

## Doctrina / Articles

### **Inteligencia artificial aplicada a la salud: viejas y nuevas cuestiones para el Derecho penal**

Artificial intelligence applied to health: old and new issues for criminal law

M<sup>a</sup> del Carmen Gómez Rivero

Catedrática de Derecho penal de la Universidad de Sevilla

---

**Sumario / Summary:** I.- Consideraciones previas; II.- La incidencia de la IA en los derechos del paciente: 1- El derecho a la información y autonomía del paciente; 2.- La eventual afectación de otros derechos del paciente; III.- Los retos del diseño de las bases de la responsabilidad penal por un resultado lesivo en el ejercicio de la medicina asistida por IA: 1.- La responsabilidad por los daños derivados de una actuación asistida por IA en la fase de diagnóstico; 2.- Determinación de la responsabilidad por el uso de la IA en fase de toma de decisiones sobre el tratamiento más adecuado para un determinado diagnóstico; 3.- Bases para la apreciación de una eventual responsabilidad penal en la fase de ejecución del tratamiento: el específico problema de los robots sanitarios.

**Resumen:** El trabajo trata una serie de cuestiones que plantea la irrupción de los sistemas de inteligencia artificial en el ámbito de la medicina. Se aborda, en concreto, su eventual incidencia en alguno de los derechos reconocidos a los pacientes, como la autonomía y privacidad, así como la cuestiones relativas a la eventual responsabilidad penal asociada al uso de tales herramientas, tanto colaborativas como autónomas.

**Palabras clave:** Inteligencia artificial, diagnóstico médico, tratamiento médico, privacidad, big data, discriminación, resultados lesivos, robots sanitarios, responsabilidad penal.

**Abstract:** This paper addresses a series of issues raised by the emergence of artificial intelligence systems in the field of medicine. Specifically, it addresses their potential impact on certain rights recognised to patients, such as autonomy and privacy, as well as issues relating to the potential criminal liability associated with the use of such tools, both collaborative and autonomous.

**Keywords:** Artificial intelligence, medical diagnosis, medical treatment, privacy, big data, discrimination, damage, healthcare robots, criminal liability.

## I.- Consideraciones previas

Hace tiempo que el término Inteligencia Artificial (IA) ha pasado a formar parte del lenguaje cotidiano, y que el contenido que intuitivamente evoca se ha incorporado al acervo común de la sociedad. Aun sin necesidad de manejar una noción exacta de su significado, la realidad que identifica se percibe como sello de identidad de nuestro tiempo y, a la vez, como un avance imparable llamado a transformar profundamente nuestro modo de vida y, aún más, el de las generaciones venideras en los más variados ámbitos. Cotidianas son algunas aplicaciones ordenadas a facilitar la vida ordinaria con diversos fines<sup>1</sup>, noticiables otras, por su originalidad<sup>2</sup>, y prometedora es en todo caso la gama de aplicaciones que en un futuro más o menos próximo es de esperar que nos acompañen en los más variados ámbitos<sup>3</sup>. Pero tal vez las mayores expectativas depositadas en el diseño de este tipo de herramientas se concentren en los avances que puede ofrecer en sectores profesionales. Se cuentan entre ellos la asistencia en la labor ordinaria de los Cuerpos y Fuerzas de Seguridad del Estado, de la Administración de Justicia<sup>4</sup>, en la enseñanza o, en lo que interesa a estas líneas, en el ámbito de la sanidad<sup>5</sup>. Si bien con altibajos en el tiempo<sup>6</sup>, las prometedoras aplicaciones de la IA han despertado encendidas expectativas sobre su rendimiento, siendo así que, si en algo existe consenso, es en reconocer que, en cualquiera de los múltiples ámbitos aplicativos, la irrupción de la llamada *cuarta revolución industrial* no tiene vuelta atrás.

Antes de entrar a tratar los retos y dificultades de la IA en el específico ámbito de la sanidad, conviene realizar algunas acotaciones conceptuales que delimiten el alcance del término como presupuesto de su uso en el orden jurídico. Sin desconocer las definiciones formuladas en la abundante literatura al respecto<sup>7</sup>, una noción precisa la ofrece el legislador europeo en el Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024<sup>8</sup>

---

<sup>1</sup> Ya sea para la búsqueda de información, como el chap GPT, o para ejecutar acciones de variado tipo -como los asistentes virtuales (entre otros, Alexa o Siri), los asistentes de conducción inclusivos del GPS, o las recomendaciones personalizadas de compras en internet-.

