPD como instrumento jurídico adecuado.

[123] Granja, Janssen, y Johansen, «Factors determining the success and failure of ehealth interventions: Systematic review of the literature».
[124] Pérez-Santonja et al., «Historia clínica electrónica : evolución de la relación médico-paciente en la consulta de Atención Primaria».
[125] Tai-Seale et al., «Electronic Health Record Logs Indicate That Physicians Split Time Evenly Between Seeing Patients And Desktop Medicine».
[126] Gardner et al., «Physician stress and burnout: the impact of health information technology».
[127] Fritz, «La transformación digital en la atención sanitaria como un reto para la autonomía y la confianza en la interacción médico-paciente», 29.
[128] Grupo de Trabajo sobre Protección de Datos del Artículo 29, «Directrices sobre la evaluación de impacto relativa a la protección de datos (EIPD) y para determinar si el tratamiento "entraña probablemente un alto riesgo" a efectos del Reglamento (UE) 2016/679», 23 y 30.
[129] En ocasiones, es posible que las explicaciones generadas por los propios sistemas de IA -en los casos en los que los sistemas están diseñados para aportar explicaciones al operador humano sobre las correlaciones realizadas para alcanzar determinado resultado-
[130] . También en esta línea, una cuestión fundamental sobre la que debe ampliarse el estudio jurídico es la relación entre la intervención/supervisión humana como mecanismo de gobernanza para sistemas de IA de alto riesgo y el régimen de responsabilidad civil y penal sanitaria derivado del uso de sistemas de IA. Es decir, en los procesos de deliberación clínicos no es únicamente el propio sistema algorítmico el que puede producir un impacto sobre la intervención humana, también hay otros aspectos sistémicos que pueden resultar en una aplicación rutinaria de los resultados automatizados. En el caso del contexto clínico, esta aplicación rutinaria fundamentada en evitar la asunción de responsabilidades jurídicas derivadas de la toma de decisiones, derivará en gran medida en prácticas de medicina defensiva. Sobre este tema, Verdicchio y Perin expresan que es necesario diseñar un modelo de distribución y asignación de responsabilidades coherente con la necesaria intervención médica humana, pero que también permita aprovechar las ventajas que ofrecen los sistemas de IA en medicina, es decir, la eficiencia y la reducción de costes. Vid. Verdicchio y Perin, «When Doctors and AI Interact: on Human Responsibility for Artificial Risks», 16; también Perin, «Estandarización y automatización en medicina: El deber de cuidado del profesional entre la legítima confianza y la debida prudencia».

2.1.2.3. Intervención humana que favorezca el cumplimiento normativo y la demostración de su cumplimiento

Por último, los trabajos preparatorios de la Comisión de 1992 sobre la Directiva de Protección de Datos señalaban el temor de que una utilización excesiva de la informática pudiera derivar en una dejación de la responsabilidad por parte de quien implementase sistemas de tratamiento de datos automatizados[131]. Así, hemos podido vincular la intervención humana como mecanismo de gobernanza con el principio de responsabilidad del RGPD, en virtud del cual el responsable del tratamiento debe cumplir y demostrar el cumplimiento del RGPD y sus principios. Y en particular, podemos encontrar en el Considerando 71 un mandato explícito para minimizar el riesgo de errores a través de la intervención humana, tanto en términos de exactitud como de los posibles efectos discriminatorios[132].

En definitiva, la intervención humana significativa en el RGPD ha de contribuir favorablemente al cumplimiento de éste y de la demostración de dicho cumplimiento.

Ha de tenerse en cuenta que, en el contexto clínico, la toma de decisiones apoyada en sistemas de tratamiento automatizado de datos ha de responder a criterios normativos que trascienden de los principios RGPD de licitud y lealtad (art. 5(1)(a)) y de exactitud (art. 5(1)(d)). Arriba se menciona el derecho al libre ejercicio de las profesiones sanitarias en la toma de decisiones, que ejercicio con plena autonomía técnica y científica de la profesión bajo determinados principios y valores jurídicos y deontológicos. Libertad de criterio que, además, pretende ser reforzada por la Carta de Derechos Digitales respecto del uso de sistemas digitales de asistencia al diagnóstico, y en particular de procesos basados en inteligencia artificial. Por ello, la intervención humana en el contexto clínico responderá también a criterios como la *lex artis ad hoc*, que el profesional debe determinar conforme a su experiencia: «*criterio valorativo de la corrección del concreto acto médico ejecutado por el profesional de la medicina -ciencia o arte médica- que tienen en cuenta las especiales características de su autor, de la profesión, de la complejidad y trascendencia vital del paciente y en su caso, de la influencia de otros factores endógenos (…), para calificar dicho acto de conforme o no con la técnica normal requerida (derivando de ello tanto el acervo de exigencias o requisitos de legitimación o actuación lícita, de la correspondiente eficacia de los servicios prestados, y, en particular, de la posible responsabilidad de su autor/médico por el resultado de su intervención o acto médico ejecutado)*» [133].

Ahora bien, si tenemos en cuenta que los sistemas de IA utilizados en este contexto estarán debidamente certificados como productos sanitarios conforme a la normativa aplicable[134], el acto médico adecuado será, por lo general y siempre que el personal sanitario lo utilice en el contexto pertinente, el recomendado por el criterio algorítmico. Una vez más, es necesario considerar el mandato de las directrices del GT29, conforme al cual la intervención humana debe tener en cuenta todos los datos pertinentes.

No podemos obviar que determinar la exactitud derivada del tratamiento de datos en el cuidado clínico puede ser muy complejo -por la propia incertidumbre que rodea la aplicación de la ciencia médica- o incluso indiferente -pongamos, por el ejercicio de la autonomía de una voluntad contraria a la recomendada por el criterio clínico-. Más allá de los casos en los que el ejercicio del derecho fundamental a la integridad (art. 15 CE) deba dejar sin efecto cualquier decisión clínica -sea ésta adoptada por una persona, por un sistema de IA o por ambos conjuntamente-; por ejemplo, en la adopción de decisiones clínicas en el final de la vida, la *lex artis ad hoc* impone al facultativo la supresión de todo tratamiento o medida asistencial: «*que implique de alguna manera una forma de la llamada "obstinación terapéutica" (o "encarnizamiento terapéutico") o la prolongación de la existencia biológica sin expectativas previsibles de recuperación*» [135]. Imaginemos que se utiliza un sistema de IA que recomienda la

[131] Comisión de las Comunidades Europeas, Propuesta modificada de Directiva «relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos», 27.
[132] Hoofnagle, van der Sloot, y Borgesius, «The European Union general data protection regulation: what it is and what it means», 92.
[133] Fundamento de Derecho segundo, Sentencia del Tribunal Supremo de 11 de marzo de 1991, núm. rec. 245/1987 (Roj: STS 13345/1991).
[134] Vid. sobre esta cuestión más adelante.
[135] Romeo Casabona y Arruego, «Lección 14. Toma de decisiones al final de la vida», 374.

mejor opción terapéutica posible, pero que no es capaz de considerar alternativas no terapéuticas en función del estado del paciente. En este caso, considerando *todos los datos pertinentes* -entendiendo por ello el contexto clínico en su conjunto y la situación personal del paciente en particular, considerando incluso aspectos como la vulnerabilidad[136]-, la intervención humana debería dejar sin efecto la propuesta terapéutica ofrecida por el sistema de IA.

En este sentido, una intervención humana significativa en el contexto clínico debe ser aquélla que contribuya al cumplimiento normativo y a su demostración en sentido amplio, considerando las normas y principios que rigen el cuidado clínico en su conjunto y no únicamente los principios del RGPD. Para asegurar que los distintos profesionales sanitarios intervienen respecto del sistema automatizado con criterios uniformes y en consideración de los datos relevantes para el contexto clínico en el que se implemente, podría resultar muy útil la protocolización del uso de estas tecnologías[137]. Además, el protocolo podría ser de utilidad para el cumplimiento de las obligaciones del responsable del tratamiento para la gestión de riesgos y, en particular, en la realización de la EIPD conforme al artículo 35 RGPD; dado que el protocolo puede servir para justificar la forma en la que la intervención humana se garantiza en el uso del sistema automatizado o de IA.

2.2. Intervención humana como medida organizativa desde el principio de responsabilidad

Tal y como he tenido ocasión de señalar anteriormente, la ubicación entre los derechos de la persona interesada de la prohibición de la toma de decisiones basada únicamente en el tratamiento automatizado o de los derechos de información y acceso, no debe inducirnos a pensar que el RGPD regula fundamentalmente la toma de decisiones automatizada a partir de dicho reconocimiento y ejercicio de derechos que permite el control individual de los datos personales. En su momento, la Directiva 95/46/CE de 24 de octubre de 1995, comenzó a desplazar el foco sobre el control del interesado acerca de la información a él que se revela o deja de revelar, hacia un control colectivo sobre la justificación de los procesos que recopilan, procesan y usan los datos[138]. Y el RGPD no hizo sino reforzar este control colectivo basándose en la responsabilidad del cumplimiento normativo[139].

Uno de los cambios normativos más relevantes entre la Directiva de 1995 y el Reglamento de 2016 es la introducción de la "*accountability*" como principio nuclear de la normativa de protección de datos[140]. Sobre el concepto de *accountability*, la clave no está en la imposición de sanciones -de forma exclusiva-, sino en el proceso de rendición de cuentas[141]. Para que este proceso de rendición de cuentas resulte efectivo, la información que se obliga a aportar al responsable sobre las operaciones de tratamiento de datos personales que realiza debe ser suficiente para determinar el cumplimiento normativo y, en su caso, permitir la imposición de las consecuencias previstas en caso de incumplimiento –como recursos legales o intervenciones para corregir el mal funcionamiento del proceso, pensemos en el derecho de rectificación, incluyendo también la posibilidad del establecimiento de sanciones, por supuesto, que corresponde a las autoridades-[142].

El proceso de rendición de cuentas concreto que establece el RGPD es el siguiente.

[136] Tanto desde la perspectiva de qué puede hacerse con los datos de estas personas, como a la hora de comprender cómo puede afectar los resultados de determinado tratamiento a las mismas, existen diferentes grupos poblacionales vulnerables. Vid. Malgieri y Niklas, «Vulnerable data subjects».
[137] El protocolo de actuación médica guarda similitudes con la lex artis ad hoc, aunque ésta debe siempre determinarse en cada caso, no pudiendo el protocolo agotar el contenido de la misma. Vid. Vázquez López, «La "Lex Artis ad hoc" como criterio valorativo para calibrar la diligencia exigible en todo acto o tratamiento médico: A propósito de un caso basado en la elección de la técnica empleada en el parto (parto vaginal vs. cesárea) », 182.
[138] Turégano Mansilla, «Los valores detrás de la privacidad», 276.
[139] Aunque no en todos los aspectos, Mayer-Schönberger destaca que el enfásis sobre el consentimiento como base legitimadora del tratamiento en el RGPD es un paso atrás en este sentido. Vid. Mayer-Schönberger, «Paradigm shift».
[140] Traducido en la versión española como "principio de responsabilidad proactiva", en este texto me referiré al principio de responsabilidad.
[141] Más ampliamente sobre la responsabilidad como mecanismo de gobernanza y su encaje en la normativa de protección de datos, vid. Hert y Lazcoz, «When GDPR-principles blind each other. Accountability, not transparency, at the heart of algorithmic governance».
[142] Vid. Cobbe, Lee, y Singh, «Reviewable Automated Decision-Making: A Framework for Accountable Algorithmic Systems».

En el artículo 5(2) RGPD el principio de responsabilidad constituye un principio nuclear que obliga a cumplir y demostrar el cumplimiento de los otros seis principios enumerados en el artículo 5(1) RGPD, esto es, el principio de licitud, lealtad y transparencia, el principio de limitación de la finalidad, el principio de minimización de datos, el principio de exactitud, el principio de limitación del plazo de conservación y el principio de integridad y confidencialidad. En desarrollo del principio de responsabilidad, el artículo 24 RGPD exige que el responsable del tratamiento tenga en cuenta: «*la naturaleza, el ámbito, el contexto y los fines del tratamiento*, así como *los riesgos de diversa probabilidad y gravedad para los derechos y libertades de las personas físicas* y que, en consecuencia, aplique *medidas técnicas y organizativas apropiadas*», a fin de garantizar y demostrar el cumplimiento normativo, debiéndose documentar dichas medidas de forma adecuada[143].

Así, dicho principio de responsabilidad construye una gobernanza mixta de la protección de datos. Por un lado, establece unos principios tasados a los que el responsable debe adherirse en el tratamiento de datos personales, si bien y, por otro lado, el modo o forma de cumplimiento de dichos objetivos normativos depende en gran medida de la propia autorregulación que establezca el responsable.

Estamos ante un modelo normativo que, tal y como reflejan distintos artículos y considerandos al referirse a "medidas adecuadas" o "apropiadas", no delimita tanto medidas o niveles de seguridad concretos, como hacía la normativa anterior, sino que parte de la idea de que cada tratamiento concreto precisará de unas medidas de seguridad que se adapten al mismo[144].

No obstante, en algunos casos, el RGPD impone determinadas medidas técnicas u organizativas debido al riesgo que plantea el tratamiento: por ejemplo, en el caso de la toma de decisiones basada únicamente en el tratamiento automatizado en virtud de las excepciones contractuales o basadas en el consentimiento (22(2)(a) y (c) RGPD), las medidas obligatorias son los derechos del interesado a obtener la intervención humana por parte del responsable del tratamiento, a expresar su punto de vista y a impugnar la decisión (22(3) RGPD). La obligación de llevar a cabo una EIPD para tratamientos que entrañen un "alto riesgo" para los derechos y libertades de las personas físicas (35 (1) RGPD), también se impone en este modelo normativo. En otros casos, el RGPD establece o sugiere una serie de medidas que puede adoptar el responsable del tratamiento cuando considere que son adecuadas por la naturaleza, el alcance, el contexto y los fines del tratamiento y sus riesgos para los derechos y libertades de las personas físicas. Por ejemplo, para garantizar la seguridad del tratamiento, el RGPD recomienda, entre otras cosas, la seudonimización y el cifrado de los datos personales (32(1)(a) RGPD)[145]. Por último, esto no cierra la puerta a otras medidas que, si bien no son requeridas o recomendadas directamente por el RGPD, pueden ser consideradas apropiadas por el responsable del tratamiento para cumplir y demostrar el cumplimiento de la ley, en función del proceso de decisiones que prefiera diseñar. Para evitar la prohibición del artículo 22, apartado 1, el responsable del tratamiento puede aumentar significativamente el nivel de intervención humana de modo que el modelo ya no sea un proceso basado únicamente en el tratamiento automatizado, generalmente prohibido por este artículo[146].

En cierto sentido, y dentro de los límites que marca la propia normativa, no puede imponerse al responsable del tratamiento una forma concreta de cumplir con el RGPD. Sin embargo, esto no relaja los requisitos del mismo para llevar a cabo un tratamiento justo, lícito y transparente y para garantizar el ejercicio de los derechos individuales de los interesados. El RGPD obliga a una gestión integral de los riesgos en el tratamiento de datos, lo cual no quiere decir que la

[143] Documentar la gestión del riesgo en el tratamiento de datos personales tiene un doble objetivo según la AEPD: en primer lugar, y siendo su objetivo más importante, dar soporte a la ejecución eficaz y eficiente de la gestión del riesgo para los derechos y libertades. En segundo lugar, y supeditado al primero, permitir demostrar que así se ha realizado. Además, la documentación de la gestión del riesgo no es un documento en sí, sino un proceso que se traduce en hechos y que se acredita documentalmente. Agencia Española de Protección de Datos (AEPD), «Gestión del riesgo y evaluación de impacto en tratamientos de datos personales», 12 y 56.

[144] Casanova Asencio, «Mecanismos de prevención del acceso indebido a la historia clínica por parte del personal sanitario y nueva legislación de protección de datos», 5.

[145] También se hace mención expresa a la seudonimización como medida apropiada para el tratamiento con fines de archivo en interés público, fines de investigación científica o histórica o fines estadísticos en el artículo 89 RGPD.

[146] Así lo contempla Grupo de Trabajo sobre Protección de Datos del Artículo 29, «Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679», 33.

realización de dicha gestión esté en todo caso sometida a las mismas obligaciones jurídicas[147]. Según la AEPD, los mayores perjuicios no se originarán de tratamientos de alto riesgo cuando éstos se encuentren bien gestionados, sino de los tratamientos mal gestionados, en los que se ignoran las amenazas y la gravedad de sus consecuencias[148]. Además, la gestión de este riesgo a través de la aplicación de las medidas oportunas y adecuadas al tratamiento de datos, exige no solo demostrar la conformidad de las actividades sino también la *eficacia* de las medidas aplicadas, conforme al considerando 74 RGPD[149].

Este sistema de gestión de riesgos basado en la responsabilidad como principio nuclear del RGPD se complementa, a su vez, con las posibilidades de vigilancia y rendición de cuentas dadas por la normativa a las personas interesadas y a las autoridades, respectivamente[150]. Por un lado, con el reconocimiento de una serie de derechos individuales que permiten a las personas interesadas, dentro de los límites del ejercicio de los mismos, examinar dicho sistema de gestión de riesgos respecto de -y exclusivamente de- sus datos personales[151], pudiendo incluso exigir la rectificación individualmente y de forma directa al responsable del tratamiento en determinados casos[152]. Por otro lado, con un sistema de vigilancia pública que recae fundamentalmente sobre la figura de las autoridades de control, cuyas competencias, funciones y poderes se establecen en el capítulo VI del RGPD, y que trata de garantizar la rendición de cuentas ante posibles infracciones normativas a través de posibilitar el acceso a recursos ante autoridades públicas -sobre todo de las de control, pero sin cerrar la puerta a otras autoridades que pudieren ser competentes y en particular autoridades judiciales- y sanciones en el capítulo VIII RGPD.

2.2.1. Obligación de realizar la EIPD para la implementación de sistemas de IA con fines de salud para la toma de decisiones clínica

La EIPD es un proceso utilizado para reforzar y demostrar el cumplimiento con el Reglamento[153]. Es, por tanto, un instrumento básico para la rendición de cuentas establecida por el principio nuclear de la responsabilidad: «*que ayudan a los responsables no solo a cumplir los requisitos del RGPD, sino también a demostrar que se han tomado medidas adecuadas para garantizar el cumplimiento del Reglamento*». La necesaria determinación de las medidas necesarias para afrontar los riesgos identificados -art. 35(7)(d) RPGD-, en palabras de la AEPD, obliga al responsable a actuar y tiene una dimensión mayor que un mero formalismo plasmado en un documento sobre el que se puedan realizar cambios mínimos para adaptarlo a cualquier tratamiento[154].

[147] Muestra de ello es la EIPD. La gestión del riesgo y la evaluación de impacto para la protección de datos son procesos íntimamente vinculados, no obstante, mientras que la gestión del riesgo es obligatoria para todo tratamiento, las obligaciones concretas que se establecen para la EIPD son obligatorias, exclusivamente, para tratamientos de alto riesgo. Agencia Española de Protección de Datos (AEPD), «Gestión del riesgo y evaluación de impacto en tratamientos de datos personales», 159.

[148] Agencia Española de Protección de Datos (AEPD), 11.

[149] Esta necesidad de demostrar la eficacia de las medidas nos acerca, como veremos más adelante, al enfoque basado en la evidencia que se ha demostrado necesario para asegurar una intervención humana significativa.

[150] Debe destacarse la importancia del delegado de protección de datos en este sistema de vigilancia y rendición de cuentas para los casos en los que su designación es obligatoria en virtud del art. 37(1) RGPD. Corresponde al delegado tanto la comunicación con los interesados a efectos del ejercicio de sus derechos individuales -art. 38(4) RGPD-, como la cooperación y actuación como punto de contacto con la autoridad de control pertinente -art. 38(1)(d) y (e) RGPD-.

[151] Aquí debería considerarse el alcance de los derechos de información y acceso en el marco de la toma de decisiones automatizada basada en la elaboración de perfiles, operando la transparencia como un requisito instrumental de la rendición de cuentas establecida por el RGPD.

[152] El RGPD permite el ejercicio del derecho de rectificación sobre los datos personales, e incluye como medida de salvaguarda de las decisiones basadas únicamente en el tratamiento automatizado el derecho a impugnar las mismas.

[153] Grupo de Trabajo sobre Protección de Datos del Artículo 29, «Directrices sobre la evaluación de impacto relativa a la protección de datos (EIPD) y para determinar si el tratamiento "entraña probablemente un alto riesgo" a efectos del Reglamento (UE) 2016/679», 4. A pesar de que el RGPD no impone una metodología concreta para realizar la EIPD, sí establece un contenido mínimo en su artículo 35(7), según Palma Ortigosa con dos objetivos*: El objetivo que se pretende con este contenido mínimo es que, por una lado se logre una cierta estandarización a la hora de llevar a cabo la EIPD por parte de distintas organizaciones y, por otro lado, se consiga una aplicación efectiva de esta medida ya que se fijan los parámetros mínimos que el legislador considera necesarios para conseguir una protección adecuada de los derechos y libertades de los interesados*. Palma Ortigosa, «Régimen jurídico de la toma de decisiones automatizadas y el uso de sistemas de inteligencia artificial en el marco del derecho a la protección de datos personales», 231.

[154] Agencia Española de Protección de Datos (AEPD), «Gestión del riesgo y evaluación de impacto en tratamientos de datos personales», 25.

Es necesario recalcar que la gestión de riesgos en el RGPD no está orientada a proteger intereses propios del responsable, sino a la protección de la persona, en su dimensión individual y social, como sujeto de los datos o afectado por el tratamiento[155]. Es decir, y en el ámbito que nos concierne, la evaluación de impacto para reforzar y demostrar el cumplimiento normativo se debe realizar desde la perspectiva de protección de los derechos y libertades de los afectados por la toma de decisiones automatizada.

Con carácter general, la EIPD es un instrumento normativo de autoevaluación[156], cuyo valor reside en conducir a la construcción de mejores sistemas en su conjunto[157], para la protección de los derechos y libertades mencionados y, en segunda instancia, para proporcionar un marco de seguridad jurídica al responsable basado en el enfoque de riesgos y la autorregulación[158]. En este sentido, el legislador es consciente de que el riesgo 0 no existe ni resulta exigible[159], y por ello la razón de ser de la EIPD reside en demostrar cómo esos riesgos son mitigados durante el ciclo de vida del tratamiento de datos personales, cómo el responsable del tratamiento ha construido un sistema de tratamiento de datos que cumple con la normativa.

Considerando ya el objeto de estudio de esta investigación, el ecosistema normativo del RGPD basado en el principio de responsabilidad no parece ofrecer dudas a la hora de determinar la obligatoriedad de realizar la EIPD. Es decir, la implementación de sistemas de IA con fines de salud que produzcan efectos jurídicos o significativos en los interesados exige la realización de la EIPD por el responsable del tratamiento. Las razones son las siguientes.

Una vez más, la EIPD debe realizarse con carácter obligatorio para todo tratamiento que probablemente entrañe un alto riesgo para los derechos y libertades de las personas[160]. Además de los criterios establecidos en el articulado del RGPD, para determinar si el tratamiento entraña esta clase de riesgo, el CEPD considera que el responsable del tratamiento debe comprobar los criterios establecidos en los considerandos 71, 75 y 91 del RGPD, en las Directrices del GT29 sobre la evaluación de impacto relativa a la protección de datos (EIPD) y para determinar si el tratamiento «entraña probablemente un alto riesgo» a efectos del Reglamento[161] y la lista de operaciones de tratamiento que pueden entrañar «un alto riesgo» adoptada a nivel nacional en virtud del artículo 35, apartado 4, que la AEPD ha establecido en su Guía para la gestión del riesgo y evaluación de impacto en tratamientos de datos personales de 2021[162]. En sucesivas ocasiones, el CEPD ha reiterado que el responsable debe considerar obligatoria la realización de la EIPD cuando se cumplan dos de los criterios establecidos, si bien, el cumplimiento de uno solo de ellos puede, en algunos casos, llevar a la misma obligación[163].

---

[155] Agencia Española de Protección de Datos (AEPD), 17.

[156] Mantelero, «AI and Big Data: A blueprint for a human rights, social and ethical impact assessment», 768; Hawath, «Regulating Automated Decision-Making: An Analysis of Control over Processing and Additional Safeguards in Article 22 of the GDPR.», 171. No existe obligación jurídica para su publicación, aunque ésta sea recomendable al menos en parte, vid. Grupo de Trabajo sobre Protección de Datos del Artículo 29, «Directrices sobre la evaluación de impacto relativa a la protección de datos (EIPD) y para determinar si el tratamiento "entraña probablemente un alto riesgo" a efectos del Reglamento (UE) 2016/679», 20. No obstante, dados los requisitos documentales y de registro que se establecen, la EIPD sido considerada una "precursora" de la aplicación normativa por parte de las autoridades de control establecidas de acuerdo al artículo 51 RGPD, en Kaminski y Malgieri, «Algorithmic impact assessments under the GDPR: producing multi-layered explanations», 132.

[157] Edwards y Veale, «Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?», 51.

[158] Agencia Española de Protección de Datos (AEPD), «Gestión del riesgo y evaluación de impacto en tratamientos de datos personales», 151.

[159] En palabra de la AEPD: *Siempre existirá un riesgo inherente o inicial implícito en cualquier tratamiento y, una vez que se hayan aplicado medidas y garantías que lo minimicen, seguirá existiendo un riesgo residual*. Agencia Española de Protección de Datos (AEPD), 20.

[160] Artículo 35(1) RGPD: *Cuando 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*.

[161] Grupo de Trabajo sobre Protección de Datos del Artículo 29, «Directrices sobre la evaluación de impacto relativa a la protección de datos (EIPD) y para determinar si el tratamiento "entraña probablemente un alto riesgo" a efectos del Reglamento (UE) 2016/679», 10 y ss.

[162] Agencia Española de Protección de Datos (AEPD), «Gestión del riesgo y evaluación de impacto en tratamientos de datos personales», 136 y ss.

[163] Comité Europeo de Protección de Datos (CEPD), «Opinion 12/2019 on the draft list of the com supervisory authority of Spain petent regarding the processing operations protection exempt from the requirement of a data impact assessment (Article 35 (5) GDPR)», 4.

OBLIGACIÓN DE REALIZAR LA EIPD
"Cuando 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"[114]
Está dentro de alguno de los supuestos establecidos en el artículo 35.3 del RGPD
Existe una norma especial que exige una EIPD para el tratamiento.
Cuando el tratamiento corresponde con alguno de los ejemplos de obligación enumerados en las Directrices WP248.
Cuando el tratamiento cumple al menos dos de las condiciones de las enumeradas en las Directrices WP248 para realizar una EIPD.
Cuando el tratamiento cumpla con dos o más criterios de las *Listas de tipos de tratamientos de datos que requieren evaluación de impacto relativa a protección de datos (art 35.4)* publicada por la AEPD.
Cuando se haya apreciado un alto riesgo teniendo en cuenta los supuestos enumerados en el artículo 28.2 de la LOPDGDD.
Cuando en alguna de las directrices publicadas por el CEPD, el tratamiento esté identificado como obligado a realizar una EIPD.
El tratamiento se encuentre sujeto a un código de conducta o a un mecanismo de certificación que exijan al responsable la realización de una evaluación de impacto.

Tabla 2. Guía AEPD 2021, p. 134

Por un lado, los usos del *big data* y la IA son claros candidatos a que dicha evaluación de impacto sea obligatoria[164]. No solo porque el uso de nuevas soluciones tecnológicas está, por sí, entre los criterios definidos, sino porque características habituales de estas soluciones tecnológicas también se encuentran entre dichos criterios, como el tratamiento de datos a gran escala o la asociación o combinación de conjuntos de datos[165].

Por otro lado, al margen de que la propia toma de decisiones automatizada con efectos jurídicos o similares es uno de los criterios a considerar por entrañar un alto riesgo[166], el artículo 35 RGPD no distingue entre toma de decisiones basada únicamente en el tratamiento automatizado y la toma de decisiones con previa intervención humana[167], tal y como recoge también el GT29 en sus directrices sobre decisiones individuales automatizadas y elaboración

[164] Cotino, «Riesgos e impactos del big data, la inteligencia artificial y la robótica. Enfoques, modelos y principios de la respuesta del Derecho», 29. Más crítica con el margen dado en este sentido por el RGPD se muestra Soriano: *el propio concepto de "alto riesgo" proporciona a responsables y encargados del tratamiento cierto margen de discrecionalidad para decidir en qué casos deben llevarse a cabo las evaluaciones de impacto*. Soriano Arnanz, «La propuesta de Reglamento de inteligencia artificial de la UE y los sistemas de alto riesgo», 115.

[165] Grupo de Trabajo sobre Protección de Datos del Artículo 29, «Directrices sobre la evaluación de impacto relativa a la protección de datos (EIPD) y para determinar si el tratamiento "entraña probablemente un alto riesgo" a efectos del Reglamento (UE) 2016/679», 11-12. También Considerando 91 RGPD: *Lo anterior debe aplicarse, en particular, a las operaciones de tratamiento a gran escala que persiguen tratar una cantidad considerable de datos personales a nivel regional, nacional o supranacional y que podrían afectar a un gran número de interesados y entrañen probablemente un alto riesgo, por ejemplo, debido a su sensibilidad, cuando, en función del nivel de conocimientos técnicos alcanzado, se haya utilizado una nueva tecnología a gran escala y a otras operaciones de tratamiento que entrañan un alto riesgo para los derechos y libertades de los interesados, en particular cuando estas operaciones hace más difícil para los interesados el ejercicio de sus derechos*.

[166] Grupo de Trabajo sobre Protección de Datos del Artículo 29, 10. Y, desde luego, resulta obligatoria cuando dicha toma de decisiones se base en la evaluación sistemática y exhaustiva de aspectos personales de personas físicas, como la elaboración de perfiles (art. 35(3)(a) RGPD).

[167] El artículo 35(3)(a) suprime el término "únicamente" en lo que respecta a la toma de decisiones: *evaluación sistemática y exhaustiva de aspectos personales de personas físicas que se base en un tratamiento automatizado, como la elaboración de perfiles, y sobre cuya base se tomen decisiones que produzcan efectos jurídicos para las personas físicas o que les afecten significativamente de modo similar*;

de perfiles[168]. Así, al contrario que los derechos de información y acceso (arts. 13(2)(f), 14(2)(g) y 15(1)(h) RGPD) o las salvaguardas del artículo 22, esta herramienta del RGPD aborda, a mi juicio, la toma de decisiones desde una perspectiva más amplia, y garantista con los derechos de las personas interesadas.

Si, además, la toma de decisiones se basa en el tratamiento de categorías especiales de datos -relativos a la salud-, la obligación de realizar la EIPD por parte del responsable que implementa el sistema automatizado para su uso en el contexto sanitario resulta indiscutible, dado que el tratamiento de esta clase de datos constituye un criterio para la determinación del alto riesgo para los interesados[169].

2.2.2.Inclusión de la intervención humana como medida organizativa en la EIPD

En cuanto a las obligaciones concretas que exige la realización de la EIPD, esta evaluación debe incluir, entre otros: «*las medidas previstas para afrontar los riesgos, incluidas garantías, medidas de seguridad y mecanismos que garanticen la protección de datos personales, y a demostrar la conformidad con el presente Reglamento, teniendo en cuenta los derechos e intereses legítimos de los interesados y de otras personas afectadas*» (art. 35(7)(d) RGPD). La toma de decisiones automatizada basada en la elaboración de perfiles que produce efectos jurídicos o significativos contiene riesgos que el responsable debe identificar en función de las operaciones de tratamiento previstas y de los fines del tratamiento. Una vez identificadas las operaciones de tratamiento y su finalidad, evaluada su necesidad y proporcionalidad e identificados los riesgos (arts. 35(7)(a),(b) y (c) RGPD), el responsable del tratamiento debe determinar las medidas adoptadas para demostrar el cumplimiento de la normativa[170].

No solo eso, el responsable debe demostrar la eficacia de dichas medidas para garantizar la protección de datos personales teniendo en cuenta los derechos y libertades de las personas que van a ser objeto de la toma de decisiones automatizada. De acuerdo con el considerando 74 RGPD, el responsable del tratamiento no solo está obligado a aplicar medidas oportunas y eficaces y ha de poder demostrar la conformidad de las actividades de tratamiento, sino que ha de poder demostrar también la eficacia de dichas medidas. Es decir, la demostración del cumplimiento normativo no es un ejercicio simplemente descriptivo del tipo de tratamiento de datos personales, sus riesgos y las medidas adoptadas para su mitigación, sino que es también un proceso de documentación empírico que debe permitir evaluar la eficacia de las medidas aplicadas.

Además, en aplicación del carácter continuo de la EIPD[171], la gestión de riesgos a la que obliga el Reglamento incluye el seguimiento y verificación de la eficacia de las medidas adoptadas, realizando un proceso de revisión y reevaluación de las medidas cuando resulte necesario, durante el ciclo de vida completo del tratamiento de datos personales a fin de garantizar que la eficacia de dichas medidas se mantiene, obteniendo los resultados esperados y sin que se produzcan alteraciones en la naturaleza, ámbito, contexto y fines del tratamiento[172]. En definitiva, las medidas adoptadas deben demostrarse eficaces no solo en el momento anterior al tratamiento, cuando el RGPD impone la realización de la EIPD (art. 35(1) RGPD), sino durante todo el ciclo vital del tratamiento.

---

[168] Grupo de Trabajo sobre Protección de Datos del Artículo 29, «Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679», 33.

[169] Grupo de Trabajo sobre Protección de Datos del Artículo 29, «Directrices sobre la evaluación de impacto relativa a la protección de datos (EIPD) y para determinar si el tratamiento "entraña probablemente un alto riesgo" a efectos del Reglamento (UE) 2016/679», 11.

[170] Algunas de estas medidas vienen impuestas por el RGPD, como la introducción de intervención humana significativa previa a la producción de efectos para la toma de decisiones no basada únicamente en el tratamiento automatizado (art. 22(1) RGPD), o el derecho a obtener intervención humana por parte del responsable, a expresar su punto de vista y a impugnar la decisión para decisiones basadas únicamente en el tratamiento automatizado (art. 22(3) RGPD). Lo cual no obsta para que el responsable pueda adoptar, además, otras medidas no expresamente previstas y que puedan resultar eficaces para mitigar los riesgos identificados.

[171] El art. 35(11) RGPD impone su revisión: *al menos cuando exista un cambio del riesgo que representen las operaciones de tratamiento*.

[172] Agencia Española de Protección de Datos (AEPD), «Gestión del riesgo y evaluación de impacto en tratamientos de datos personales», 22 y 50.

Como ya se ha mencionado brevemente, la intervención humana ha de considerarse entre las medidas organizativas cuya efectividad debe demostrarse de forma continuada.

Parte de la doctrina ya había identificado la EIPD como una herramienta útil a la hora de determinar el grado de humanización realmente existente en la toma de decisiones automatizada y si dichas decisiones se basan o no únicamente en el tratamiento automatizado[173]. Particular importancia adquiere, a mi juicio[174], analizar el vínculo que, en base al principio nuclear de responsabilidad del RGPD, pueda existir entre la intervención humana como mecanismo de gobernanza en el artículo 22 –con fundamento en dicho principio conforme al análisis realizado anteriormente– y la EIPD como materialización normativa más importante de este principio[175].

Si atendemos a la prohibición contenida en el artículo 22(1) RGPD -y 22(4)- de tomar decisiones basadas únicamente en el tratamiento automatizado que produzcan efectos jurídicos o significativos, conforme al análisis realizado anteriormente podemos concluir que hay dos formas de cumplir con dicha prohibición[176]. O bien se acude a alguna de las excepciones -y salvaguardas- contenidas en el artículo 22 RGPD -y se justifica debidamente-, o bien el responsable de tratamiento puede: «*diseñar un «modelo» de decisiones basadas en la elaboración de perfiles, aumentando significativamente el nivel de intervención humana de forma que el modelo ya no sea un proceso de toma de decisiones totalmente automatizadas*»[177].

En cada caso, el responsable deberá justificar cómo sirve a los objetivos normativos, cómo resulta significativa la intervención humana y, por ende, por qué es eficaz y cómo se demuestra como tal para mitigar los riesgos del tratamiento en la toma de decisiones automatizada.

Consideremos el objeto de este estudio: la implementación por el responsable del tratamiento de sistemas de IA con fines de salud para el apoyo a la toma de decisiones clínica. Al diseñar un sistema de apoyo a la toma de decisiones, este modelo debe incluir una intervención humana significativa previa -*in the loop*- a la producción de un efecto jurídico o similar en el interesado para cumplir con el RGPD y, por ende, el responsable del tratamiento deberá demostrar igualmente que dicha intervención se produce en estos términos en la EIPD, puesto que constituye una medida organizativa indispensable -y de obligado cumplimiento- para evitar los riesgos que puedan identificarse en el contexto particular de la toma de decisiones basada en la elaboración de perfiles[178].

Como sabemos, el RGPD no impone un determinado modelo para la toma de decisiones apoyada en sistemas automatizados de elaboración de perfiles, sino que para demostrar el cumplimiento con el artículo 22(1), la intervención humana debe cumplir un requisito temporal -tener lugar de forma previa a la producción de efectos jurídicos o significativos- y un requisito cualitativo -intervención humana significativa-. Incluso el responsable del tratamiento puede decidir incluir un agente humano en la toma de decisiones, *in the loop*, como medida/garantía

[173] Veale y Edwards, «Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling»; también en este sentido, Palma Ortigosa, «Automated Decision-Making in the GDPR. Algorithms in the Scope of the Data Protection».

[174] Este vínculo e intersección entre la regulación de la toma de decisiones automatizada y las obligaciones del EIPD sobre el principio de responsabilidad ya ha sido explorado en otras investigaciones aunque no en detalle sobre los mecanismos de intervención humana. Vid. Kaminski y Malgieri, «Multi-Layered Explanations from Algorithmic Impact Assessments in the GDPR»; Janssen, «An approach for a fundamental rights impact assessment to automated decision-making».