<sup>2</sup> Es el caso de la noticia que en septiembre de 2025 daba cuenta del nombramiento en Albania de una ministra asistente digital, de nombre Diella, encargada de todas las cuestiones relativas a la licitación pública, como medio para evitar la corrupción, información disponible en: <https://es.euronews.com/next/2025/09/13/albania-nombra-a-la-primer-ministra-de-inteligencia-artificial-del-mundo> (última consulta: 20 de septiembre de 2025).

<sup>3</sup> A modo de ejemplo, como la puesta en el mercado de vehículos de conducción autónoma o el diseño de drones para la entrega rápida de mercancías.

<sup>4</sup> Se trata de un ámbito especialmente atractivo a la vista de las posibilidades que ofrece tanto en sede de investigación policial como en lo que se refiere propiamente a la labor de la Administración de Justicia. Por todos, MIRÓ LLINARES, Fernando, "Inteligencia artificial y justicia penal: Más allá de los resultados lesivos causados por robots", *Revista de Derecho Penal y Criminología*, núm. 20, 2018, pp. 87 ss.

<sup>5</sup> Así lo señalaba ya el *Informe sobre IA y salud de la Oficina de Ciencia y Tecnología del Congreso de los Diputados*, Informe C: *Inteligencia artificial y salud*. 2022: "La salud es uno de los ámbitos donde se espera una mayor contribución de la inteligencia artificial (IA), con implicaciones en el coste, calidad y eficiencia de la atención médica, incluida la capacidad de preparación y respuesta ante emergencias sanitarias". Disponible en: [https://www.congreso.es/backoffice\\_doc/prensa/notas\\_prensa/94234\\_1668420226544.pdf](https://www.congreso.es/backoffice_doc/prensa/notas_prensa/94234_1668420226544.pdf) (última consulta: 30 de mayo de 2025).

<sup>6</sup> Se trata de los llamados periodos de "invierno" de la IA, provocados por cierto desaliento en las posibilidades de estas técnicas, debido en buena medida a las sobre expectativas depositadas en ellas. De forma detallada AMUNÁTEGUI PERELLÓ, Carlos, *Arcana Technicae. El Derecho y la inteligencia artificial*, Arcana Technicae. *El Derecho y la inteligencia artificial*. Valencia, Tirant lo Blanch, 2020, pp. 18 s.

<sup>7</sup> A título de ejemplo, GONZÁLEZ CABANES, Francisco / DÍAZ DÍAZ, Norberto, quienes la definen como "la disciplina que intenta replicar y desarrollarla inteligencia natural, de humanos y seres vivos, a través de computadoras o máquinas", en "¿Qué es la inteligencia artificial?", Gamero Casado, Francisco Luis, (ed.), *Inteligencia artificial y sector público: Retos, límites y medios*. Valencia, 2023, p. 40. Entre otras definiciones similares se cuenta la propuesta por AMUNÁTEGUI PERELLÓ, Carlos, *Arcana Technicae. El Derecho y la inteligencia artificial*, op. cit., p. 51, para quien se trata del "conjunto de técnicas que permiten resolver de manera automatizada problemas que para los seres humanos requieren intelecto", si bien subrayando las dificultades de ofrecer un concepto más allá de esa definición y del detalle que permite la atención a la perspectiva histórica, pp. 13 ss. Otras definiciones especifican las distintas secuencias que comprende la IA. Señala DE LA CUESTA AGUADO, Paz, que "El concepto de inteligencia artificial hace referencia tanto a una rama de la ciencia de la computación que trata de diseñar agentes artificiales inteligentes como a la inteligencia de tales agentes", "Inteligencia artificial y responsabilidad penal", *Revista Penal México*, núm. 16-17, 2019, p. 52. Véase también sobre el origen del concepto y su significado actual VALLS PRIETO, Javier, "Inteligencia artificial, derechos humanos y bienes jurídicos", Aranzadi, Pamplona, 2021, pp. 17 ss.

<sup>8</sup> Por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n° 300/2008, (UE) n° 167/2013, (UE) n° 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828.

(Reglamento IA). Su art. 3.1 la define como “un sistema basado en una máquina que está diseñado para funcionar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación tras el despliegue, y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales”<sup>9</sup>.