[175] Sin olvidar el deber de protección de datos por diseño, Martínez, «Cuestiones de ética jurídica al abordar proyectos de Big Data. El contexto del Reglamento general de protección de datos», 160.

[176] En cualquiera de los casos, el GT29 parece sugerir que la EIPD requeriría evaluar el grado de participación humana en el modelo de toma de decisiones: *Como parte de la EIPD, el responsable del tratamiento debe identificar y registrar el grado de participación humana en el proceso de toma de decisiones y en qué punto se produce esta*. Grupo de Trabajo sobre Protección de Datos del Artículo 29, «Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679», 23.

[177] Grupo de Trabajo sobre Protección de Datos del Artículo 29, 33.

[178] Dado que a los sistemas de toma de decisiones no basados únicamente en el tratamiento automatizado no son aplicables salvaguardas como el derecho a expresar su punto de vista e impugnar la decisión, la intervención humana como componente esencial de la toma de decisiones y, por ende, la evaluación sobre su cumplimiento conforme al RGPD adquiere una importancia determinante en la mitigación de los riesgos que esta clase de tratamiento pueda tener sobre los interesados.

para la gestión del riesgo en el tratamiento -aunque tuviese la habilitación normativa para tomar dichas decisiones de forma plenamente automatizada-[179].

Ahora, cuando se trata de demostrar la eficacia de la intervención humana como medida organizativa para la reducción del riesgo del tratamiento de datos personales -esquivando de forma legítima la prohibición del artículo 22 RGPD-, se pone de manifiesto, una vez más, que los responsables del tratamiento necesitan mayor seguridad jurídica a la hora de determinar qué puede entenderse por intervención humana significativa -y poder justificarlo debidamente en la EIPD-.

2.2.3. Diseño de una intervención humana demostrable para la toma de decisiones clínica

De lo desarrollado hasta el momento en relación con las disposiciones relativas a la realización de EIPD por el responsable del tratamiento, podemos entender que un responsable del tratamiento que pretenda implementar un sistema de IA con fines de salud para la toma de decisiones clínica -objeto de estudio en esta investigación- habrá de cumplir con la obligación de realizar una EIPD conforme al artículo 35 RGPD[180]. Que la realización de la EIPD requiere justificar el cumplimiento normativo del proceso de toma de decisiones, incluyendo las medidas adoptadas para abordar los riesgos de este tratamiento de datos para los derechos y libertades de las personas interesadas, así como la eficacia de estas medidas. Entre las medidas organizativas a adoptar de forma obligatoria, y cuya eficacia debe probarse -debe ser "demostrable"-, encontramos la intervención humana como componente esencial para la toma de decisiones no basada únicamente en el tratamiento automatizado -legítimas conforme a 22(1) RGPD-. A su vez, el RGPD exige una intervención humana significativa, ¿cómo pueden los responsables del tratamiento diseñar procesos de toma de decisiones con intervención humana significativa y demostrable para la toma de decisiones clínica?

Una de las principales críticas relacionadas con la EIPD es que el texto del Reglamento resulta más bien vago en lo que respecta a la metodología concreta que debe aplicarse, limitándose a establecer los contenidos mínimos que debe contener[181]. Aunque las autoridades de control y el CEPD han paliado parcialmente esta problemática en los últimos años a través de la publicación de guías y directrices, el análisis de la intervención humana es residual en este sentido, a pesar de la importancia destacada que adquiere en el marco de la toma de decisiones automatizada[182].

En el análisis previo sobre la intervención humana significativa en el contexto asistencial, se han extraído una serie de criterios que pretenden servir de base para aportar seguridad jurídica en la interpretación de este término, los cuales pueden resumirse en los siguientes puntos: [1] la intervención humana significativa debe realizarse por personal clínico con autoridad y competencia para modificar el resultado algorítmico; [2] la aplicación rutinaria de los resultados algorítmicos conlleva una dejación de la responsabilidad incompatible con la normativa de

[179] Así recoge la AEPD que ese control humano puede establecerse como una medida que actúa en la propia definición de la naturaleza, ámbito, contexto o fines del tratamiento, en las siguientes formas: *Reemplazar tratamientos automatizados por tratamientos manuales que incorporen procedimientos de supervisión y control; Llevar a cabo la supervisión humana de las decisiones automatizadas; Utilizar personal especialmente cualificado en determinadas fases del tratamiento, especialmente en su supervisión*. Agencia Española de Protección de Datos (AEPD), «Gestión del riesgo y evaluación de impacto en tratamientos de datos personales», 104.

[180] Salvo contadas excepciones. Si el responsable considerase que la toma de decisiones automatizada que pretende implementar no entraña un alto riesgo, debe justificar y documentar los motivos por los que no se realiza una EIPD e incluir/registrar las opiniones del delegado de protección de datos. En caso de duda, la recomendación es que se realice la EIPD: *En los casos en los que no esté claro si se requiere una EIPD, el GT29 recomienda realizar una, ya que esta evaluación representa un instrumento práctico para ayudar a los responsables del tratamiento a cumplir la legislación de protección de datos*. Grupo de Trabajo sobre Protección de Datos del Artículo 29, «Directrices sobre la evaluación de impacto relativa a la protección de datos (EIPD) y para determinar si el tratamiento "entraña probablemente un alto riesgo" a efectos del Reglamento (UE) 2016/679», 9 y 13.

[181] Gstrein, «European AI Regulation: Brussels Effect versus Human Dignity?», 13.

[182] La propia AEPD declara la necesidad de gestionar los errores técnicos derivados de la implementación automatizada de tratamientos de datos, incluyendo tanto sistemas plenamente automatizados como sistemas de apoyo a la toma de decisiones, reconociendo que la experiencia demuestra que la estimación del impacto que tienen estos fallos y errores en los datos personales no se suele tener en cuenta (pudiendo tener efectos nocivos sobre la calidad de los datos o efectos discriminatorios), en Agencia Española de Protección de Datos (AEPD), «Gestión del riesgo y evaluación de impacto en tratamientos de datos personales», 43.

protección de datos, por ende, la intervención humana significativa del personal clínico debe tener una influencia real sobre el proceso decisorio evitando que éste se convierta, de facto, en un sistema plenamente automatizado; [3] la intervención humana significativa está fundamentada en que el responsable del tratamiento pueda mantener la responsabilidad sobre el tratamiento automatizado de datos personales y, por ende, teniendo en cuenta el contexto clínico que rodea al paciente en su sentido más amplio, la intervención del personal clínico al cumplimiento normativo y su demostración. Acerca de estos criterios hay varias cuestiones que me gustaría resaltar en relación con la demostración de su cumplimiento en la EIPD.

A pesar de que, en coherencia con los principios de protección de datos desde el diseño y por defecto, la EIPD debe realizarse antes del tratamiento -art. 35(1) y (10) RGPD-, no se trata de una obligación puntual: «*La actualización de la EIPD a lo largo del proyecto de ciclo de vida garantizará que se tenga en cuenta la protección de los datos y la intimidad y propiciará la creación de soluciones que fomenten el cumplimiento*»[183]. La EIPD es un proceso continuo, y dicha continuidad adquiere especial relevancia en la demostración del cumplimiento de la prohibición de la toma de decisiones automatizada. Mientras que la autoridad y competencia del personal clínico son características de un carácter más formal-institucional, la evaluación del resto de criterios -si su intervención no es rutinaria y contribuye al cumplimiento normativo y su demostración- requieren de una evaluación más continua que encaja en la propia naturaleza de la EIPD.

Para Noto La Diega, la evaluación sobre si la intervención humana es o no significativa solo podría determinarse caso por caso y, por tanto, la aplicación de las decisiones del 22(1) también debería depender de ese caso por caso[184]. No obstante, hay razones de peso para considerar que el carácter significativo o no de la intervención humana debe determinarse en una evaluación sistémica y continuada del modelo de toma de decisiones. Por un lado, determinar caso por caso el carácter significativo de la intervención humana y, por ende, la aplicación de la prohibición del 22(1) RGPD provocaría una inseguridad jurídica inasumible tanto para responsables del tratamiento como para los interesados[185]. Por otro lado, es insostenible observar en el caso por caso si el personal clínico realiza una aplicación rutinaria de los resultados algorítmicos, es decir, no puede determinarse si el modelo de decisiones está afectado por un sesgo de automatización por el hecho de que el agente humano haya seguido o no una decisión particular. Es necesario realizar una evaluación institucional o sistemática de la intervención humana en el proceso de toma de decisiones clínico en su conjunto.

En este punto pueden ser muy útiles los ítems planteados por Cabitza, Campagner y Datteri para la evaluación de sistemas de IA aplicados al ámbito clínico-asistencial[186]:

- Cuántas veces los responsables de la toma de decisiones clínica (RC) y el sistema están de acuerdo en un caso (concordancia)
- Cuántas veces los RC creen que el sistema tiene razón (confianza);
- Cuántas veces se demuestra que el sistema tiene razón después de los hechos (precisión);

[183] Grupo de Trabajo sobre Protección de Datos del Artículo 29, «Directrices sobre la evaluación de impacto relativa a la protección de datos (EIPD) y para determinar si el tratamiento "entraña probablemente un alto riesgo" a efectos del Reglamento (UE) 2016/679», 16.

[184] Noto La Diega, «Against the Dehumanisation of Decision-Making – Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information», 19.

[185] Por supuesto que en el caso concreto se podrá determinar que la intervención humana no ha corregido un error del resultado algorítmico o incluso que ha cambiado un resultado algorítmico que resultaba más apropiado en términos normativos, por ello es tan relevante la posibilidad de impugnar tanto la inferencia algorítmica como la decisión final, esté la misma precedida o no de intervención humana. Sin embargo, dichos errores en el caso concreto no pueden llevarnos a la consideración de que la intervención humana en un modelo de decisiones concreto sea o no significativa. Solo una evaluación sistemática puede determinar este aspecto. Un ejemplo sencillo. Pongamos que estamos evaluando la exactitud del sistema sobre un total de 100 decisiones, el humano ha seguido el criterio algorítmico en 80 de las cuales 75 eran correctas -es decir, el humano se tenía que haber apartado en 5 ocasiones y no lo hizo-, mientras que de las 20 veces que se apartó del criterio algorítmico, la máquina había errado en 15 ocasiones -es decir, el humano debía haber seguido el criterio en otras 5-. Por tanto, el sistema antes de la intervención humana sería exacto en 80/100 ocasiones, mientras que tras la intervención lo sería en 90/100 ocasiones, ¿tendría sentido sostener que la intervención humana no ha sido significativa en las 10 ocasiones en las que se equivocó?

[186] Cabitza, Campagner, y Datteri, «To Err is (only) Human. Reflections on How to Move from Accuracy to Trust for Medical AI». La traducción es mía.

- Cuántas veces los RC han cambiado de opinión por el consejo del sistema (impacto en el rendimiento);
- Cuántas veces los RC han percibido que el sistema es útil (utilidad);
- Cuántas veces los RC han creído recibir elementos interesantes para tomar sus decisiones (utilidad);
- Cuántas veces creen haber sido más rápidos en su toma de decisiones o, más bien, obstaculizados por el sistema (satisfacción);
- Cuántas veces el resultado del sistema ha facilitado o censurado la discusión con sus colegas o con los pacientes (impacto colaborativo);
- Cuántas veces ha facilitado el aprendizaje o ha aliviado de la "carga" el recuerdo, el razonamiento analógico y la inferencia deductiva (impacto cognitivo):
- Cuántas veces los usuarios creen que una herramienta de este tipo puede haber alimentado el sesgo de confirmación, la medicina defensiva, y otros sesgos (como el sesgo de automatización y la complacencia).

Además, esta evaluación institucional de la intervención humana puede ayudar a los responsables del tratamiento de datos a prevenir los efectos secundarios de esos sistemas en la sociedad que van más allá del nivel individual[187] y, por tanto, a comprobar cómo puede influir la intervención humana en la mitigación de dichos efectos[188]. Determinar desde este punto de vista si la intervención humana es significativa puede resultar complejo, como se ha puesto de manifiesto más arriba, en todo contexto normativo complejo -y el ámbito clínico lo es- encontraremos distintos valores compitiendo entre sí[189], lo cual puede derivar en que la intervención humana derive -por ejemplo- en un sistema menos exacto, pero más justo, o viceversa. O, como hemos visto en el ámbito clínico, decisiones que afectan a valores más cercanos al ejercicio de derechos fundamentales y al libre desarrollo de la personalidad en el marco bioético, ¿es posible resolver aquí esta tensión?

Parece complicado y el trabajo jurídico por realizar en este campo es amplio.

Llegados a este punto, parece importante recalcar las limitaciones que tanto la EIPD como la intervención humana tienen como mecanismos de gobernanza de los sistemas algorítmicos en el RGPD. Tal y como se señalaba anteriormente, la EIPD es un instrumento normativo de autoevaluación con una publicidad muy limitada, cuyo valor reside en conducir a la construcción de mejores sistemas en su conjunto responsabilizando al responsable del tratamiento del cumplimiento del RGPD y su demostración. El rol de la intervención humana debe entenderse en ese conjunto. Del mismo modo que la evaluación de la intervención humana significativa es sólo una de las medidas que deben/pueden adoptarse y justificarse en el diseño del proceso de toma de decisiones automatizada, y que pueden servir no sólo para evitar errores, sesgos y discriminaciones, sino también para legitimar un sistema o incluso respetar la dignidad de un individuo dentro de él[190].

Además, el RGPD no impone un grado y clase concreto de intervención humana[191]. Con lo cual, los responsables del tratamiento cuentan con un amplio abanico de posibilidades para

[187] Tamò-Larrieux, «Decision-making by machines: Is the 'Law of Everything' enough?», 14-15.

[188] La aplicación de medidas técnicas y organizativas apropiadas para garantizar y demostrar el cumplimiento del RGPD se debe hacer teniendo en cuenta *la naturaleza, el ámbito, el contexto y los fines del tratamiento así como los riesgos de diversa probabilidad y gravedad para los derechos y libertades de las personas físicas*. Grupo de Trabajo sobre Protección de Datos del Artículo 29, «Directrices sobre la evaluación de impacto relativa a la protección de datos (EIPD) y para determinar si el tratamiento "entraña probablemente un alto riesgo" a efectos del Reglamento (UE) 2016/679», 22. Aunque puede considerarse que la evaluación de riesgos sobre los derechos y libertades de las personas físicas en la EIPD (arts. 35(1) y 35(7)(c)), incluyen los derechos fundamentales, el RGPD no incluye enfoques detallados sobre cómo o con qué nivel de detalle deben evaluarse el impacto sobre dichos derechos fundamentales, vid. Janssen, «An approach for a fundamental rights impact assessment to automated decision-making», 85.

[189] Brennan-Marquez y Henderson, «Artificial Intelligence and Role-Reversible Judgment», 142.

[190] Kaminski y Malgieri, «Algorithmic impact assessments under the GDPR: producing multi-layered explanations», 133.

[191] Tampoco los tribunales deben imponer su propia visión sobre cómo cumplir con los requisitos del GDPR en esta materia. Así lo expresaba, en primera instancia, la Sentencia del *England and Wales High Court*, de 4 de septiembre de 2019, en el caso *R (on the application of Edward Bridges) v the Chief Constable of South Wales Police*. [2019] EWHC 2341 (Admin), caso núm. CO/4085/2018: *(146) On a complaint about a failure to comply with section 64 Data Protection Act 2018, it is for the Court to decide whether the data controller has discharged that obligation. What is required is compliance itself, i.e. not simply an attempt to comply that falls within a range of reasonable conduct. However, when determining whether the steps taken by the data controller meet the requirements of section 64, the Court will not necessarily substitute its own view for that of the data controller on all matters.*

cumplir con la intervención humana como componente esencial de la toma de decisiones impuesta por el Reglamento y demostrar su cumplimiento en la EIPD. Coincidiendo plenamente con las palabras de Binns hay un modelo de colaboración humano-máquina implícito en el artículo 22(1) RGPD[192] y la EIPD es, a mi juicio, una herramienta adecuada para evaluar, y reevaluar si fuera necesario, el mejor modelo de colaboración posible dentro de los amplios márgenes establecidos en dicha disposición[193].

En el ámbito clínico, a mi entender, esta búsqueda del mejor modelo de colaboración humano-máquina requerirá del establecimiento de protocolos clínicos -guías para la práctica clínica- que incorporen en su metodología la aplicación de medidas organizativas para la gestión del riesgo en el tratamiento de datos personales. El diseño de protocolos clínicos que estipulen cómo los decisores humanos deben interactuar con los sistemas de IA que sirven de apoyo a la toma de decisiones puede contribuir al aumento de la precisión y la eficacia en la toma de decisiones[194]. Además, podrían incorporarse sistemas de evaluación institucional como los mencionados arriba a la fase de evaluación o monitorización del cumplimiento del protocolo clínico[195] que, a su vez, podrían servir para demostrar la eficacia de la intervención humana como medida organizativa en la EIPD de forma continua.

En resumen, el artículo 35 del RGPD exige a los responsables del tratamiento que demuestren de forma continua cómo se introduce la intervención humana de forma significativa para cumplir con el derecho del interesado a no ser sometido a una toma de decisiones individual automatizada. Esto no significa que el Reglamento imponga un modelo específico de toma de decisiones, ni que deba entenderse al margen del resto de medidas organizativas y salvaguardas para demostrar que dicha intervención humana resulta significativa. En el ámbito clínico debe considerarse el potencial de combinar la protocolización clínica de la colaboración humano-máquina con la gestión de riesgos desde el punto de vista de protección de datos.

3. Limitaciones del modelo normativo actual: gobernanza y supervisión humana de los sistemas de IA durante todo su ciclo de vida

A pesar de que la normativa de protección de datos, como se ha podido ver, exige al responsable del tratamiento el diseño de una intervención humana significativa y demostrable para el uso de sistemas de IA en el contexto clínico, es necesario referirse ahora a la indisoluble relación existente entre el diseño y desarrollo de un modelo y de cómo esta fase modula y constriñe las posibilidades de supervisión humana en la fase de implementación del sistema automatizado en los procesos de toma de decisiones. Los beneficios previstos al implantar la automatización por desarrolladores y quienes deciden utilizarlos -aumento de la eficiencia, mejora de la seguridad, aumento de la flexibilidad de las operaciones, reducción de la carga de trabajo de los operarios, etc.- no siempre se materializan y pueden verse contrarrestados por los costes de rendimiento humano asociados al uso inadaptado de una automatización mal diseñada o inadecuadamente formada[196].

Por ello, resulta indispensable asegurar la utilización de los modelos de IA conforme al fin y uso para el que fueron diseñados y desarrollados, es decir, la configuración del elemento humano en el diseño y desarrollo juega un papel clave en la futura implementación en el contexto clínico -y en cualquier otro contexto-[197]. En este sentido, es necesario que el responsable del tratamiento evalúe correctamente las características bajo las que fue desarrollado y validado el sistema, de forma que pueda considerar qué cambios ha de introducir en el entorno clínico (pre)existente en el que se implementará para asegurar un funcionamiento adecuado del

[192] Binns, «Human Judgment in algorithmic loops: Individual justice and automated decision-making», 9.
[193] Selbst también se muestra optimista sobre el potencial de las evaluaciones de impacto en el diseño de colaboraciones humano-máquina para el uso de sistemas automatizados en la prevención del crimen. Vid. Selbst, «Disparate Impact in Big Data Policing».
[194] Vid. al respecto Cabitza, Campagner, y Sconfienza, «Studying human-AI collaboration protocols: the case of the Kasparov's law in radiological double reading».
[195] Vid. fases del protocolo clínico en Saura Llamas y Saturno Hernández, «Protocolos clínicos: ¿cómo se construyen? Propuesta de un modelo para su diseño y elaboración».
[196] Parasuraman y Manzey, «Complacency and Bias in Human Use of Automation: An Attentional Integration», 381.
[197] Jamieson y Goldfarb, «Clinical considerations when applying machine learning to decision-support tasks versus automation», 779.

sistema (y considere también el coste -en términos económicos y humanos- de dichos cambios).