Las Directrices interpretativas formuladas por la UE con fecha de 29 de julio de 2025 para facilitar la aplicación del Reglamento, aclaran los siete elementos básicos de la definición<sup>10</sup>. En primer lugar, con la indicación de que se trata de un sistema basado en una máquina pone de relieve que debe estar dirigido de forma computacional y basarse en operaciones de máquina, integrada por componentes de hardware y software. En segundo lugar, la referencia a que el sistema esté «diseñado para funcionar con distintos niveles de autonomía», supone a su vez el reconocimiento de su capacidad para generar resultados de salida, con mayor o menor grado de independencia respecto a la actuación humana. En tercer lugar, la definición señala como eventual rasgo, si bien no indispensable, la capacidad de adaptación del sistema, entendiendo por tal la programación de la herramienta para el autoaprendizaje, lo que implica su capacidad para cambiar y adaptar su comportamiento. El cuarto elemento hace referencia a los objetivos del sistema, explícitos o implícitos, aspecto que remite a las tareas a realizar. El quinto elemento, considerado esencial, reclama la capacidad de la herramienta para inferir resultados de salida a partir de los datos de entrada. La sexta característica refiere dicha capacidad a la formulación de predicciones, contenidos, recomendaciones o decisiones, resultados en todo caso obtenidos a partir de la información de entrada y basados en distintos sistemas, entre los que se cuentan los de aprendizaje automático<sup>11</sup>. Por último, el séptimo rasgo de la definición apunta a la potencial influencia del sistema en entornos físicos (por ejemplo, un robot) o virtuales.

Cumpliendo tales características para ser considerados como sistemas inteligentes, su aplicación al ámbito de la medicina ha coincidido prácticamente con la irrupción de estas herramientas<sup>12</sup>, generalizándose su uso de modo paulatino hasta el punto de poder decirse que en la situación actual suponen la puesta en manos de los sanitarios de un revolucionario instrumento para la realización de buena parte de sus funciones, y prometen aún más en el futuro una innovación integral de la práctica médica. Su rumbo se desvía de los derroteros propios de la medicina clásica, basada en exclusiva en el método de la evidencia científica como técnica para obtener diagnósticos, diseñar tratamientos y en general pautas de actuación. Como subraya la ya referida definición europea de la IA, en lugar de ello, la obtención de los resultados parte de la incorporación y posterior procesamiento de ingentes cantidades de datos, cuya interrelación se sitúa en condiciones de ofrecer un panorama completo tanto sobre las condiciones médicas en general como sobre su aplicación a un paciente concreto<sup>13</sup>.

Los prometedores rendimientos de los sistemas inteligentes en medicina aspiran a proyectarse en un doble plano, que no necesariamente tienen que comprenderse de un modo secuencial o lineal. Por una parte, alcanzar resultados equiparables en precisión a la actuación humana. Por otra, superar las posibilidades de ésta, lo que habría de traducirse en un diagnóstico, así como en unas decisiones que ofrecieran mayor fiabilidad de la hasta ahora conseguida por la vieja metodología basada en la evidencia científica. En este sentido es abundante la literatura médica especializada que da cuenta del importante cambio de paradigma que en el ámbito de la sanidad supone la llegada de aquella tecnología, que parece traer consigo una renovación en

---

<sup>9</sup> La definición del Reglamento entró en vigor el 2 de febrero de 2025.

<sup>10</sup> *Communication from the commission Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)*, disponible en: <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application> (última consulta: 25 de mayo de 2025).

<sup>11</sup> Las citadas Directrices interpretativas ofrecen una clasificación de los sistemas. Junto a los de aprendizaje automático se contemplan los basados en la lógica y los inferidos a partir de conocimientos codificados. A su vez, dentro de los sistemas de aprendizaje automático cabe distinguir el supervisado, el no supervisado, el autosupervisado y el aprendizaje por refuerzo, pp. 7 ss.

<sup>12</sup> Suele identificarse su primera aplicación en el sistema desarrollado a principio de los años 70, *Mycin*. Fue programado para diagnosticar enfermedades infecciosas detectables en sangre, proponiendo adicionalmente un tratamiento personalizado para el paciente.

<sup>13</sup> Véase en este sentido el Considerando 157 del Reglamento General de Protección de Datos.

prácticamente todos los aspectos relacionados con la salud<sup>14</sup>. Sin desconocer su gama de aplicaciones<sup>15</sup>, las consideraciones que siguen se centran en las relacionadas con la práctica clínica, entendiendo por tal la prestación asistencial en cada una de las fases en que se desarrolla la relación médico-paciente: la de prevención, diagnóstico, tratamiento y seguimiento de la enfermedad<sup>16</sup>.