Por supuesto, esta evaluación resulta apropiada para demostrar el cumplimiento normativo en el RGPD: se trata de asegurar que un sistema de IA es utilizado conforme al fin y uso para el que fue diseñado -y validado-. No obstante, esta posibilidad encuentra severas limitaciones en la normativa de protección de datos. A pesar de que podemos considerar aplicable el RGPD a fases de diseño y desarrollo de sistemas de IA cuando se produzca un tratamiento de datos personales en el entrenamiento o validación de los mismos, las disposiciones que afectan a este tipo de tratamiento lo hacen de forma más bien tangencial y, desde luego, no hay disposición alguna que vincule los mecanismos de intervención humana en el RGPD con las fases de diseño y desarrollo de los modelos. La regulación de la toma de decisiones automatizada la normativa de protección de datos establece sus obligaciones para la fase de implementación de los sistemas y limita la responsabilidad únicamente a quien hace uso de estos sistemas bajo su autoridad -responsable del tratamiento-. Como consecuencia, no puede considerarse que la intervención humana en el RGPD garantice la supervisión humana de los sistemas de IA durante todo su ciclo de vida, tal y como se ha reclamado en las distintas iniciativas para la regulación de la IA en el contexto europeo[198].

Ahora bien, el diseño y desarrollo de determinados sistemas de IA puede estar sujeto a normativa sectorial que podría complementar en este aspecto la normativa de protección de datos. Este es el caso de los sistemas de IA con fines de salud objeto de esta investigación, a los cuales -como se verá a continuación- resulta aplicable con carácter general la normativa europea productos sanitarios, en la que se incluye el Reglamento 2017/745 sobre los productos sanitarios (MDR por sus siglas en inglés, en adelante)[199] y el Reglamento 2017/746 sobre los productos sanitarios para diagnóstico in vitro (IVDR, en adelante)[200].

Por ello, en primera instancia, este apartado realiza un análisis de esta normativa y, en particular, de las obligaciones que se establecen en la fase de diseño y desarrollo de los sistemas previa a su certificación como productos sanitarios. En segunda instancia, a partir de las limitaciones identificadas en la normativa de productos sanitarios se realiza un análisis de la repercusión que el Reglamento de IA tendrá en este contexto a partir de su plena aplicabilidad. En particular, este análisis indaga en las intersecciones regulatorias entre el RGPD, la normativa de productos sanitarios y el Reglamento de IA, para garantizar la supervisión humana de los sistemas de IA durante todo su ciclo de vida.

3.1. Sistemas de inteligencia artificial como productos sanitarios

Para poder dar respuesta al interrogante sobre si los sistemas de IA pueden considerarse productos sanitarios -y, por ende, se debe aplicar a aquéllos la normativa que regula estos últimos- debemos conocer qué entiende dicha normativa por producto sanitario y en qué medida esta definición engloba los distintos sistemas de IA que son objeto de estudio en esta investigación[201]. Con carácter general, la normativa de productos sanitarios establece un sistema de calificación (¿qué productos han de ser considerados como productos sanitarios a efectos normativos?) y un sistema de clasificación basado en el riesgo (¿qué obligaciones jurídicas han de aplicarse a un producto sanitario en base al tipo de riesgo que representa?).

A efectos del artículo 2(1) MDR[202], para calificar un artilugio como producto sanitario, éste debe estar destinado a ser utilizado en personas con un fin médico específico. En cuanto al artilugio

[198] Vid. más arriba el epígrafe 1.2. La importancia de la supervisión humana en este contexto.
[199] 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.° 178/2002 y el Reglamento (CE) n.° 1223/2009 y por el que se derogan las Directivas 90/385/CEE y 93/42/CEE del Consejo
[200] 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
[201] Se tomará por referencia en este sentido el MDR por tratarse de la regulación de carácter más general, aunque se realizarán puntualizaciones relativas a la aplicación del IVDR.
[202] Art. 2(1) MDR, «producto sanitario»*: todo instrumento, dispositivo, equipo, programa informático, implante, reactivo, material u otro artículo destinado por el fabricante a ser utilizado en personas, por separado o en combinación, con alguno de los siguientes fines médicos específicos:*
— diagnóstico, prevención, seguimiento, predicción, pronóstico, tratamiento o alivio de una enfermedad,

en sí, no resulta determinante, aunque la disposición incluye la posibilidad expresa de que se trate de un programa informático. Y los sistemas de IA son, antes de nada, programas informáticos[203], por tanto, no hay impedimento alguno para considerar que los sistemas de IA son programas informáticos o *software* en los términos expresados por el MDR[204].

Desde luego, la cuestión determinante es la de que el programa informático tenga un fin médico específico. A estos efectos, deben descartarse aquellos programas informáticos que se limitan a tareas de almacenaje, archivado, comunicación o búsqueda[205], ahora bien, si el programa realiza alguna acción sobre los datos que sobrepase en algún sentido estas funciones es probable que tenga un fin médico específico. Entre otros, diagnóstico, prevención, seguimiento, predicción, pronóstico, tratamiento o alivio de una enfermedad o discapacidad, y siempre que no ejerza su acción principal prevista en el interior o en la superficie del cuerpo humano por mecanismos farmacológicos, inmunológicos o metabólicos[206]. Como puede verse, casi con toda probabilidad un sistema de IA que realiza inferencias sobre el estado de salud de un paciente -elaboración de perfiles a efectos del RGPD-, entrará -a su vez- en el ámbito de aplicación del MDR al realizar esa inferencia sobre su estado de salud. Es decir, la amplitud de los fines médicos definidos en el artículo 2(1) MDR debería ser, en principio, una garantía de que prácticamente todo sistema de IA que pueda ser utilizado para el diagnóstico, prevención, seguimiento, prognosis o tratamiento sea calificado como producto sanitario[207].

Por otro lado, mientras que el sistema de calificación en el MDR tiene su fundamento en la finalidad a la que se destina el producto sanitario, el sistema de clasificación normativa -que afecta a los productos sanitarios ya calificados como tales- tiene su fundamento en el riesgo que pueda representar el uso de dicho producto. Ello quiere decir que las obligaciones concretas en la normativa de productos sanitarios dependen del riesgo del producto a calificar,

---

— diagnóstico, seguimiento, tratamiento, alivio o compensación de una lesión o de una discapacidad,
— investigación, sustitución o modificación de la anatomía o de un proceso o estado fisiológico o patológico,
— obtención de información mediante el examen in vitro de muestras procedentes del cuerpo humano, incluyendo donaciones de órganos, sangre y tejidos,
y que no ejerce su acción principal prevista en el interior o en la superficie del cuerpo humano por mecanismos farmacológicos, inmunológicos ni metabólicos, pero a cuya función puedan contribuir tales mecanismos. (…)

[203] Artículo 3(1) RIA, «Sistema de inteligencia artificial (sistema de IA)»: *un sistema basado en una máquina que está diseñado para funcionar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación tras el despliegue, y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales.*

[204] En este mismo sentido se pronuncia Kiseleva, en Kiseleva, «AI as a Medical Device: Is it Enough to Ensure Performance Transparency and Accountability?», 9-10. Además, la definición de software aportada por el MDCG es, si cabe, más amplia*: se entiende por programa informática una serie de instrucciones que procesan datos de entrada creando datos de salida*. Grupo de Coordinación de Productos Sanitarios (MDGC), «Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 – MDR and Regulation (EU) 2017/746 – IVDR (MDCG 2019-11)», 6.

[205] El MDCG aporta como ejemplos para la distinción entre estas acciones que, por ejemplo, busque en la imagen hallazgos que apoyen una hipótesis clínica en cuanto al diagnóstico o la evolución de la terapia o que amplíe localmente el contraste del hallazgo en una pantalla de imagen para que sirva de apoyo a la decisión o sugiera una acción a realizar por el usuario. Sin embargo, la alteración de la representación de los datos con fines de embellecimiento/cosméticos o de compatibilidad no permitiría calificar dicho programa como sanitario. Grupo de Coordinación de Productos Sanitarios (MDGC), «Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 – MDR and Regulation (EU) 2017/746 – IVDR (MDCG 2019-11)», 7. Por ejemplo, un sistema de procesamiento de lenguaje natural (NLP) basado en aprendizaje profundo para automatizar la recogida de la HC, quedaría fuera de la calificación de producto sanitario, al igual que los procesadores de texto actuales en los que se recoge de forma manual por las médicas de atención primaria. También se situarían al margen de la regulación del MDR los sistemas hospitalarios de información, como un sistema automatizado para establecer las citas de consultas externas, salvo que alguno de los módulos de estos sistemas cumpla con un fin médico específico, en cuyo caso dicho módulo sería calificado como producto sanitario.

[206] La simple utilización del programa en el contexto clínico no es suficiente para considerarlo un producto sanitario. Así lo interpretaba el TJUE en relación con la anterior Directiva 93/42 sobre productos sanitarios: *Así pues, el legislador ha dejado claro que, en relación con los programas informáticos, para que éstos queden comprendidos en el ámbito de aplicación de la Directiva 93/42, no basta con que se utilicen en un contexto sanitario, sino que es necesario que su finalidad, definida por su fabricante, sea específicamente médica*. Sentencia del Tribunal de Justicia (Sala Tercera) de 22 de noviembre de 2012, *Brain Products GmbH contra BioSemi VOF y otros*, Asunto C-219/11, par. 17.

[207] Es necesario añadir que el MDR no distingue entre productos destinados a ser utilizados como apoyo a la toma de decisiones y productos destinados a ser utilizados directamente de forma plenamente automatizada. Lo relevante es, en cualquier caso, el fin médico al que se destine.

esto es, cuanto más alto sea el riesgo para los pacientes, más alta la clasificación y más estrictas las obligaciones impuestas al proveedor del producto sanitario[208].

La clasificación en el MDR se realiza de la siguiente manera. Un programa informático sanitario puede ser clasificado entre los niveles o clases I, IIa, IIb y III, que determinan de menor a mayor el grado de intervención regulatoria.

Con carácter general, para los programas informáticos independientes de cualquier otro producto es aplicable la Regla 11 del anexo VIII del MDR[209]. En virtud de la misma, los programas destinados a proporcionar información que se utiliza para tomar decisiones con fines terapéuticos o de diagnóstico se clasifican en la clase IIa, salvo que las decisiones tengan un impacto que pueda causar: muerte o deterioro irreversible (clase III), o deterioro grave de la salud o intervención quirúrgica (clase IIb). Pueden distinguirse también por su finalidad aquellos programas que, destinados a observar procesos fisiológicos –clase IIa con carácter general, salvo que observe parámetros fisiológicos vitales, en cuyo caso se clasificaría como clase IIb-[210], y el resto de programas (clase I)[211].

Esta clasificación determina fundamentalmente el tipo de evaluación clínica requerido para la certificación del producto (CE) o el seguimiento poscomercialización por el que se rige cada producto[212], que son las dos figuras fundamentales por las que se establecen las obligaciones que afectan a los fabricantes de productos sanitarios.

3.1.1.Certificación y seguimiento poscomercialización de los productos sanitarios

El marcado CE es una certificación característica de la legislación armonizada de la Unión Europea, a través del cual un fabricante declara que un producto comercializado cumple con una determinada normativa. En el caso de los productos sanitarios, el fabricante o

[208] Kolfschooten, «EU regulation of artificial intelligence: Challenges for patients' rights», 103.

[209] Regla 11, anexo VII MDR: *Los programas informáticos destinados a proporcionar información que se utiliza para tomar decisiones con fines terapéuticos o de diagnóstico se clasifican en la clase IIa, salvo si estas decisiones tienen un impacto que pueda causar: — la muerte o un deterioro irreversible del estado de salud de una persona, en cuyo caso se clasifican en la clase III, o; — un deterioro grave del estado de salud de una persona o una intervención quirúrgica, en cuyo caso se clasifican en la clase IIb.*
Los programas informáticos destinados a observar procesos fisiológicos se clasifican en la clase IIa, salvo si se destinan a observar parámetros fisiológicos vitales, cuando la índole de las variaciones de dichos parámetros sea tal que pudiera dar lugar a un peligro inmediato para el paciente, en cuyo caso se clasifican en la clase IIb.
Todos los demás programas informáticos se clasifican en la clase I.

[210] Los procesos y parámetros fisiológicos vitales incluyen, por ejemplo, la respiración, el ritmo cardíaco, las funciones cerebrales, los gases sanguíneos, la presión arterial y la temperatura corporal. Grupo de Coordinación de Productos Sanitarios (MDGC), «Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 – MDR and Regulation (EU) 2017/746 – IVDR (MDCG 2019-11)», 14.

[211] La guía del MDCG sobre calificación y clasificación de programas informáticos sanitarios aporta algunos ejemplos que pueden ser de utilidad a la hora de interpretar el sistema establecido por el MDR, y aclara que, cualquier cambio tanto en la finalidad prevista, como en el contexto/situación de la atención clínica en la que se utiliza ese mismo producto, podría dar lugar a una clase de riesgo diferente. Algunos ejemplos destacados por su posible desarrollo basado en IA, que pueden clasificarse respectivamente en clase III, IIb, IIa y I, en Grupo de Coordinación de Productos Sanitarios (MDGC), 28-29.:

- Un programa informático destinado a realizar diagnósticos mediante el análisis de imágenes para tomar decisiones de tratamiento decisiones de tratamiento en pacientes con ictus agudo debe clasificarse como clase III según la Regla 11(a).
- Un programa informático de diagnóstico destinado a puntuar la depresión basándose en los datos introducidos sobre los síntomas de un paciente (por ejemplo, el estado de ánimo, la ansiedad) deben clasificarse como clase IIb en virtud de la Regla 11(a).
- Un programa informático destinado a clasificar las sugerencias terapéuticas para un profesional de la salud sobre la base de la historia del paciente, los resultados de las pruebas de imagen y las características del paciente, por ejemplo, que enumera y clasifica todas las opciones de quimioterapia disponibles para las personas BRCA-positivas, debe clasificarse como clase IIa según la Regla 11(a).
- Una aplicación pretende ayudar a la concepción calculando el estado de fertilidad de la usuaria basándose en un algoritmo estadístico validado. La usuaria introduce datos de salud, como la temperatura corporal basal (TB) y los días de menstruación, para seguir y predecir la ovulación. El estado de fertilidad del día actual se refleja en uno de los tres indicadores luminosos: rojo (fértil), verde (infértil) o amarillo (fase de fluctuación del ciclo). Esta aplicación debe clasificarse como clase I según la Regla 11(c).

[212] Por ejemplo, la evaluación clínica y su documentación deben actualizarse durante todo el ciclo de vida del producto con datos clínicos obtenidos que se obtendrán de la aplicación por el fabricante del plan del seguimiento clínico poscomercialización (art. 61.11 primero MDR); si bien, en el caso de los productos de clase III, el informe de evaluación del seguimiento clínico poscomercialización y el resumen sobre seguridad y funcionamiento clínico del producto deben actualizarse, al menos, una vez al año (art. 61.11 segundo MDR).

desarrollador indica que un producto es conforme con los requisitos aplicables del MDR y de otra legislación que sea aplicable en su caso (art. 2(43) MDR). Conforme al artículo 5 MDR, la introducción en el mercado debe cumplir con los requisitos generales de seguridad y funcionamiento que figuran en el anexo I que les sean aplicables, demostrando su conformidad con dichos requisitos a través de una evaluación clínica[213].

La evaluación clínica es un proceso estructurado, transparente, iterativo y continuo que forma parte del sistema de gestión de la calidad de un producto[214]. Se trata también de un proceso agravado en función del riesgo[215]. Este proceso incluye la confirmación de la conformidad con los requisitos generales de seguridad y funcionamiento que figuran en el anexo I MDR, así como la evaluación de los efectos secundarios indeseables[216] y de la aceptabilidad de la relación beneficio-riesgo (art. 61(1) MDR)[217]. Asimismo, es responsabilidad del fabricante o desarrollador especificar y justificar el nivel de las pruebas clínicas necesario para demostrar la conformidad y la adecuación a las características del producto y su finalidad prevista (art. 61(1) MDR)[218].

La evaluación clínica de los programas informáticos o *software* sanitario tiene características ya de por sí particulares. Estas particularidades son evidentes cuando el propio MDR las prevé expresamente; al igual que existen unas normas particulares para la clasificación del software sanitario (Regla VIII del Anexo III MDR), se prevén requisitos de seguridad y funcionamiento también particulares (por ejemplo, punto 17 del Anexo I MDR)[219]. No obstante, en determinados aspectos también es necesaria una interpretación particular de algunos aspectos genéricos del MDR. Por ejemplo, el MDCG recalca que el concepto de beneficio clínico[220] para el software puede desviarse del que se aplica en el caso de los productos farmacéuticos u otros productos sanitarios, ya que el beneficio de aquél puede residir en el suministro de información médica precisa sobre los pacientes, en su caso, evaluada con respecto a la información médica obtenida mediante el uso de otras opciones y tecnologías de diagnóstico; mientras que el resultado clínico final para el paciente depende de otras opciones de diagnóstico y/o terapéuticas que podrían estar disponibles[221].

Por otro lado, el proceso de seguimiento poscomercialización en el MDR se constituye a partir del establecimiento de obligaciones de planificación, actualización, documentación y registro para los fabricantes, que deben rendir cuentas sobre el cumplimiento de las mismas ante las autoridades nacionales competentes encargadas de las actividades de vigilancia y control del mercado.

El seguimiento poscomercialización forma parte del sistema de gestión de la calidad que los fabricantes deben cumplir conforme a las obligaciones generales establecidas en el artículo 10

[213] Art. 2(44) MDR: *«evaluación clínica»: un proceso sistemático y planificado para generar, recoger, analizar y evaluar de forma continua los datos clínicos relativos a un producto para verificar su seguridad y funcionamiento, incluidos los beneficios clínicos, cuando se utilice conforme a la finalidad prevista por el fabricante;*

[214] Grupo de Coordinación de Productos Sanitarios (MDGC), «Guidance on Clinical Evaluation (MDR)/Performance Evaluation (IVDR) of Medical Device Software (MDCG 2020-1)», 10.

[215] Considerando 60 MDR: *Los procedimientos de evaluación de la conformidad para los productos de la clase I deben llevarse a cabo, generalmente, bajo la exclusiva responsabilidad de los fabricantes, dado el bajo grado de vulnerabilidad asociado a estos productos. En el caso de los productos de las clases IIa, IIb y III, debe ser obligatorio un nivel apropiado de intervención de un organismo notificado*.