En relación con la prevención, singularmente atractivas se presentan las posibilidades de la IA de cara a la predicción de patologías sobre la base de una serie de algoritmos que no sólo toman en cuenta la predisposición genética del individuo, sino patrones de repetición, e incluso la interpretación de los registros de la actividad del sujeto<sup>17</sup>, patrones y predicciones que impulsan el avance de la llamada medicina personalizada<sup>18</sup>. Por su parte, en la fase de diagnóstico la IA se revela como una herramienta extraordinariamente útil a la vista de su capacidad para analizar una ingente cantidad de información de los historiales clínicos ordenados a la detección de enfermedades, importancia paralela a la que también en esta fase cobran los algoritmos diseñados para interpretar imágenes médicas (radiografías, electrocardiogramas,

<sup>14</sup> Entre la abundante literatura al respecto puede verse, entre otros, DEL RÍO SOLÁ, María Lourdes, LÓPEZ SANTOS, José María /VAQUERO PUERTA, Carlos, "La inteligencia artificial en el ámbito médico", *Revista Española de Investigaciones Quirúrgicas*, vol. XXI, núm. 3, 2018, pp. 113 ss; GÓMEZ GONZÁLEZ, Emilio/ GÓMEZ Emilia/ MARQUEZ-RIVAS, Javier, et al. *Inteligencia Artificial en Medicina y Salud: revisión y clasificación de las aplicaciones actuales y del futuro cercano y su impacto ético y social (traducción del original en inglés)*, arXiv: 2001.09778, 2020, disponible en: [https://5a8fe19c-a1f8-41fe-938d-8c3f23f975c9.filesusr.com/ugd/d42f6e\\_2fc902643d594faa8e80d82d7f43d2b6.pdf](https://5a8fe19c-a1f8-41fe-938d-8c3f23f975c9.filesusr.com/ugd/d42f6e_2fc902643d594faa8e80d82d7f43d2b6.pdf) pp. 1 ss. (última consulta: 12 de mayo de 2025), donde ofrecen una exhaustiva clasificación de la IA y sus aplicaciones en el ámbito de la salud, pp. 3 ss., 6 ss; HE Jianxing./ BAXTER Sally/ XU Jie/ XU Jiming/ ZHOU Xingtao/ ZHANG Kang "The practical implementation of artificial intelligence technologies in medicine", *Naturemedicine*, núm. 24, 2019, disponible en <https://pubmed.ncbi.nlm.nih.gov/30617336/> (última consulta 15 de julio de 2025). Interesante es la consulta de la Declaración del Comité Ejecutivo de la Asociación Médica alemana, adoptada en su reunión de 14 de enero de 2025, que ofrece un recorrido por las distintas aplicaciones de la IA en medicina, "Stellungnahme kunstliche Intelligenz in der Medizin", en *Deutsches Ärzteblatt* 4/2025, pp. 33 ss. Disponible en: <https://www.aerzteblatt.de/archiv/stellungnahme-kuenstliche-intelligenz-in-der-medizin-1581a074-7da3-4724-875e-09f2be35a9a3> (última consulta: 20 de mayo de 2025).

<sup>15</sup> Entre ellas, el ámbito de la docencia, la investigación, gestión de recursos materiales y humanos o la de la calidad asistencial en los centros hospitalarios y, en general, en unidades asistenciales. En cualquiera de ellos ofrece una optimización de los recursos sanitarios, humanos y técnicos, así como economía de tiempo. Sirva como muestra la aplicación que permite la transcripción de los datos relevantes de la consulta, permitiendo prescindir del tiempo dedicado al teclado de los datos clínicos del paciente. Es, por ejemplo, el caso de la herramienta *Scribe*, implementada ya en algunos centros hospitalarios, que permite transcribir y procesar la conversación entre médico y paciente, de modo que la consulta médica se centra en la atención a éste, sin dedicar tiempo a labores administrativas. La aplicación genera informes precisos a partir de la identificación de los datos relevantes y sugiere tratamientos y pruebas de diagnóstico. Información disponible en: <https://isanidad.com/311979/scribe-un-proyecto-con-ia-para-mirar-a-los-ojos-del-paciente-en-la-consulta/> (última consulta: 8 de septiembre de 2025).

Sobre las distintas aplicaciones de la IA al ámbito sanitario puede verse, por ejemplo, ÁVILA-TOMÁS, José Francisco/ MAYER-PUJADAS, Miguel Ángel/ QUESADA VARELA Víctor Julio, "La inteligencia artificial y sus aplicaciones en medicina II: Importancia actual y aplicaciones prácticas", *op. cit.*, pp. 83 ss.

<sup>16</sup> El potencial de la IA en materia sanitaria en absoluto se ciñe a la estricta relación asistencial médico paciente, esto es, a la prevención y tratamiento de enfermedades de un sujeto en cuestión que demanda una prestación asistencial. Alcanza también al ejercicio en general de todo lo relacionado con la sanidad. Bastaría con referir su interés desde el punto de vista del desarrollo de nuevos medicamentos o del diseño de políticas públicas de salud, entre otros aspectos. Véase al respecto el Informe del Parlamento Europeo "Artificial intelligence in healthcare Applications, risks, and ethical and societal impacts", 2022. El documento identifica cuatro grandes campos en los que es esperable que la IA biomédica realice las aportaciones más significativas: el primero es la práctica clínica, al que se ciñen las consideraciones de esta contribución; el segundo la investigación biomédica, comprensiva del descubrimiento de nuevos fármacos, ensayos clínicos y diseño de medicina personalizada, entendiendo por tal la que pronostica el efecto del medicamento en un individuo a la vista de sus características, inclusive las genéticas; el tercero, el de la salud pública, que a partir de grandes magnitudes de datos ofrece resultados sobre la predicción y el tratamiento de pandemias o evolución del comportamiento de determinadas enfermedades; y el cuarto la administración sanitaria o gestión del sistema sanitario. Disponible en: [https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729512/EPRS\\_STU\(2022\)729512\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729512/EPRS_STU(2022)729512_EN.pdf) (última consulta: 2 de mayo de 2025).

<sup>17</sup> De ello dan cuenta ÁVILA-TOMÁS, José Francisco/ MAYER-PUJADAS, Miguel Ángel/ QUESADA VARELA, Víctor Julio, destacando, entre otras, la posibilidad de generar algoritmos que predicen el riesgo de psicopatologías a partir del análisis del contenido de mensajes publicados en redes sociales, o de identificar patrones que pudieran sugerir una progresión del Parkinson a través de la interpretación de datos que registran nuestros movimientos (sensores en el teléfono móvil o reloj inteligente), "La inteligencia artificial y sus aplicaciones en medicina II: Importancia actual y aplicaciones prácticas", *Atención Primaria*, núm. 53, 2020, p. 83.

<sup>18</sup> A ello se refiere el Informe del Parlamento Europeo "Artificial intelligence in healthcare Applications, risks, and ethical and societal impacts", *op. cit.*, p. 12. Señala el documento cómo la medicina personalizada se basa en gran medida en la comprensión científica del modo en que las características únicas de un paciente individual, así como los perfiles moleculares y genéticos lo hacen vulnerable a una determinada enfermedad y sensible a un específico tratamiento terapéutico. Concluye del siguiente modo: "Las herramientas de IA pueden mejorar los avances en medicina personalizada evaluando el beneficio clínico de diferentes métodos de investigación y múltiples tipos de datos", p. 12.

encefalocardiogramas, etc)<sup>19</sup>. Es más, en el caso del diagnóstico, las ventajas del uso de estas herramientas no se agotan en su precisión, sino también en la rapidez de sus resultados, algo crucial en determinados escenarios que reclaman una actuación inmediata<sup>20</sup>. Atractivo es también su empleo para elegir el tratamiento más adecuado, fase en que los sistemas de IA pueden marcar pautas de actuación tomando en cuenta la predicción de la respuesta del paciente en función de sus características. No menos importante es el valor de aquellas herramientas en la fase de ejecución del tratamiento, a la vista del avance que ya supone y que se potenciará en el futuro de la mano de la intervención de los llamados robot cirujanos, cuyo desarrollo aspira a alcanzar plenos niveles de autonomía<sup>21</sup>.

Los beneficios de tan sugerente presente y futuro contrastan con los riesgos potencialmente asociados a su uso. Obvio es que esos temores no son predicables en el mismo grado de todas las aplicaciones de la IA en el ámbito sanitario, siendo su nivel más alto conforme mayor sea el protagonismo decisional del sistema, así como más directa y elevada sea también su incidencia en bienes o derechos de la persona. El primero de los aspectos remite directamente a la diferencia entre los conocidos como sistemas de IA fuertes o débiles, atenta a su grado de autonomía respecto a la actuación humana. Mientras los sistemas débiles centran su objetivo en la realización de competencias puntuales predefinidas a partir de una instrucción de entrada, los adjetivados como fuertes se ordenan a la toma de decisiones autónomas, recreando la capacidad humana de aprender, razonar y tomar decisiones<sup>22</sup>.