[216] Anexo I MDR. Capítulo I. Requisitos generales: *Todos los riesgos conocidos y previsibles y efectos secundarios indeseables se reducirán al mínimo, y habrán de ser aceptables en relación con los beneficios evaluados para el paciente o el usuario procedentes del funcionamiento del producto en condiciones normales de uso.*

[217] Art. 2(24) MDR, «determinación de la relación beneficio-riesgo»*: el análisis de todas las evaluaciones del beneficio y del riesgo que puedan ser pertinentes a efectos del uso del producto para la finalidad prevista, cuando se utilice de acuerdo con la finalidad prevista por el fabricante;*

[218] En este sentido, son muy exhaustivas las obligaciones de documentación técnica, también sobre seguimiento poscomercialización, en los anexos II y III MDR, y que debe presentarse de forma clara, organizada, fácil de buscar e inequívoca.

[219] Incluye, entre otros, la obligación de que los sistemas informáticos se desarrollen basándose en el estado actual de la técnica, teniendo en cuenta los principios de ciclo de vida del desarrollo, gestión de los riesgos, incluida la seguridad de la información, validación y verificación.

[220] Art. 2(53) MDR, «beneficio clínico»: *el efecto positivo de un producto sobre la salud de una persona, expresado en términos de resultados clínicos significativos y mensurables pertinentes para el paciente, incluidos los resultados relacionados con el diagnóstico o con un efecto positivo en la gestión de pacientes o en la salud pública;*

[221] Grupo de Coordinación de Productos Sanitarios (MDGC), «Guidance on Clinical Evaluation (MDR)/Performance Evaluation (IVDR) of Medical Device Software (MDCG 2020-1)», 3.

MDR[222]. Esta obligación extiende la evaluación clínica más allá del diseño y desarrollo de un sistema, de forma que el fabricante debe continuar obteniendo datos clínicos para actualizar esta evaluación durante todo el ciclo de vida del producto (art. 61(11) MDR)[223], esto es, durante la fase de despliegue e implementación del sistema de IA.

Al igual que el proceso de evaluación clínica, el seguimiento poscomercialización se establece en función del riesgo del producto y debe ser adecuado a cada tipo de producto: «*adecuado para recabar, conservar y analizar activa y sistemáticamente datos pertinentes sobre calidad, funcionamiento y seguridad de un producto durante todo su ciclo de vida, extraer las conclusiones pertinentes y determinar, aplicar y supervisar cualquier actuación preventiva y correctiva*»(art. 83 (1) y (2) MDR)[224].

Dentro del seguimiento poscomercialización, los fabricantes de productos clases IIa, IIb y III están obligados a emitir un informe periódico de seguridad actualizado («PSUR») -mínimo uno cada año para clases IIb y III y mínimo uno cada dos años para clase IIa-. Dicho informe debe actualizar durante todo el ciclo de vida del producto, entre otros, las conclusiones que deben utilizarse en la determinación de la relación beneficio-riesgo y los principales resultados del informe de evaluación sobre el seguimiento clínico poscomercialización (art. 86(1) MDR). La revisión de estos aspectos puede ser particularmente útil para los sistemas de IA, habida cuenta las diferencias en el rendimiento que se vienen observando entre las fases de desarrollo y su implementación en distintos contextos del mundo real.

3.1.2. Limitaciones del modelo normativo de productos sanitarios para la regulación de los sistemas de inteligencia artificial

El modelo normativo de productos sanitarios definido aquí a grandes rasgos tiene, sin lugar a duda, aspectos positivos cuando evaluamos cómo complementa en todo el ciclo de vida de un sistema de IA la gobernanza establecida por el RGPD. Entre otros, pueden destacarse su ámbito de aplicación -donde confluyen el perfilado sobre el estado de salud de una persona con los fines médicos específicos- o la gobernanza basada en la responsabilidad sobre el cumplimiento normativo y su demostración -que guardan muchas similitudes en fases distintas para fabricantes de productos y responsables del tratamiento, respectivamente-. No obstante, y más allá de estos rasgos genéricos, es necesario poner de manifiesto algunas notas más críticas respecto de la regulación de los sistemas de IA por la normativa de productos sanitarios, así como de la comunicación entre ambas normativas -protección de datos y productos sanitarios-.

Por ejemplo, la gobernanza de los datos de entrenamiento y validación -que serán en gran medida de carácter personal-, dadas las características de los sistemas de IA, es esencial para, una vez más, garantizar ese elevado nivel de seguridad y de protección de la salud. Bien es cierto que MDR e IVDR recogen la obligación genérica de que los productos sean seguros y eficaces y no comprometan el estado clínico o la seguridad de los pacientes, y que ello debería obligar a evaluar e inspeccionar los datos de entrenamiento y validación de los sistemas de IA para cumplir con dicha obligación. Sin embargo, Stöger et al. señalan que es difícil concluir que esta obligación genérica esté, en la actualidad, siendo un mecanismo de gobernanza de datos efectivo[225].

Asimismo, parece obvio que para poder generar confianza en los sistemas de IA sanitaria es necesario que el sistema normativo genere, a su vez, la transparencia suficiente entre los actores que participan en el ciclo de vida del sistema y los reguladores encargados de velar por el cumplimiento normativo[226]. Los sistemas de evaluación de conformidad y seguimiento

[222] El modelo normativo del MDR e IVDR centra las obligaciones jurídicas y la responsabilidad en los desarrolladores o fabricantes, sin embargo, el papel de los responsables del despliegue de los sistemas y la coordinación entre ambos está muy limitada en el contexto del seguimiento poscomercialización actual.
[223] Con una periodicidad al menos anual para productos de la clase III y los productos implantables (art. 61(11)(2) MDR).
[224] Y debe basarse en un plan de seguimiento poscomercialización cuyos requisitos se establecen en la sección 1.1 del anexo III MDR.
[225] Stöger et al., «Legal aspects of data cleansing in medical AI», 8.
[226] Ello requiere de evitar formas de opacidad deliberada, como la que habitualmente se produce respecto de las bases de datos que sirven para el entrenamiento y validación de estos sistemas de IA sanitaria. Rajpurkar et al., «AI in health

poscomercial en el MDR e IVDR establecen unos requerimientos de registro y documentación exhaustivos que contribuyen a establecer un control transparente del desarrollo y funcionamiento de los sistemas de IA. No obstante, las obligaciones de documentación son excesivamente genéricas y dejan un amplio margen de interpretación a la hora de abarcar determinadas funcionalidades de los sistemas de IA que pueden ser determinantes a la hora de asegurar un funcionamiento seguro de los mismos; entre otros, los efectos de las distintas formas de opacidad que pueden afectar al sistema en su implementación.

También en este sentido, el funcionamiento seguro de los sistemas de IA depende en gran medida de la coordinación entre los intervinientes en fases de diseño y desarrollo y en su implementación en el mundo real[227]. El modelo normativo del MDR e IVDR centra las obligaciones jurídicas y la responsabilidad en los desarrolladores o fabricantes, sin embargo, el papel de los usuarios de los sistemas y la coordinación entre ambos está muy limitada en el contexto del seguimiento poscomercialización actual[228].

Por último, si nos fijamos en la regulación de esta normativa de la interacción humana y producto sanitario -sistema de IA, en lo que a nosotros interesa-, parece que el modelo regulatorio actual resulta muy poco satisfactorio en este sentido. Entre las obligaciones particulares para los sistemas informáticos no se incluye ninguna disposición sobre la interfaz humano-máquina[229]. El MDCG sí incluye los factores humanos como un elemento a considerar en la evaluación del rendimiento técnico de un software sanitario[230], aunque el alcance de esta recomendación en una guía del MDCG parece, una vez más, limitado dada la importancia que la interacción humano-IA adquiere en el funcionamiento seguro de estos sistemas[231].

En definitiva, puede concluirse que el modelo europeo actual de regulación de los productos sanitarios resulta insuficiente para hacer frente a los desafíos que plantean los sistemas de IA en salud. Ello, por sí, podría comprometer el cumplimiento de los objetivos normativos de los reglamentos MDR e IVDR, a saber, garantizar un elevado nivel de seguridad y de protección de la salud y el buen funcionamiento del mercado interior. Además, si atendemos a la intersección normativa entre los productos sanitarios y la protección de datos, no existe una coordinación/comunicación normativa entre los responsables intervinientes en las fases de desarrollo -fabricante de productos- y de implementación -responsable del tratamiento- que resulte satisfactoria para la gobernanza de los sistemas de IA en todo su ciclo de vida. Y, por último, si consideramos en particular la regulación de la interacción humano-máquina en el desarrollo de sistemas de IA, la normativa de productos sanitarios no ofrece obligaciones jurídicas que puedan tener un impacto suficiente para complementar las carencias de la normativa de protección de datos para garantizar la supervisión humana desde el diseño y desarrollo de los sistemas de IA.

3.2. Reglamento de IA: tendiendo puentes entre la regulación del desarrollo de los sistemas de IA (MDR/IVDR) y su implementación (RGPD)

---

and medicine», 31. Así como de afrontar estrategias de validación satisfactorias para sistemas de IA que no resulten interpretables, vid. Hall y Ordish, «Regulating algorithms in healthcare. The GDPR and IVDR in practice».

[227] Kiseleva, «AI as a Medical Device: Is it Enough to Ensure Performance Transparency and Accountability?», 2.

[228] En este sentido, también se ha reclamado un diálogo más fluido y mayor coordinación entre las autoridades de control de protección de datos, los organismos notificados de productos sanitarios e incluso los comités de ética implicados para mejorar el funcionamiento y seguridad de los sistemas de IA concebidos como productos sanitarios. Meszaros, Compagnucci, y Minssen, «The Interaction of the Medical Device Regulation and the GDPR: Do European Rules on Privacy and Scientific Research Impair the Safety and Performance of AI Medical Devices?», 89.

[229] Vid. Anexo I (17) MDR

[230] El MDCG entiende por "human factors engineering" la aplicación de los conocimientos sobre el comportamiento humano, las capacidades, las limitaciones y otras características al diseño y a las interacciones con un software sanitaria para lograr una adecuada "usabilidad". Entendida, a su vez, como la característica de la interfaz de usuario que establece la eficacia, la eficiencia y la facilidad de aprendizaje del usuario y su satisfacción. Grupo de Coordinación de Productos Sanitarios (MDGC), «Guidance on Clinical Evaluation (MDR)/Performance Evaluation (IVDR) of Medical Device Software (MDCG 2020-1)», 10.

[231] Debe tenerse en cuenta que el rendimiento de los sistemas de IA dependerá en gran medida de la forma en que los seres humanos proporcionen datos de entrada e interpreten los datos de salida, con lo cual, cobra relevancia normativa la comprobación de los factores humanos y la formación adecuada de los usuarios humanos de los sistemas de IA médica. Vid. Gerke et al., «The need for a system view to regulate artificial intelligence/machine learning-based software as medical device». Además, la utilidad de la colaboración entre humanos e inteligencia artificial dependerá probablemente de las características específicas de la tarea y del contexto clínico. de Hond et al., «Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review», 9; Rajpurkar et al., «AI in health and medicine», 34.

3.3. Una vez identificadas las limitaciones de la normativa de protección de datos para garantizar la supervisión humana de los sistemas de IA durante todo su ciclo de vida, considerando también la normativa de productos sanitarios aplicable en fase de diseño y desarrollo de sistemas de IA con fines de salud, resulta oportuno preguntarse por el encaje del Reglamento de IA en esta cuestión. Debe tenerse en cuenta que, a efectos de vigencia normativa, las obligaciones relativas a los sistemas de IA de alto riesgo en relación con normativa sectorial -como es el caso-, no serán aplicables hasta agosto de 2027. A continuación, se expondrá cómo este Reglamento se introduce directamente en la intersección normativa entre la protección de datos y los productos sanitarios en lo que se refiere al objeto de estudio de esta investigación.

Aunque ya se ha introducido anteriormente en esta investigación, resulta oportuno recordar y ampliar algunas características básicas del RIA. La primera normativa sobre inteligencia artificial a nivel europeo[232], distingue entre distintos niveles de riesgoy centra el grueso de su regulación en los sistemas de alto riesgo a través del establecimiento de unos requisitos de obligado cumplimiento para el diseño y desarrollo de estos sistemas de forma previa a su comercialización en el mercado europeo.

Para ello, desarrolla una serie de requisitos obligatorios para los sistemas de IA de alto riesgo[233], que resultarán a su vez aplicables para productos sanitarios de IA como veremos más adelante. Aunque desarrolla otras figuras jurídicas como la de importadores o distribuidores, la más relevante en lo que respecta al cumplimiento con los requisitos obligatorios y su demostración recae sobre el proveedor de sistemas[234]. Asimismo, es también relevante para este análisis la figura del responsable del despliegue[235], no solo por las obligaciones que recaen sobre el mismo en este reglamento, sino por las obligaciones que derivan de otras normativas en el uso y despliegue de los sistemas. De especial relevancia en este marco es la normativa de protección de datos, bajo la cual el responsable del despliegue de un sistema de IA de alto riesgo será, a su vez, responsable del tratamiento para la toma de decisiones automatizada basada en las decisiones de dicho sistema[236].

El modelo de gobernanza por el que el Reglamento opta para los sistemas de IA de alto riesgo se basa en la asunción de la responsabilidad por parte del proveedor mediante un sistema de gestión de riesgos exhaustivamente documentado que se somete a la evaluación de conformidad previa a la introducción en el mercado o puesta en servicio del sistema[237]. Es un modelo de gobernanza que, por tanto, guarda ciertas similitudes con las líneas generales de

[232] Aunque parezca una cuestión menor, uno de los temas más discutidos durante la tramitación legislativa del Reglamento fue la propia definición de IA. Vid. al respecto, Viljanen, «Is EU regulating the right AI?». Disponible en: https://blogit.utu.fi/utu/2022/03/14/is-eu-regulating-the-right-ai/

[233] En resumen, se establece la obligación de establecer, implantar, documentar y mantener un sistema de gestión de riesgos durante todo el ciclo de vida de un sistema de IA de alto riesgo (art. 9 RIA); obligaciones sobre datos y gobernanza de datos destinadas a mantener criterios de calidad en el entrenamiento, validación y prueba de los sistemas (art. 10 RIA); obligaciones de documentación técnica que permitan la demostración del cumplimiento normativo (art. 11 RIA); obligaciones de registro automatizado para garantizar la trazabilidad del funcionamiento del sistema a lo largo de todo el ciclo vital del sistema (art. 12 RIA); obligaciones de transparencia y comunicación de información a los usuarios para garantizar un uso correcto de la información de salida en el uso y despliegue de los sistemas (art. 13 RIA); obligaciones de supervisión humana para garantizar que los sistemas sean diseñados y desarrollados de forma que puedan ser supervisados de manera efectiva por personas físicas en su uso y despliegue (art. 14 RIA); y obligaciones relativas a alcanzar un nivel adecuado de precisión, solidez y ciberseguridad y funcionen de manera consistente en esos sentidos durante todo su ciclo de vida (art. 15 RIA).

[234] Art. 3(2) RIA, «Proveedor»: *una persona física o jurídica, autoridad pública, órgano u organismo que desarrolle un sistema de IA o un modelo de IA de uso general o para el que se desarrolle un sistema de IA o un modelo de IA de uso general y lo introduzca en el mercado o ponga en servicio el sistema de IA con su propio nombre o marca, previo pago o gratuitamente.*

[235] «Responsable del despliegue»: *una persona física o jurídica, o autoridad pública, órgano u organismo que utilice un sistema de IA bajo su propia autoridad, salvo cuando su uso se enmarque en una actividad personal de carácter no profesional.*

[236] Se trata de una equivalencia de la que el RIA es plenamente consciente como puede comprobarse, por ejemplo, en el artículo 26(9) a la que me referiré con mayor detalle más adelante.

[237] Para van Kolfschooten esto resulta, por un lado, contradictorio con otros instrumentos de la UE que regulan los productos y servicios en el mercado interior, donde la posición de los usuarios finales es mucho más central. Por otro lado, esto es especialmente perjudicial en el contexto clínico, ya que los pacientes son particularmente susceptibles a los riesgos de la IA debido a la dependencia inherente y a las asimetrías de información en la relación médico-paciente, lo cual supondría un riesgo un riesgo de cosificación de los pacientes contraria a la dignidad de los mismos protegida por el ordenamiento jurídico, en Kolfschooten, «EU regulation of artificial intelligence: Challenges for patients' rights», 107.

los modelos de gestión de riesgos basados en la responsabilidad de las normativas de productos sanitarios y de protección de datos.

Más arriba se ha señalado que el RIA se introduce directamente en la intersección normativa entre la protección de datos y los productos sanitarios en lo que se refiere al objeto de estudio de esta investigación, ¿por qué?

Por un lado, el Reglamento de IA trata de reforzar la armonía de esta regulación con la legislación de la UE existente. Aunque elabora una lista ad hoc de sistemas que considera de alto riesgo -anexo III-, la mayoría de sistemas de alto riesgo son concebidos como tales por tratarse de un producto -o el componente de seguridad de un producto- que, conforme a la legislación de la UE recogida en el anexo I del texto, debe someterse a una evaluación de la conformidad realizada por un organismo independiente para su introducción en el mercado o puesta en servicio (art. 6(1)(b) RIA). Entre la legislación recogida en el anexo I encontramos el MDR y el IVDR[238]. Ello se traduce en que los productos sanitarios regulados por el MDR e IVDR son clasificados como de alto riesgo por el RIA o, lo que es lo mismo, que el ámbito de aplicación del MDR e IVDR para los programas informáticos coincide con el ámbito de aplicación de los sistemas de IA sanitarios de alto riesgo en el RIA -siempre que dicho programa informático cumpla con la definición de sistema de IA en el Reglamento de IA-[239]. Este esfuerzo de armonización tiene también su reflejo en la evaluación de conformidad, tratando de evitar duplicidades[240], aunque la integración de los requerimientos de ambas normas resulta compleja[241], y es por ello necesaria la aportación de guías que aporten seguridad jurídica a la aplicación normativa[242].

En definitiva, se configura entre ambos un marco regulatorio *ex ante* cuyo objetivo es garantizar que solo se comercialicen (o se pongan en servicio) dispositivos médicos basados en IA seguros y eficaces[243].