El segundo aspecto señalado lleva a extremar las precauciones y el consiguiente nivel de riesgo allí donde la actuación del sistema pueda incidir en aspectos o derechos esenciales de la persona. Es precisamente la atención a estos extremos lo que ha llevado al legislador europeo a identificar los denominados sistemas de alto riesgo dentro de la clasificación que ofrece<sup>23</sup>, sometidos a un régimen especial de gestión de riesgos<sup>24</sup>. Básicamente, distingue la Ley entre aquellos que suponen un riesgo relevante con arreglo a la legislación armonizada de la Unión en materia de seguridad de los productos; y los que pueden afectar significativamente a la salud, la seguridad o los derechos fundamentales de las personas en casos de los usos específicos enumerados en el Reglamento IA<sup>25</sup>. Serán una serie de directrices futuras emanadas de la

<sup>19</sup> Sobre sus posibilidades, así como sobre algunos ensayos realizados hasta la fecha, puede verse GRUBE, Karl, "Is the future of medical diagnosis in computer algorithms?", *The Lancet Digital Health*, vol. 1, mayo 2019, pp. 15 s.

<sup>20</sup> Paradigma al respecto es la situación que se plantea allí donde en el curso de una intervención quirúrgica urge el conocimiento del diagnóstico, para conocer la necesidad, en su caso, de extirpar un tumor o la zona potencialmente afectada. Un modelo de IA diseñado precisamente para proporcionar al cirujano información rápida es el conocido como *Sturgeon*. Utilizado en el caso de intervenciones de tumores cerebrales proporciona información al sanitario para decidir sobre la zona a extirpar. El sistema escanea segmentos del ADN del paciente, de modo que atendiendo a determinadas modificaciones ofrece información relevante sobre la agresividad del tumor y, con ello, sobre el modo de actuar en tal caso. Sus ventajas no se agotan en la precisión de la información, sino en la capacidad de suministrarla durante el tiempo de la intervención. Información disponible al respecto en: [https://www.clarin.com/new-york-times-international-weekly/nueva-herramienta-ia-agnostica-tumores-cerebrales-quiropano\\_0\\_6vgqyHc2XO.html?srsid=AfmBOoqKzQ0GNX6Nd5P6pchoDb2r](https://www.clarin.com/new-york-times-international-weekly/nueva-herramienta-ia-agnostica-tumores-cerebrales-quiropano_0_6vgqyHc2XO.html?srsid=AfmBOoqKzQ0GNX6Nd5P6pchoDb2r) (última consulta: 2 de agosto de 2025).

<sup>21</sup> Todas esas aplicaciones se han agrupado conforme a cuatro áreas básicas: las de relación, ordenadas a la relación del ser humano con el sistema, las de decisión, que ayudan a los profesionales a la toma de decisiones, las de descubrimiento, que se orientan a la interpretación de las enfermedades, y las de operaciones, que procuran mejorar la eficiencia del sistema. Véase al respecto DíEZ, Juan José, "Aplicaciones médico-sanitarias de la inteligencia artificial: una visión desde la práctica clínica", *Anales de la Real Academia de Doctores de España*, vol. 8, núm. 4, 2023, pp. 877 ss. Disponible en: [https://www.rade.es/imageslib/PUBLICACIONES/ARTICULOS/V8N4%20-%2012%20-%20CON%20-%20DIEZ\\_IA%20medicina.pdf](https://www.rade.es/imageslib/PUBLICACIONES/ARTICULOS/V8N4%20-%2012%20-%20CON%20-%20DIEZ_IA%20medicina.pdf) (última consulta: 12 de mayo de 2025).

<sup>22</sup> Puede verse al respecto el Dictamen del Comité Económico y Social Europeo de 31 de mayo y 1 de junio de 2017 (DOUE de 31 de agosto de 2017). Punto 2.2 "En términos generales, se puede distinguir entre IA débil (narrow AI) e IA fuerte (general AI). La IA débil es capaz de realizar tareas específicas. La IA fuerte es capaz de realizar las mismas tareas intelectuales que un ser humano".

<sup>23</sup> Distingue entre prácticas prohibidas -que incluyen los sistemas de manipulación subliminal (Título II, art. 5)-, los sistemas de alto riesgo -permitidos si superan un procedimiento de evaluación (Título II, Capítulo I, Anexos II y III)-, sistemas con riesgos relacionados con la transparencia -permitidos pero sujetos a procedimientos de evaluación y transparencia (art. 52)-, y de riesgo mínimo -no sujeto a restricciones (art. 69)-.

<sup>24</sup> De su definición se ocupa el ya citado Reglamento IA, que establece en su art. 6 los sistemas de IA considerados como de alto riesgo, siendo previamente definido el riesgo en su art. 3 como "la combinación de la probabilidad de que se produzca un perjuicio y la gravedad de dicho perjuicio".

<sup>25</sup> Considera la norma como de alto riesgo un sistema que reúne las condiciones que especifica; en concreto, incluye dos grupos de casos. El primero remite a los productos enumerados en su anexo, que incorpora un listado de actos legislativos de armonización de la Unión basados en el nuevo marco legislativo. Conforme al art. 6.1, la remisión tiene lugar en un doble sentido: bien que esté destinado a ser utilizado como un componente de seguridad de un producto de los allí enumerados, bien que el propio sistema sea uno de esos productos. En la relación normativa comprendida en



Comisión, las que habrán de proponer un listado exhaustivo de ejemplos y casos prácticos de uso de sistemas de IA de alto riesgo<sup>26</sup>.

Una de las principales consecuencias asociadas en el reglamento a la identificación del sistema como de alto riesgo es la exigencia de que no actúe de modo por completo independiente o autónomo, sino siempre bajo supervisión humana. Así lo deja sentado su art. 14, conforme al cual, tales sistemas “se diseñarán y desarrollarán de modo que puedan ser vigilados de manera efectiva por personas físicas durante el período que estén en uso, lo que incluye dotarlos de herramientas de interfaz humano-máquina adecuadas”. El apartado segundo del mismo artículo traza los objetivos de aquella supervisión, que serán, en concreto, los de “prevenir o reducir al mínimo los riesgos para la salud, la seguridad o los derechos fundamentales que pueden surgir cuando se utiliza un sistema de IA de alto riesgo conforme a su finalidad prevista o cuando se le da un uso indebido razonablemente previsible, en particular cuando dichos riesgos persistan a pesar de la aplicación de otros requisitos establecidos en la presente sección”.

Como habrá ocasión de ir comprobando en las líneas que siguen, este principio, reconocido por lo demás en otros documentos<sup>27</sup>, resulta esencial en orden a sentar los parámetros mínimos de actuación que reclama la permisibilidad del uso de tales herramientas en el ámbito sanitario. Interesa tan solo de momento dejar sentado que conforme a aquella premisa su implementación pasa, ante todo, por reclamar el componente humano en su uso, al menos en términos de supervisión y seguimiento de los resultados que arroje<sup>28</sup>, de modo que, como se ha expresado gráficamente en la doctrina, el ser humano siga siendo “dueño de sus propias decisiones”<sup>29</sup>.

---

dicho anexo se cuenta el Reglamento (UE) 2017/746 del Parlamento Europeo y del Consejo, de 5 de abril de 2017, sobre los productos sanitarios para diagnóstico in vitro y por el que se derogan la Directiva 98/79/CE y la Decisión 2010/227/UE de la Comisión. De acuerdo con el apartado tercero del mismo art. 6, las previsiones del apartado primero serán de aplicación cuando se cumpla cualquiera de las condiciones que enuncia el precepto, entre las que se incluyen que el sistema de IA esté destinado a realizar una tarea de procedimiento limitada; esté destinado a mejorar el resultado de una actividad humana previamente realizada, o esté destinado a detectar patrones de toma de decisiones o desviaciones con respecto a patrones de toma de decisiones anteriores y no esté destinado a sustituir la valoración humana previamente realizada sin una revisión humana adecuada, ni a influir en ella.

El segundo bloque de casos que conforme al Reglamento IA lleva a considerar el sistema como de alto riesgo, remite a los contemplados en el anexo III, cuando planteen un riesgo importante de causar un perjuicio a la salud, la seguridad o los derechos fundamentales de las personas físicas, también al no influir sustancialmente en el resultado de la toma de decisiones (art. 6.3). Se considerará en todo caso de alto riesgo el sistema cuando lleve a cabo la elaboración de perfiles de personas físicas (art. 6.3). Entre otros, se incluyen la relación del anexo III, los “Sistemas de IA destinados a ser utilizados por las autoridades públicas o en su nombre para evaluar la admisibilidad de las personas físicas para beneficiarse de servicios y prestaciones esenciales de asistencia pública, incluidos los servicios de asistencia sanitaria, así como para conceder, reducir o retirar dichos servicios y prestaciones o reclamar su devolución”, así como los “Sistemas de IA destinados a ser utilizados para la evaluación y la clasificación de las llamadas de emergencia realizadas por personas físicas o para el envío o el establecimiento de prioridades en el envío de servicios de primera intervención en situaciones de emergencia, por ejemplo, policía, bomberos y servicios de asistencia médica, y en sistemas de triaje de pacientes en el contexto de la asistencia sanitaria de urgencia”.