Por otro lado, la relación entre la normativa de protección de datos y el Reglamento de IA es notoria, en tanto el desarrollo y la utilización de sistemas de IA implicarán en la mayoría de casos el tratamiento de datos personales[244]. En cuanto al objeto de estudio de esta investigación, tratándose en todo caso de sistemas de IA que elaboran perfiles sobre el estado de salud de personas físicas y, por tanto, se produce un tratamiento de datos personales, la coincidencia es total. Resulta por tanto necesario examinar cómo se interrelaciona el

[238] Anexo I. Sección A – Lista de la legislación de armonización de la Unión basada en el nuevo marco legislativo. (…) *11.Reglamento (UE) 2017/745 del Parlamento Europeo y del Consejo, de 5 de abril de 2017, sobre los productos sanitarios, por el que se modifican la Directiva 2001/83/CE, el Reglamento (CE) n.º 178/2002 y el Reglamento (CE) n.º 1223/2009 y por el que se derogan las Directivas 90/385/CEE y 93/42/CEE del Consejo (DO L 117 de 5.5.2017, p. 1); 12.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 (DO L 117 de 5.5.2017, p. 176).*

[239] Kolfschooten, «EU regulation of artificial intelligence: Challenges for patients' rights», 106. A excepción de los programas informáticos que pudieren clasificarse como clase I, dado que estos se certifican bajo la exclusiva responsabilidad de los fabricantes. Considerando 60 MDR: *Los procedimientos de evaluación de la conformidad para los productos de la clase I deben llevarse a cabo, generalmente, bajo la exclusiva responsabilidad de los fabricantes, dado el bajo grado de vulnerabilidad asociado a estos productos. En el caso de los productos de las clases IIa, IIb y III, debe ser obligatorio un nivel apropiado de intervención de un organismo notificado*. Tal y como se ha señalado anteriormente, todo programa destinado a proporcionar información que se utiliza para tomar decisiones con fines terapéuticos o de diagnóstico debe ser clasificado, al menos, como clase IIa.

[240] Considerando 124 RIA: Para reducir al mínimo la carga que deben soportar los operadores y evitar posibles duplicidades, conviene, en el caso de los sistemas de IA de alto riesgo asociados a productos regulados por la legislación de armonización de la Unión vigente basada en el nuevo marco legislativo, que la conformidad de dichos sistemas de IA con los requisitos establecidos en el presente Reglamento se evalúe en el marco de la evaluación de la conformidad ya prevista en dicha legislación.*(…)*.

[241] Vid. Gennari, «O Complementarity, Where Art Thou? Wading through the Medical Device Regulation and the AI Act Compliance: The Case of Software As a Medical Device. A Primer».

[242] En este sentido, muy relevante la primera versión -y las futuras- de la guía conjunta del Comité Europeo de Inteligencia Artificial y del Grupo de Coordinación de Productos Sanitarios. Vid. Comité Europeo de Inteligencia Artificial (AIB) y Grupo de Coordinación de Productos Sanitarios (MDCG) «Interplay between the Medical Devices Regulation (MDR) & In vitro Diagnostic Medical Devices Regulation (IVDR) and the Artificial Intelligence Act (AIA) (AIB 2025-1 / MDCG 2025-6)».

[243] Parziale, Andrea. « AI-powered Medical Devices and the Development Risk Defense Under the Revised Product Liability Directive», 283.

[244] Comité Europeo de Protección de Datos (CEPD) y Supervisor Europeo de Protección de Datos (SEPD), «Dictamen conjunto 5/2021 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial)», 9.

cumplimiento de las distintas obligaciones impuestas por ambos reglamentos a los distintos actores involucrados en el ciclo de vida de los sistemas de IA de alto riesgo que elaboren perfiles con fines médicos.

A continuación, se indagará en la particular relación entre la supervisión humana como requerimiento de obligado cumplimiento en el Reglamento de IA y la aplicación de los mecanismos de intervención humana del artículo 22 RGPD y las disposiciones relativas a esta intervención en la EIPD. Y finalmente, se analizarán también las nuevas obligaciones del RIA que afectan a los responsables del despliegue, figura que se corresponderá habitualmente con el responsable del tratamiento en la fase de uso del sistema como veremos.

3.3.1. Obligaciones para el proveedor y el requisito de supervisión humana en el RIA a la luz de la intervención humana requerida por el RGPD

El reconocimiento de requerimientos de obligado cumplimiento por esta norma que complementen las obligaciones ya establecidas para la certificación de sistemas bajo el MDR e IVDR, contribuirá a paliar algunas de las limitaciones arriba indicadas[245]. En este epígrafe se analizará con detalle el impacto que podría tener el reconocimiento de la supervisión humana como uno de estos requerimientos de obligado cumplimiento.

El artículo 14 del RIA dispone que los sistemas de alto riesgo deberán diseñarse y desarrollarse de modo que puedan ser supervisados de forma efectiva por personas físicas durante su fase de uso (art. 14(1) RIA). El fundamento de esta supervisión efectiva está en prevenir o reducir al mínimo los riesgos para la salud, la seguridad o los derechos fundamentales, tanto cuando se utiliza un sistema de IA conforme a su finalidad prevista o se le da un uso indebido *razonablemente previsible* (art. 14(2) RIA)[246]. Así, el Reglamento obliga al proveedor a adoptar determinadas medidas para garantizar esta supervisión desde el diseño y desarrollo del modelo, o bien definiéndolas e integrándolas en el sistema de IA -cuando sea técnicamente posible-, o bien definiéndolas para que las lleve a cabo el propio responsable del despliegue del sistema (art. 14(3) RIA).

Las "cualidades" a partir de las cuales se establece el concepto de "supervisión efectiva" que permita al proveedor cumplir con dicho requisito, también vienen definidas en el articulado; así, las personas a quienes se encomiende la supervisión del sistema deben ser capaces de, entre otros, entender por completo las capacidades y limitaciones del sistema, ser conscientes del sesgo de automatización, interpretar correctamente la información de salida del sistema, desestimar, invalidar o revertir dicha información o interrumpir el sistema accionando un botón específicamente destinado a tal fin (art. 14(4) RIA)[247].

La doctrina resalta que la integración de enfoques centrados en el ser humano y en el usuario a lo largo de todo el proceso de desarrollo de la IA permitirá diseñar algoritmos de IA que reflejen mejor las necesidades y la cultura del personal sanitario, al tiempo que permitirá identificar y abordar los posibles riesgos en una fase temprana[248]. Aunque, a estos efectos, no puede dejarse de lado la importancia del entrenamiento y formación de los propios seres humanos -

[245] Por ejemplo, el requisito de gobernanza de datos incluye la realización de prácticas adecuadas sobre los conjuntos de datos de entrenamiento, validación y prueba, entre otras, la elección de un diseño adecuado, recopilación de datos, actividades de preprocesamiento para la limpieza de los datos, evaluación previa de los datos disponibles o examen de sesgos. Esta disposición añade obligaciones específicas y que pueden resultar críticas para el funcionamiento seguro de los sistemas de IA sanitario y que, en el mejor de los casos bajo la actual normativa de productos sanitarios, podrían considerarse prácticas adecuadas-voluntarias para la gestión del riesgo.

[246] Esta previsibilidad se vincula necesariamente con el sistema de gestión de riesgos establecido por el RIA, otorgando además a la supervisión humana un carácter de salvaguarda última cuando este apartado añade: *en particular cuando dichos riesgos persisten a pesar de aplicar otros requisitos establecidos en el presente capítulo*.

[247] Para estos fines será necesario el establecimiento de mecanismos de validación que permitan al proveedor realizar validación en entornos clínicos reales. A estos efectos, pueden resultar determinantes: (1) los espacios controlados de pruebas para la IA concebidos en los artículos 58-60 RIA y (2) el Espacio Europeo de Datos de Salud (EEDS). Kiseleva y de Hert reclaman el desarrollo conjunto de la normativa sobre EEDS y RIA en estos aspectos, sin olvidar que la normativa sectorial sobre productos sanitarios añadirá requerimientos específicos a este marco normativo. Kiseleva y de Hert, «Creating a European Health Data Space: Obstacles in Four Key Legal Areas», 14.

[248] Lekadir et al., «Artificial intelligence in healthcare: Applications, risks, and ethical and societal impacts», 8.

personal clínico en este caso- para el uso de estos sistemas[249], lo cual conlleva establecer también ulteriores obligaciones como veremos.

Por el momento, es relevante poder analizar la interrelación de las obligaciones establecidas para el proveedor en la fase de desarrollo del sistema en materia de supervisión humana, con los mecanismos de gobernanza basados en la intervención humana del RGPD para el caso de sistemas de IA que realicen perfiles sobre el tratamiento de datos personales, que establecen obligaciones para el responsable del tratamiento en la fase de uso del sistema. Debe también resaltarse que este último -el responsable del tratamiento en fase de uso según RGPD-, coincidirá habitualmente con el denominado responsable del despliegue en el RIA; esto es, quien utilice un sistema de IA bajo su propia autoridad (art. 3(4) RIA), será también por lo general quien determine los fines y medios del tratamiento (art. 4(7) RGPD) en su utilización.

Por un lado, es muy relevante la conexión entre las medidas de transparencia y las de supervisión humana en el RIA[250]. Las medidas de supervisión humana se incluyen específicamente entre la información que los proveedores deben proporcionar a los usuarios/responsables en virtud del requisito de transparencia, incluidas las medidas técnicas establecidas para facilitar la interpretación de los resultados de los sistemas de IA por parte de los usuarios (art. 13(3)(d) RIA). Además, se establece la obligación expresa para responsables del despliegue de, cuando proceda, utilizar la información facilitada conforme al artículo 13 del RIA para cumplir la obligación de llevar a cabo una evaluación de impacto relativa a la protección de datos conforme al artículo 35 RGPD (art. 26(9) RIA).

Por lo tanto, en virtud del RIA, los responsables del despliegue -y del tratamiento- deberán utilizar la información facilitada por los proveedores -que permite a las personas a las que se asigna la supervisión humana por mandato del artículo 22 RGPD comprender las capacidades y limitaciones del sistema- para cumplir, a su vez, con el mandato del artículo 35 del RGPD. A mi modo de ver, el RIA ampliará las posibilidades de que los responsables del despliegue proporcionen y demuestren una intervención humana significativa como responsables del tratamiento de datos en el RGPD. En el contexto clínico, la información facilitada por los proveedores podría resultar determinante a la hora de diseñar los protocolos clínicos que construyan la mejor colaboración humano-máquina posible.Por otro lado, las obligaciones de seguimiento poscomercialización para proveedores deben también tenerse en cuenta en este contexto. Un sistema de seguimiento posterior a la comercialización de los sistemas de IA, es decir para la fase de implementación y uso de los mismos -como el diseñado en el Reglamento de IA[251]- podría ayudar a identificar y modificar los sistemas de IA de alto riesgo que no puedan ser supervisados eficazmente por personas físicas y, por tanto, no permiten a los responsables del tratamiento cumplir con la intervención humana en virtud del artículo 22 RGPD, por no haber intervención humana significativa posible dadas las limitaciones introducidas desde el diseño y desarrollo del modelo. Al llevar a cabo una EIPD, los responsables del tratamiento recopilarán datos y pruebas en el contexto clínico sobre si es posible que la intervención humana contribuya al cumplimiento del RGPD -en los términos arriba planteados- durante el uso de un sistema de IA de alto riesgo[252]. Y, como usuarios en virtud del RIA, los responsables del tratamiento podrían compartir esa información recopilada con los desarrolladores para permitirles evaluar el cumplimiento del sistema con el requisito de supervisión humana a ojos del RIA[253], es decir, para demostrar que sus sistemas de IA de alto riesgo pueden ser supervisados eficazmente por personas físicas.

[249] Vid. Gerke, «Health AI for Good Rather Than Evil? The Need for a New Regulatory Framework for AI-Based Medical Devices», 506.
[250] Blatti, Andrea, y Tramacere, Stefano, «The Transparency and Liability Issues Associated with AI-Based Medical Systems», 298.
[251] Capítulo IX RIA. Que reforzaría, a su vez, el sistema de seguimiento poscomercial del MDR e IVDR que más arriba se ha considerado limitado para el seguimiento del funcionamiento de sistemas de IA.
[252] En este sentido, entre otras pruebas que podrían recopilarse podríamos encontrar la evaluación institucional mencionada en el apartado 2.2.3. Diseño de una intervención humana demostrable para la toma de decisiones clínica, recogida de Cabitza, Campagner, y Datteri, «To Err is (only) Human. Reflections on How to Move from Accuracy to Trust for Medical AI».
[253] Art. 72(2) RIA: *El sistema de vigilancia poscomercialización recopilará, documentará y analizará de manera activa y sistemática los datos pertinentes que pueden facilitar los responsables del despliegue o que pueden recopilarse a través de otras fuentes sobre el funcionamiento de los sistemas de IA de alto riesgo durante toda su vida útil, y que*

3.3.2. Obligaciones para responsables del despliegue en relación con la supervisión humana en el RIA y con la intervención humana requerida por el RGPD

La versión inicial del Reglamento de IA propuesto por la Comisión en abril de 2021 no incluyó obligaciones de relevancia para quienes utilizasen un sistema de IA bajo su propia autoridad, esto es, para los responsables del despliegue. No obstante, a lo largo de su tramitación legislativa, resultó necesaria la ampliación de dichas obligaciones no solo a la fase de uso -el seguimiento poscomercialización afecta a dicha fase-, sino también a los responsables del despliegue -y no únicamente a los proveedores-. De esta forma, algunas de las obligaciones establecidas en el artículo 26 RIA, tienen una notable repercusión en la intersección entre las obligaciones de supervisión humana en el artículo 14 RIA para los proveedores, y las que recaen también sobre los responsables del despliegue en virtud de los artículos 22 y 35 del RGPD -aquí como responsables del tratamiento-. Se trata, fundamentalmente, de tres obligaciones:

- Los responsables del despliegue deben implementar medidas técnicas y organizativas adecuadas para garantizar que los sistemas sean supervisados de acuerdo a las instrucciones y medidas implementadas por el proveedor (art. 26(1) RIA). Es decir, se obliga a los responsables del despliegue a aplicar las medidas de supervisión humana indicadas por el proveedor y según las instrucciones específicas[254]. Teniendo en cuenta que algunas de las medidas de supervisión humana adoptadas por los proveedores pueden ser simplemente medidas a poner en práctica por el responsable del despliegue (art. 14(3)(b) RIA), esta obligación para el responsable es simplemente consecuencia de la anterior.

- La supervisión humana debe ser encomendada a personas físicas que tengan la competencia, la formación y la autoridad necesarias, y con el apoyo necesario del responsable del despliegue (art. 26(2) RIA). Esta obligación es en realidad la cristalización normativa de los criterios ya establecidos para la intervención humana significativa, como hemos podido ver, pero extendida a cualquier tipo de uso de sistemas de IA de alto riesgo.

- El cumplimiento de la supervisión humana por diseño del artículo 14 RIA no debe afectar al cumplimiento de obligaciones normativas nacionales y de la UE relativas a la toma de decisiones en la fase de uso, ni a la libertad para organizar los recursos y actividades del responsable del despliegue (art. 26(3) RIA). Esta es una alusión directa, por ejemplo, a las disposiciones del artículo 22 RGPD que afectarán a la toma de decisiones automatizada basada en sistemas de IA de alto riesgo, supuesto que es precisamente el que se ha trabajado en la presente investigación.

Con sus limitaciones, el modelo de regulación para los sistemas de IA propuesto por lel RIA en combinación con el RGPD refuerza la gobernanza y supervisión humana de los procesos de toma de decisiones automatizada basada en la elaboración de perfiles en dos direcciones distintas[255].

Primero en lo que se refiere a las obligaciones de los proveedores. Por un lado, el establecimiento de la supervisión humana como un requisito de obligado cumplimiento desde el diseño y desarrollo de los modelos -para los desarrolladores- podrá contribuir de forma efectiva al cumplimiento normativo -por parte de los responsables del tratamiento- y su demostración -a través de la EIPD- de los mecanismos de gobernanza basados en la intervención humana

---

permiten al proveedor evaluar el cumplimiento permanente de los requisitos establecidos en el capítulo III, sección 2, por parte de los sistemas de IA. (…)

[254] Schwemer, Tomada, y Pasini, «Legal AI Systems in the EU's proposed Artificial Intelligence Act», 7.

[255] En esta línea, una editorial en The Lancet Digital Health reflexionaba sobre la distribución de la responsabilidad para los sistemas de IA en salud; mientras que parte de la literatura entiende que los desarrolladores de IA son responsables de proporcionar orientación sobre la gestión de sus herramientas, incluyendo cómo y cuándo comprobar el rendimiento del sistema, e identificar las vulnerabilidades que puedan surgir después de su puesta en práctica; otra parte sostiene que no toda la responsabilidad recae en los desarrolladores de IA, y que los proveedores de salud deben probar los sistemas de IA con nuevos datos para verificar su utilidad y evaluar las posibles vulnerabilidades. Parece que el RIA encaja en esta segunda visión. Vid. The Lancet Digital Health, «Holding artificial intelligence to account».

aplicables a la fase de implementación del sistema -art. 22 RGPD-. Por otro lado, un sistema de seguimiento posterior a la comercialización que promueva la comunicación de los datos recabados en la fase de implementación del sistema -por parte del responsable del tratamiento a través de la EIPD- permitirá al proveedor evaluar el cumplimiento -por parte de los proveedores- de los requisitos de obligado cumplimiento -entre otros, de la supervisión humana- para el desarrollo y sistemas de IA de alto riesgo durante todo el ciclo de vida del sistema.

Segundo, en relación con las obligaciones para responsables del despliegue, es decir, con los responsables del tratamiento en el uso de estos sistemas, fase en la que las obligaciones del RIA y del RGPD se deben aplicar conjuntamente. Por ello, el propio RIA establece que el cumplimiento del requisito de supervisión humana del RIA no debe afectar al cumplimiento de obligaciones normativas nacionales y de la UE relativas a la toma de decisiones en la fase de uso, ni a la libertad para organizar los recursos y actividades del responsable del despliegue. Y las obligaciones establecidas aquí para los responsables del despliegue encajan además perfectamente con el cumplimiento de la intervención humana significativa y su relación con el principio de responsabilidad proactiva descrito anteriormente. Por un lado, con la obligación de adoptar medidas por los responsables que permitan a los clínicos un uso de los sistemas de IA conforme al diseño llevado a cabo por los proveedores. Y, por otro lado, encomendando la labor de supervisión/intervención humana a personal clínico con la competencia, la formación y la autoridad necesarias en función del tipo de sistema de IA de alto riesgo utilizado.

En definitiva, mitigando en múltiples aspectos las limitaciones señaladas en la intersección normativa vigente entre la protección de datos y los productos sanitarios para la regulación de los sistemas de IA con fines de salud destinados a ser utilizados en el contexto clínico.

4. Conclusiones

La implementación de sistemas de IA en la asistencia sanitaria podría ser un factor determinante en la mejora de la salud y bienestar general de la población mundial. No obstante, este momento incipiente de su implementación es crucial para abordar las cuestiones relativas a su seguridad y buen funcionamiento[256]. A su vez, las distintas iniciativas regulatorias de la IA han puesto de manifiesto que es indispensable garantizar la supervisión humana durante todo el ciclo de vida de los sistemas de IA para prevenir o reducir al mínimo los riesgos para la salud, la seguridad o los derechos fundamentales.

Este es el marco de partida de esta investigación jurídica sobre la supervisión humana los sistemas de IA que, a partir del tratamiento de datos personales, arrojan inferencias sobre el estado de salud de los pacientes y pueden utilizarse para fines de diagnóstico, prognosis o de elección terapéutica como apoyo a la toma de decisiones en la asistencia sanitaria.

El primero de los resultados de este trabajo pone de manifiesto el papel central que el Reglamento General de Protección de Datos debe ocupar en este contexto. La prohibición de la toma de decisiones basada únicamente en el tratamiento automatizado exige al responsable del tratamiento la introducción de una intervención humana significativa -del personal clínico-, y no meramente formal, en el proceso decisorio previa a la producción de efectos jurídicos o significativos en el interesado -paciente-.

A partir del análisis de las Directrices sobre decisiones individuales automatizadas y elaboración de perfiles refrendadas por el Comité Europeo de Protección de Datos[257], esta investigación concluye que el término significativo en el contexto asistencial debe traducirse en: [1] intervención humana por personal clínico con autoridad y competencia para modificar el resultado algorítmico; [2] la aplicación rutinaria de los resultados algorítmicos conlleva una dejación de la responsabilidad incompatible con la normativa de protección de datos, por ende, la intervención humana del personal clínico debe tener una influencia real sobre el proceso decisorio evitando que éste se convierta, de facto, en un sistema plenamente automatizado; [3] la intervención humana está fundamentada en que el responsable del tratamiento pueda

[256] Liu et al., «The medical algorithmic audit».
[257] Grupo de Trabajo sobre Protección de Datos del Artículo 29, «Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679».

mantener la responsabilidad sobre el tratamiento automatizado de datos personales y, por ende, teniendo en cuenta el contexto clínico que rodea al paciente en su sentido más amplio, la intervención del personal clínico al cumplimiento normativo y su demostración. Este análisis no es más que una aproximación al concepto de intervención humana significativa en el contexto asistencial, el trabajo jurídico por desarrollar aquí es todavía muy amplio.

No obstante, esta investigación concluye que la intervención humana se trata únicamente de un mecanismo de gobernanza contenido en el artículo 22, sino también de una medida organizativa parte del sistema de gestión de riesgos al que obliga el RGPD sobre la base del principio de responsabilidad proactiva. En particular, la EIPD obliga al responsable del tratamiento a justificar cómo introduce la intervención humana en el proceso decisorio y a demostrar su eficacia como medida organizativa. Aquí, para la construcción de procesos decisorio con el mejor modelo de colaboración humano-máquina, se destaca el potencial del establecimiento de protocolos clínicos que incorporen en su metodología la aplicación de medidas organizativas para la gestión del riesgo en el tratamiento de datos personales. Una vez más, el análisis en este aspecto solo puede considerarse una primera aproximación a la materia.

En definitiva, partiendo del principio de responsabilidad, la intervención humana en el RGPD como mecanismo de gobernanza del artículo 22 y medida organizativa en la gestión del riesgo, ofrece un marco flexible y garantista para la supervisión humana de sistemas de IA con fines de salud en el contexto asistencial. Ahora bien, la principal limitación de la normativa de protección de datos es que no ofrece garantías para la supervisión humana de estos sistemas durante todo el ciclo de vida de IA, dado que su fase de diseño y desarrollo no se aborda de forma específica.

Por ello, esta investigación repara en la normativa europea de productos sanitarios -MDR e IVDR- para analizar si la legislación vigente ofrece desde esta perspectiva sectorial soluciones para las limitaciones del RGPD a la hora de garantizar la supervisión humana desde el diseño y desarrollo de los sistemas de IA. Sin embargo, el sistema de certificación y seguimiento poscomercialización de los productos sanitarios tampoco resulta satisfactorio y, en particular, entre las obligaciones particulares para los sistemas informáticos no se incluye ninguna disposición sobre la interfaz humano-máquina que pueda reforzar la supervisión humana de los sistemas de IA desde fases tempranas.

Por último, se evalúa la repercusión que tendrá el Reglamento de IA en este contexto normativo. Los resultados en este sentido son positivos con la inclusión de la supervisión humana como requisito de obligado cumplimiento para sistemas de IA de alto riesgo. El modelo de gobernanza del RIA analizado permite concluir que el cumplimiento por los proveedores del requisito de supervisión humana en fase de desarrollo contribuirá al cumplimiento normativo y su demostración, por parte del responsable del despliegue/tratamiento que implante el sistema de IA, de los mandatos del artículo 22 y 35 RGPD sobre intervención humana para procesos de toma de decisiones automatizada. Es decir, contribuirá a garantizar una supervisión humana -desde el diseño y desarrollo y durante todo su ciclo de vida- de los sistemas de IA con fines de salud que se implementen en procesos de decisión asistenciales.

Para finalizar, a pesar del potencial de la intersección entre las normativas de protección de datos, productos sanitarios y sistemas de IA para garantizar una supervisión humana durante todo el ciclo de vida del sistema, me gustaría profundizar en una crítica que ha sobrevolado en cierta medida el análisis jurídico realizado. Conforme al mismo, los 3 modelos normativos analizados comparten -en líneas generales- la centralidad del modelo de responsabilidad sobre el cumplimiento normativo y su demostración.

La introducción de la responsabilidad sobre el cumplimiento normativo y su demostración como principio nuclear del RGPD ha consolidado el control colectivo sobre la justificación de los procesos que recopilan, procesan y usan los datos personales. En esta investigación se ha resaltado el valor de este control colectivo para los procesos de toma de decisiones automatizada basados en la elaboración de perfiles, frente a modelos normativos de control individual que desplazan la responsabilidad a personas de por sí atomizadas por estos procesos. No obstante, el sistema normativo de responsabilidad del RGPD puede resultar

insuficiente, especialmente cuando el control colectivo se aleja de mecanismos directos y abiertos a la sociedad civil[258].

Lo mismo podría decirse de la normativa de productos sanitarios vigente y, en este sentido, el Reglamento de IA corre el riesgo de calcificar algunos de los defectos ya existentes en el modelo normativo de responsabilidad adoptado por el RGPD. En particular, si la gestión de riesgos y el proceso de demostración del cumplimiento normativo -en particular sobre los requerimientos de obligado cumplimiento- no es accesible en modo alguno a mecanismos directos y abiertos a la sociedad civil. Esto podría ser particularmente pernicioso para los grupos más vulnerables y expuestos a las formas de discriminación implícitas en el funcionamiento de los sistemas de IA.

A estos efectos, resulta necesario reforzar la publicidad [transparencia] para favorecer la deliberación colectiva [permeabilidad] de estos modelos normativos durante el ciclo de vida completo de los sistemas de IA y de su tratamiento de datos personales. Ello no debe traducirse en la publicidad de todo el proceso de cumplimiento normativo y de su demostración, sino únicamente de los aspectos que resulten determinantes para comprender, por un lado, los aspectos fundamentales del diseño y desarrollo de los sistemas de IA de alto riesgo en relación con el contexto particular en el que van a ser implementados y, por otro, la justificación del tratamiento de datos personales en el contexto particular en el que se implementan los procesos de toma de decisiones automatizados.

Con dicho objetivo, creo que podría resultar de extraordinaria utilidad la disponibilidad de acceso por la sociedad civil a espacios controlados de pruebas. De forma que no se comprometa ningún derecho e interés propio -de desarrolladores o responsables del tratamiento según la fase del ciclo de vida del sistema- ni de terceros sobre el proceso de cumplimiento normativo y su demostración. Pero, a su vez, se permita el testeo y acceso a sus resultados del funcionamiento de los sistemas de alto riesgo en contextos particulares en los que vayan a ser implementados, pudiendo la sociedad civil examinar el funcionamiento de los sistemas sobre grupos vulnerables.

[258] El proceso de demostración del cumplimiento al que obliga el RGPD puede ser controlado de forma indirecta por autoridades competentes y judiciales, pero no prevé mecanismos de control directo sobre el mismo que permitan una deliberación más abierta sobre estos procesos de tratamiento.

Bibliografía

Agencia de los Derechos Fundamentales de la Unión Europea (FRA). «Getting the future right Artificial intelligence and fundamental rights». Luxemburgo, 2020. https://doi.org/10.2811/774118.

Agencia Española de Protección de Datos (AEPD). «Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción», 2020.

———. «Gestión del riesgo y evaluación de impacto en tratamientos de datos personales», 2021.

Almada, Marco. «Human Intervention in Automated Decision-making: Toward the Construction of Contestable Systems». En *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*, 2-11. ICAIL '19. New York, NY, USA: ACM, 2019. https://doi.org/10.1145/3322640.3326699.

Beunza Nuin, Juan José, Enrique Puertas Sanz, y Emilia Condés Moreno. *Manual práctico de inteligencia artificial en entornos sanitarios*. Elsevier, 2020.

Binns, Reuben. «Human Judgment in algorithmic loops: Individual justice and automated decision-making». *Regulation & Governance*, 7 de octubre de 2020. https://doi.org/https://doi.org/10.1111/rego.12358.

Bjerring, Jens Christian, y Jacob Busch. «Artificial Intelligence and Patient-Centered Decision-Making». *Philosophy & Technology*, 2020. https://doi.org/10.1007/s13347-019-00391-6.

Blatti, Andrea, y Tramacere, Stefano. «The Transparency and Liability Issues Associated with AI-Based Medical Systems». En F. Casarosa, F. Gennari, & A. Rossi (Eds.), *Enabling and Safeguarding Personalized Medicine*. Springer Nature Switzerland (2025): 291-315. https://doi.org/10.1007/978-3-031-99709-9_15

Brennan-Marquez, Kiel, y Stephen Henderson. «Artificial Intelligence and Role-Reversible Judgment». *Journal of Criminal Law and Criminology* 109, n.º 2 (1 de enero de 2019).

Brkan, Maja. «Do Algorithms Rule the World? Algorithmic Decision-Making in the Framework of the GDPR and Beyond». *International Journal of Law and Information Technology* 27, n.º January (2019): 91-121. https://doi.org/https://doi.org/10.1093/ijlit/eay017.

Cabitza, Federico, Andrea Campagner, y Clara Balsano. «Bridging the "Last Mile" Gap between AI Implementation and Operation: "Data Awareness" That Matters». *Annals of Translational Medicine* 8, n.º 7 (abril de 2020): 501. https://doi.org/10.21037/atm.2020.03.63.

Cabitza, Federico, Andrea Campagner, y Edoardo Datteri. «To Err is (only) Human. Reflections on How to Move from Accuracy to Trust for Medical AI». En *ITAIS 2020, the XVII Conference of the Italian Chapter of AIS Organizing in a digitized world: Diversity, Equality and Inclusion*, editado por Federica Ceci, Andrea Prencipe, y Paolo Spagnoletti, 36-49. Cham: Springer International Publishing, 2021.

Cabitza, Federico, Andrea Campagner, y Luca Maria Sconfienza. «Studying human-AI collaboration protocols: the case of the Kasparov's law in radiological double reading». *Health Information Science and Systems* 9, n.º 1 (2021): 8. https://doi.org/10.1007/s13755-021-00138-8.

Cabitza, Federico, Andrea Campagner, Felipe Soares, Luis García de Guadiana-Romualdo, Feyissa Challa, Adela Sulejmani, Michela Seghezzi, y Anna Carobene. «The importance of being external. methodological insights for the external validation of machine learning models in medicine». *Computer Methods and Programs in Biomedicine* 208 (2021): 106288. https://doi.org/https://doi.org/10.1016/j.cmpb.2021.106288.

Cabitza, Federico, Davide Ciucci, y Raffaele Rasoini. «A Giant with Feet of Clay: On the Validity of the Data that Feed Machine Learning in Medicine». En *Organizing for the Digital World*,

editado por Federico Cabitza, Carlo Batini, y Massimo Magni, 121-36. Cham: Springer International Publishing, 2019.

Cabitza, Federico, Angela Locoro, Camilla Alderighi, Raffaele Rasoini, Domenico Compagnone, y Pedro Berjano. «The elephant in the record: On the multiplicity of data recording work». *Health Informatics Journal* 25, n.º 3 (2019): 475-90. https://doi.org/10.1177/1460458218824705.

Casanova Asencio, Andrea Salud. «Mecanismos de prevención del acceso indebido a la historia clínica por parte del personal sanitario y nueva legislación de protección de datos». *Bioderecho.es*, n.º 7 (31 de marzo de 2018). https://doi.org/10.6018/bioderecho.360771.

Challen, Robert, Joshua Denny, Martin Pitt, Luke Gompels, Tom Edwards, y Krasimira Tsaneva-Atanasova. «Artificial intelligence, bias and clinical safety». *BMJ Quality & Safety* 28, n.º 3 (2019): 231-37. https://doi.org/10.1136/bmjqs-2018-008370.

Cobbe, Jennifer, Michelle Seng Ah Lee, y Jatinder Singh. «Reviewable Automated Decision-Making: A Framework for Accountable Algorithmic Systems». En *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 598–609. FAccT '21. New York, NY, USA: Association for Computing Machinery, 2021. https://doi.org/10.1145/3442188.3445921.

Coiera, Enrico. «The Last Mile: Where Artificial Intelligence Meets Reality». *Journal of Medical Internet Research* 21, n.º 11 (8 de noviembre de 2019): e16323. https://doi.org/10.2196/16323.

Comisión Europea. Libro Blanco sobre la inteligencia artificial (2020).

Comité Europeo de Inteligencia Artificial (AIB) y Grupo de Coordinación de Productos Sanitarios (MDCG) «Interplay between the Medical Devices Regulation (MDR) & In vitro Diagnostic Medical Devices Regulation (IVDR) and the Artificial Intelligence Act (AIA) (AIB 2025-1 / MDCG 2025-6)», 2025.

Comité Europeo de Protección de Datos (CEPD). «Opinion 12/2019 on the draft list of the com supervisory authority of Spain petent regarding the processing operations protection exempt from the requirement of a data impact assessment (Article 35 (5) GDPR)», 2019.

Comité Europeo de Protección de Datos (CEPD), y Supervisor Europeo de Protección de Datos (SEPD). «Dictamen conjunto 5/2021 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial)», 2021.

Cotino, Lorenzo. «Riesgos e impactos del big data, la inteligencia artificial y la robótica. Enfoques, modelos y principios de la respuesta del Derecho». *Revista General de Derecho Administrativo* 50 (2019): 1-37.

Cotino, Lorenzo, José Antonio Castillo Parrilla, Idoia Salazar, Richard Benjamins, María Cumbreras, y Adaya María Esteban. «Un análisis crítico constructivo de la Propuesta de Reglamento de la Unión Europea por el que se establecen normas armonizadas sobre la Inteligencia Artificial (Artificial Intelligence Act)». *Diario La Ley*, 2021, 1-15.

Davenport, Thomas, y Ravi Kalakota. «The Potential for Artificial Intelligence in Healthcare». *Future Healthcare Journal* 6, n.º 2 (junio de 2019): 94-98. https://doi.org/10.7861/futurehosp.6-2-94.

Dick, Vincent, Christoph Sinz, Martina Mittlböck, Harald Kittler, y Philipp Tschandl. «Accuracy of Computer-Aided Diagnosis of Melanoma: A Meta-analysis». *JAMA Dermatology* 155, n.º 11 (1 de noviembre de 2019): 1291-99. https://doi.org/10.1001/jamadermatol.2019.1375.

Durán, Juan Manuel, y Karin Rolanda Jongsma. «Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI». *Journal of Medical Ethics* 47, n.º 5 (1 de mayo de 2021): 329-35. https://doi.org/10.1136/medethics-2020-106820.

Edwards, L, y M Veale. «Enslaving the Algorithm: From a "Right to an Explanation" to a "Right

to Better Decisions"?» *IEEE Security & Privacy* 16, n.º 3 (2018): 46-54. https://doi.org/10.1109/MSP.2018.2701152.

European Group on Ethics in Science and New Technologies. «Future of Work, Future of Society», 2018. https://doi.org/doi:10.2777/029088.

Favaretto, Maddalena, Eva De Clercq, y Bernice Simone Elger. «Big Data and discrimination: perils, promises and solutions. A systematic review». *Journal of Big Data* 6, n.º 1 (2019): 1-27. https://doi.org/10.1186/s40537-019-0177-4.

Ferretti, Agata, Manuel Schneider, y Alessandro Blasimme. «Machine learning in medicine: Opening the New Data Protection Black Box». *European Data Protection Law Review* 4, n.º 3 (2018): 320-32. https://doi.org/10.21552/edpl/2018/3/10.

Fosch-Villaronga, Eduard, Hadassah Drukarch, Pranav Khanna, Tessa Verhoef, y Bart Custers. «Accounting for diversity in AI for medicine». *Computer Law & Security Review* 47 (2022): 105735. https://doi.org/https://doi.org/10.1016/j.clsr.2022.105735.

Fosch-Villaronga, Eduard, Pranav Khanna, Hadassah Drukarch, y Bart H M Custers. «A human in the loop in surgery automation». *Nature Machine Intelligence* 3, n.º 5 (2021): 368-69. https://doi.org/10.1038/s42256-021-00349-4.

Freeman, Karoline, Jacqueline Dinnes, Naomi Chuchu, Yemisi Takwoingi, Sue E Bayliss, Rubeta N Matin, Abhilash Jain, Fiona M Walter, Hywel C Williams, y Jonathan J Deeks. «Algorithm based smartphone apps to assess risk of skin cancer in adults: systematic review of diagnostic accuracy studies». *BMJ* 368 (10 de febrero de 2020): m127. https://doi.org/10.1136/bmj.m127.

Freeman, Karoline, Julia Geppert, Chris Stinton, Daniel Todkill, Samantha Johnson, Aileen Clarke, y Sian Taylor-Phillips. «Use of artificial intelligence for image analysis in breast cancer screening programmes: systematic review of test accuracy». *BMJ* 374 (2 de septiembre de 2021): n1872. https://doi.org/10.1136/bmj.n1872.

Fritz, Alexis. «La transformación digital en la atención sanitaria como un reto para la autonomía y la confianza en la interacción médico-paciente». *Dilemata* 0, n.º 32 SE-Debate (26 de mayo de 2020): 17-36.

García Garmendia, J L. «Spanish influenza score: Predictive power without giving up the classic». *Medicina Intensiva (English Edition)* 45, n.º 2 (2021): 67-68. https://doi.org/10.1016/j.medine.2020.09.005.

Gardner, Rebekah L., Emily Cooper, Jacqueline Haskell, Daniel A. Harris, Sara Poplau, Philip J. Kroth, y Mark Linzer. «Physician stress and burnout: the impact of health information technology». *Journal of the American Medical Informatics Association : JAMIA* 26, n.º 2 (2019): 106-14. https://doi.org/10.1093/jamia/ocy145.

Garnacho-Montero, José, y Ignacio Martín-Loeches. «Clinical management of sepsis can be improved by artificial intelligence: no». *Intensive Care Medicine* 46, n.º 2 (2020): 378-80. https://doi.org/10.1007/s00134-020-05947-1.

Gennari, Francesca. «O Complementarity, Where Art Thou? Wading through the Medical Device Regulation and the AI Act Compliance: The Case of Software As a Medical Device. A Primer». *BioLaw Journal - Rivista Di BioDiritto*, n.º 3, (2024): 411-53. https://doi.org/10.15168/2284-4503-3207.

Gerke, Sara. «Health AI for Good Rather Than Evil? The Need for a New Regulatory Framework for AI-Based Medical Devices». *Yale Journal of Health Policy, Law and Ethics* 20, n.º 2 (2021): 433-513.

Gerke, Sara, Boris Babic, Theodoros Evgeniou, y I Glenn Cohen. «The need for a system view to regulate artificial intelligence/machine learning-based software as medical device». *npj Digital Medicine* 3, n.º 1 (2020): 53. https://doi.org/10.1038/s41746-020-0262-2.

Gil González, Elena, y Paul de Hert. «Understanding the legal provisions that allow processing and profiling of personal data—an analysis of GDPR provisions and principles». *ERA*

Forum 19, n.º 4 (2019): 597-621. https://doi.org/10.1007/s12027-018-0546-z.

Granja, Conceição, Wouter Janssen, y Monika Alise Johansen. «Factors determining the success and failure of ehealth interventions: Systematic review of the literature». *Journal of Medical Internet Research*, 2018. https://doi.org/10.2196/10235.

Grupo de Coordinación de Productos Sanitarios (MDGC). «Guidance on Clinical Evaluation (MDR)/Performance Evaluation (IVDR) of Medical Device Software (MDCG 2020-1)», 2020.

———. «Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 – MDR and Regulation (EU) 2017/746 – IVDR (MDCG 2019-11)», 2019.

Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial de la Comisión Europea (HLEG-AI). *Directrices éticas para una IA fiable*. Editado por Comisión Europea. Bruselas: Oficina de Publicaciones de la Comisión Europea, 2019. https://doi.org/doi/10.2759/14078.

Grupo de Trabajo sobre Protección de Datos del Artículo 29. «Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679». Bruselas, 2018.

———. «Directrices sobre la evaluación de impacto relativa a la protección de datos (EIPD) y para determinar si el tratamiento "entraña probablemente un alto riesgo" a efectos del Reglamento (UE) 2016/679», 2017.

Gstrein, Oskar Josef. «European AI Regulation: Brussels Effect versus Human Dignity?» *Forthcoming, Zeitschrift für Europarechtliche Studien (ZEuS)*, n.º 4 (2022): 1-24.

Hall, Alison, y Johan Ordish. «Regulating algorithms in healthcare. The GDPR and IVDR in practice», 2019.

Harvey, H, y F Cabitza. «Algorithms are the new drugs? Reflections for a culture of impact assessment and vigilance». *MCCSIS. IADIS International Conference ICT, Society and Human Beings 2018*, 2018, 281-85.

Hawath, Mariam. «Regulating Automated Decision-Making: An Analysis of Control over Processing and Additional Safeguards in Article 22 of the GDPR.» *European Data Protection Law Review* 7, n.º 2 (2021): 161-73.

Hert, Paul de, y Guillermo Lazcoz. «When GDPR-principles blind each other. Accountability, not transparency, at the heart of algorithmic governance». *European Data Protection Law Review* 8, n.º 1 (2022): 1-10.

Hond, Anne A H de, Artuur M Leeuwenberg, Lotty Hooft, Ilse M J Kant, Steven W J Nijman, Hendrikus J A van Os, Jiska J Aardoom, et al. «Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review». *npj Digital Medicine* 5, n.º 1 (2022): 2. https://doi.org/10.1038/s41746-021-00549-7.

Hoofnagle, Chris Jay, Bart van der Sloot, y Frederik Zuiderveen Borgesius. «The European Union general data protection regulation: what it is and what it means». *Information & Communications Technology Law* 28, n.º 1 (2 de enero de 2019): 65-98. https://doi.org/10.1080/13600834.2019.1573501.

Huq, Aziz Z. «A Right to a Human Decision». *Virginia Law Review* 106, n.º 3 (2020): 611-88.

ICO. «Guide to the UK General Data Protection Regulation (UK GDPR)», 2020.

Jamieson, Trevor, y Avi Goldfarb. «Clinical considerations when applying machine learning to decision-support tasks versus automation». *BMJ Quality & Safety* 28, n.º 10 (1 de octubre de 2019): 778 LP - 781. https://doi.org/10.1136/bmjqs-2019-009514.

Janssen, Heleen L. «An approach for a fundamental rights impact assessment to automated decision-making». *International Data Privacy Law* 10, n.º 1 (1 de febrero de 2020): 76-106. https://doi.org/10.1093/idpl/ipz028.

Jones, Meg Leta. «The right to a human in the loop: Political constructions of computer

automation and personhood». *Social Studies of Science* 47, n.º 2 (2017): 216-39. https://doi.org/10.1177/0306312717699716.

Kaminski, Margot E, y Gianclaudio Malgieri. «Algorithmic impact assessments under the GDPR: producing multi-layered explanations». *International Data Privacy Law* 11, n.º 2 (1 de abril de 2021): 125-44. https://doi.org/10.1093/idpl/ipaa020.

———. «Multi-Layered Explanations from Algorithmic Impact Assessments in the GDPR». En *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 68–79. FAT* '20. New York, NY, USA: Association for Computing Machinery, 2020. https://doi.org/10.1145/3351095.3372875.

Kim, Dong Wook, Hye Young Jang, Kyung Won Kim, Youngbin Shin, y Seong Ho Park. «Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: Results from recently published papers». *Korean Journal of Radiology* 20, n.º 3 (2019): 405-10. https://doi.org/10.3348/kjr.2019.0025.

Kiseleva, Anastasiya. «AI as a Medical Device: Is it Enough to Ensure Performance Transparency and Accountability?» *European Pharmaceutical Law Review* 4, n.º 1 (2020).

Kiseleva, Anastasiya, y Paul de Hert. «Creating a European Health Data Space: Obstacles in Four Key Legal Areas». *European Pharmaceutical Law Review* 5, n.º 1 (2021): 21-36.

Kolfschooten, Hannah Van. «EU regulation of artificial intelligence: Challenges for patients' rights». *Common Market Law Review* 59, n.º 1 (2022): 81-112.

Lehr, David, y Paul Ohm. «Playing with the data: what legal scholars should learn about machine learning». *UC Davis Law Review* 51 (2017): 653-717.

Lekadir, Karim, Gianluca Quaglio, Anna Tselioudis Garmendia, y Catherine Gallin. «Artificial intelligence in healthcare: Applications, risks, and ethical and societal impacts». Bruselas, 2022. https://doi.org/10.2861/568473.

Liu, Xiaoxuan, Livia Faes, Aditya U Kale, Siegfried K Wagner, Dun Jack Fu, Alice Bruynseels, Thushika Mahendiran, et al. «A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis». *The Lancet Digital Health* 1, n.º 6 (25 de septiembre de 2019): 271-97. https://doi.org/10.1016/S2589-7500(19)30123-2.

Liu, Xiaoxuan, Ben Glocker, Melissa M McCradden, Marzyeh Ghassemi, Alastair K Denniston, y Lauren Oakden-Rayner. «The medical algorithmic audit». *The Lancet Digital Health*, 2022. https://doi.org/https://doi.org/10.1016/S2589-7500(22)00003-6.

Malgieri, Gianclaudio. «Automated decision-making in the EU Member States: The right to explanation and other "suitable safeguards" in the national legislations». *Computer Law & Security Review*, 2019. https://doi.org/https://doi.org/10.1016/j.clsr.2019.05.002.

Malgieri, Gianclaudio, y Giovanni Comandé. «Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation». *International Data Privacy Law* 7, n.º 4 (1 de noviembre de 2017): 243-65. https://doi.org/10.1093/idpl/ipx019.

Malgieri, Gianclaudio, y Jędrzej Niklas. «Vulnerable data subjects». *Computer Law & Security Review* 37 (2020): 105415. https://doi.org/https://doi.org/10.1016/j.clsr.2020.105415.

Mantelero, Alessandro. «AI and Big Data: A blueprint for a human rights, social and ethical impact assessment». *Computer Law & Security Review* 34, n.º 4 (2018): 754-72. https://doi.org/https://doi.org/10.1016/j.clsr.2018.05.017.

Martinez-Millana, Antonio, Aida Saez-Saez, Roberto Tornero-Costa, Natasha Azzopardi-Muscat, Vicente Traver, y David Novillo-Ortiz. «Artificial intelligence and its impact on the domains of universal health coverage, health emergencies and health promotion: An overview of systematic reviews». *International Journal of Medical Informatics* 166 (2022): 104855. https://doi.org/https://doi.org/10.1016/j.ijmedinf.2022.104855.

Martínez, Ricard. «Cuestiones de ética jurídica al abordar proyectos de Big Data. El contexto del Reglamento general de protección de datos». *Dilemata* 24 (2017): 151-64.

Mayer-Schönberger, Viktor. «Paradigm shift». *Computer Law and Security Review* 40 (2021). https://doi.org/10.1016/j.clsr.2020.105515.

Mendoza, Isak, y Lee A Bygrave. «The Right Not to be Subject to Automated Decisions Based on Profiling BT -». En *EU Internet Law: Regulation and Enforcement*, editado por Tatiana-Eleni Synodinou, Philippe Jougleux, Christiana Markou, y Thalia Prastitou, 77-98. Cham: Springer International Publishing, 2017. https://doi.org/10.1007/978-3-319-64955-9_4.

Meszaros, Janos, Marcelo Corrales Compagnucci, y Timo Minssen. «The Interaction of the Medical Device Regulation and the GDPR: Do European Rules on Privacy and Scientific Research Impair the Safety and Performance of AI Medical Devices?» En *The Future of Medical Device Regulation: Innovation and Protection*, editado por Carmel Shachar, Christopher Robertson, I Glenn Cohen, Timo Minssen, y W Nicholson Price II, 77-90. Cambridge: Cambridge University Press, 2022. https://doi.org/DOI: 10.1017/9781108975452.007.

Mittelstadt, Brent. «"The doctor will not see you now ": The algorithmic displacement of virtuous medicine». En *3TH1CS - The reinvention of ethics in the digital age*, editado por P. Otto y E. Gräf, 1-8. Berlin: iRights, 2017.

Nagendran, Myura, Yang Chen, Christopher A Lovejoy, Anthony C Gordon, Matthieu Komorowski, Hugh Harvey, Eric J Topol, John P A Ioannidis, Gary S Collins, y Mahiben Maruthappu. «Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies». *BMJ* 368 (25 de marzo de 2020): m689. https://doi.org/10.1136/bmj.m689.

Noto La Diega, Guido. «Against the Dehumanisation of Decision-Making – Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information». *JIPITEC* 9, n.º 1 (2018).

Núñez Reiz, A, y M Sánchez García. «En respuesta a "Big Data Analysis y Machine Learning en medicina intensiva: identificando nuevos retos ético-jurídicos"». *Medicina Intensiva* 44, n.º 5 (2020): 320. https://doi.org/10.1016/j.medin.2020.01.004.

Palma Ortigosa, Adrian. «Automated Decision-Making in the GDPR. Algorithms in the Scope of the Data Protection». *Revista General De Derecho Administrativo*, n.º 50 (2019).

———. «Régimen jurídico de la toma de decisiones automatizadas y el uso de sistemas de inteligencia artificial en el marco del derecho a la protección de datos personales». Universidad de Valencia, 2021.

Parasuraman, Raja, y Dietrich H Manzey. «Complacency and Bias in Human Use of Automation: An Attentional Integration». *Human Factors* 52, n.º 3 (1 de junio de 2010): 381-410. https://doi.org/10.1177/0018720810376055.

Parziale, Andrea. « AI-powered Medical Devices and the Development Risk Defense Under the Revised Product Liability Directive». En F. Casarosa, F. Gennari, & A. Rossi (Eds.), *Enabling and Safeguarding Personalized Medicine*. Springer Nature Switzerland (2025): 279-290.

Pérez-Santonja, T, L Gómez-Paredes, S Álvarez-Montero, L Cabello-Ballesteros, y M.T Mombiela-Muruzabal. «Historia clínica electrónica : evolución de la relación médico-paciente en la consulta de Atención Primaria». *Semergen* 43, n.º 3 (2017): 175-81. https://doi.org/http://dx.doi.org/10.1016/j.semerg.2016.03.022.

Perin, Andrea. «Estandarización y automatización en medicina: El deber de cuidado del profesional entre la legítima confianza y la debida prudencia». *Revista Chilena de Derecho y Tecnología* 8, n.º 1 (2019): 3-28. https://doi.org/10.5354/0719-2584.2019.52560.

Perkonigg, Matthias, Johannes Hofmanninger, Christian J Herold, James A Brink, Oleg Pianykh, Helmut Prosch, y Georg Langs. «Dynamic memory to alleviate catastrophic

forgetting in continual learning with medical imaging». *Nature Communications* 12, n.º 1 (2021): 5678. https://doi.org/10.1038/s41467-021-25858-z.

Plana, Deborah, Dennis L Shung, Alyssa A Grimshaw, Anurag Saraf, Joseph J Y Sung, y Benjamin H Kann. «Randomized Clinical Trials of Machine Learning Interventions in Health Care: A Systematic Review». *JAMA Network Open* 5, n.º 9 (29 de septiembre de 2022). https://doi.org/10.1001/jamanetworkopen.2022.33946.

Prainsack, Barbara. «The political economy of digital data: introduction to the special issue». *Policy Studies* 41, n.º 5 (2020). https://doi.org/10.1080/01442872.2020.1723519.

Rajpurkar, Pranav, Emma Chen, Oishi Banerjee, y Eric J Topol. «AI in health and medicine». *Nature Medicine* 28, n.º 1 (2022): 31-38. https://doi.org/10.1038/s41591-021-01614-0.

Roig, Antoni. «Safeguards for the right not to be subject to a decision based solely on automated processing (Article 22 GDPR)». *European Journal of Law and Technology* 8, n.º 3 (2017).

Romeo Casabona, Carlos María. «Revisión de las categorías jurídicas de la normativa europea ante la tecnología del big data aplicada a la salud». *Revista de Derecho y Genoma Humano : Genética, Biotecnología y Medicina Avanzada* Extra, n.º 1 (2019): 85-127.

Romeo Casabona, Carlos María, y Gonzalo Arruego. «Lección 14. Toma de decisiones al final de la vida». En *Manual de Bioderecho: (adaptado a la docencia en ciencias, ciencias de la salud y ciencias sociales y jurídicas)*, 361-86, 2022.

Romeo Casabona, Carlos María, Encarnación Guillén, José Manuel Jeréz, Iñigo de Miguel Beriain, y Pilar Nicolás Jiménez. «Informe Anticipando, Inteligencia Artificial en salud: Retos éticos y legales», 2020.

Sartor, Giovanni, y Francesca Lagioia. «The impact of the General Data Protection Regulation (GDPR) on artificial intelligence (PE 641.530)», 2020.

Saura Llamas, J, y P Saturno Hernández. «Protocolos clínicos: ¿cómo se construyen? Propuesta de un modelo para su diseño y elaboración». *Atención Primaria* 18, n.º 2 (1996): 94-96.

Schoenbaum, S C, y L K Gottlieb. «Algorithm based improvement of clinical quality.» *British Medical Journal* 301, n.º 6765 (15 de diciembre de 1990): 1374 LP - 1376. https://doi.org/10.1136/bmj.301.6765.1374.

Schwemer, Sebastian Felix, Letizia Tomada, y Tommaso Pasini. «Legal AI Systems in the EU's proposed Artificial Intelligence Act». En *Proceedings of the Second International Workshop on AI and Intelligent Assistance for Legal Professionals in the Digital Workplace*, 1-8, 2021.

Selbst, Andrew D. «Disparate Impact in Big Data Policing». *Georgia Law Review* 52 (2017): 109-95.

Sittig, Dean F, y Hardeep Singh. «A New Sociotechnical Model for Studying Health Information Technology in Complex Adaptive Healthcare Systems». *Quality & Safety in Health Care* 19, n.º Suppl 3 (octubre de 2010): i68-74. https://doi.org/10.1136/qshc.2010.042085.

Sniderman, Allan D, Ralph B D'Agostino Sr, y Michael J Pencina. «The Role of Physicians in the Era of Predictive Analytics». *JAMA* 314, n.º 1 (7 de julio de 2015): 25-26. https://doi.org/10.1001/jama.2015.6177.

Soriano Arnanz, Alba. «La propuesta de Reglamento de inteligencia artificial de la UE y los sistemas de alto riesgo». *Revista General de Derecho de los Sectores Regulados* 8 (2021): 1-24.

Stöger, Karl, David Schneeberger, Peter Kieseberg, y Andreas Holzinger. «Legal aspects of data cleansing in medical AI». *Computer Law & Security Review* 42 (2021): 105587. https://doi.org/https://doi.org/10.1016/j.clsr.2021.105587.

Supervisor Europeo de Protección de Datos (SEPD). «Dictamen 3/2015 -La gran oportunidad

de Europa. Recomendaciones del SEPD sobre las opciones de la UE en cuanto a la reforma de la protección de datos-», 2015.

Tai-Seale, Ming, Cliff W Olson, Jinnan Li, Albert S Chan, Criss Morikawa, Meg Durbin, Wei Wang, y Harold S Luft. «Electronic Health Record Logs Indicate That Physicians Split Time Evenly Between Seeing Patients And Desktop Medicine». *Health Affairs* 36, n.º 4 (2017): 655-62. https://doi.org/10.1377/hlthaff.2016.0811.

Tamò-Larrieux, Aurelia. «Decision-making by machines: Is the 'Law of Everything' enough?» *Computer Law & Security Review* 41 (2021): 105541. https://doi.org/https://doi.org/10.1016/j.clsr.2021.105541.

The Lancet Digital Health. «Holding artificial intelligence to account». *The Lancet Digital Health*, 2022. https://doi.org/https://doi.org/10.1016/S2589-7500(22)00068-1.

Topol, Eric J. *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. Basic Books, Inc., 2019.

———. «High-performance medicine: the convergence of human and artificial intelligence». *Nature Medicine* 25, n.º 1 (2019): 44-56. https://doi.org/10.1038/s41591-018-0300-7.

Turégano Mansilla, Isabel. «Los valores detrás de la privacidad». *Doxa. Cuadernos de Filosofía del Derecho* 43 (2020): 255-83. https://doi.org/10.14198/DOXA2020.43.10.

Vázquez López, J Enrique. «La "Lex Artis ad hoc" como criterio valorativo para calibrar la diligencia exigible en todo acto o tratamiento médico: A propósito de un caso basado en la elección de la técnica empleada en el parto (parto vaginal vs. cesárea) ». *Cuadernos de Medicina Forense* . scieloes , 2010.

Veale, Michael, y Lilian Edwards. «Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling». *Computer Law & Security Review* 34, n.º 2 (1 de abril de 2018): 398-404. https://doi.org/10.1016/J.CLSR.2017.12.002.

Veale, Michael, y Frederik Zuiderveen Borgesius. «Demystifying the Draft EU Artificial Intelligence Act — Analysing the good, the bad, and the unclear elements of the proposed approach». *Computer Law Review International* 22, n.º 4 (2021): 97-112. https://doi.org/doi:10.9785/cri-2021-220402.

Verdicchio, Mario, y Andrea Perin. «When Doctors and AI Interact: on Human Responsibility for Artificial Risks». *Philosophy & Technology* 35, n.º 1 (2022): 11. https://doi.org/10.1007/s13347-022-00506-6.

Viljanen, Mika. «Is EU regulating the right AI?» *Turun yliopiston blogi*, marzo de 2022.

Wachter, Sandra, Brent Mittelstadt, y Luciano Floridi. «Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation». *International Data Privacy Law* 7, n.º 2 (1 de mayo de 2017): 76-99. https://doi.org/10.1093/idpl/ipx005.

Wagner, Ben. «Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision-Making Systems». *Policy & Internet* 11, n.º 1 (2019): 104-22. https://doi.org/10.1002/poi3.198.