<sup>26</sup> El 5 de junio de 2025 la Comisión puso en marcha una consulta pública para recabar información sobre la aplicación de las normas del Reglamento relativas a los sistemas de IA de alto riesgo, consulta que abierta hasta el 18 de julio de este mismo año. Durante ese período se trataba de recopilar información de distintos agentes, entre ellos, los proveedores y desarrolladores de sistemas de IA de alto riesgo, pero también la ciudadanía en general. Información disponible en: <https://digital-strategy.ec.europa.eu/es/news/commission-launches-public-consultation-high-risk-ai-systems>, (última consulta: 12 de junio 2025).

<sup>27</sup> Es el caso de Carta de derechos digitales, un documento elaborado por el Ministerio de Asuntos Económicos y Transformación Digital de España en 2021 con el objetivo de garantizar los derechos en el ámbito digital. Disponible en: [https://portal.mineco.gob.es/es-es/ministerio/participacionpublica/audienciapublica/Paginas/SEDIA\\_Carta\\_Derechos\\_Digitales.aspx](https://portal.mineco.gob.es/es-es/ministerio/participacionpublica/audienciapublica/Paginas/SEDIA_Carta_Derechos_Digitales.aspx), (última consulta: 20 de junio de 2025). Ya en su capítulo XXII, punto 4 declara “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”. Véase también el Libro Blanco de la Comisión Europea, que contempla la necesidad de la supervisión humana en las aplicaciones de IA de alto riesgo (Human oversight) “El objetivo de una IA digna de confianza, ética y centrada en el ser humano sólo puede alcanzarse garantizando una participación adecuada de los seres humanos en relación con las aplicaciones de IA de alto riesgo”, p. 21. Disponible en: [https://www.industriaconectada40.gob.es/SiteCollectionDocuments/libro\\_blanco\\_I.A.\\_comision\\_%20europeacommission\\_feb2020\\_en.pdf](https://www.industriaconectada40.gob.es/SiteCollectionDocuments/libro_blanco_I.A._comision_%20europeacommission_feb2020_en.pdf), (última consulta: 20 de junio de 2025).

<sup>28</sup> Véase al respecto la Declaración del Comité Ejecutivo de la Asociación Médica alemana, adoptada en su reunión de 14 de enero de 2025, que hace un recorrido por las distintas aplicaciones de la IA en medicina, *Deutsches Ärzteblatt*, 4/2025 op. cit. pp. 7 ss.

<sup>29</sup> ROMEO CASABONA, Carlos María, /NICOLÁS JIMÉNEZ, Pilar, *Retos éticos y necesidades normativas en la actividad asistencial en medicina personalizada*, Fundación Instituto Roche, Madrid, 2018. Abordando la cuestión desde la óptica de la medicina personalizada señalan: “El ser humano ha de permanecer como señor de sus propias decisiones, respecto a las que las diversas herramientas de precisión sobre el diagnóstico, pronóstico y tratamiento no deberían sobrepasar

Obviamente lo anterior es tan sólo el punto de partida de la admisibilidad del recurso a este tipo de sistemas, sin que el respeto de tal regla suponga ni mucho menos la eliminación de los riesgos asociados a su uso en el ámbito de la medicina. De hecho, de aquellos riesgos ofrece buena cuenta no sólo una abundante literatura sobre la materia, sino diversos documentos oficiales que han puesto sobre el tapete la necesidad de estar prevenidos frente, al menos, una dualidad de frentes problemáticos. El primero, apunta al posible retroceso que el empleo de tal tecnología puede producir en el reconocimiento de alguno de los derechos del paciente ampliamente consolidados. Se cuentan entre ellos el derecho a su autonomía o el derecho a la información, presupuesto de la validez del consentimiento a cualquier acto médico. El segundo, advierte de la necesidad de adaptar los clásicos esquemas de responsabilidad por mala praxis médica plasmada en un resultado lesivo al nuevo escenario que ofrece de la irrupción de la IA. Las consideraciones que siguen se ocupan de todo ello, planteando en relación con cada una de estas cuestiones tanto sus principales dificultades como una serie de propuestas para afrontarlas.

## **II.- La incidencia de la IA en los derechos del paciente**

### **1- El derecho a la información y autonomía del paciente**

El actual modelo de relación médico-paciente que inspira nuestro sistema se basa decididamente en la comprensión de éste como un sujeto portador de una serie de derechos que vertebran su relación con los distintos agentes implicados en el proceso asistencial. El más básico de todos ellos es el de estar informado sobre su estado de salud, la terapia que en su caso precise, las distintas opciones de las que disponga, los riesgos de cada acto médico, así como los asociados a la ausencia de tratamiento. Son, en definitiva, aspectos que se vinculan con el conocimiento del proceso decisorio del médico y que, aun expresados en un lenguaje no científico, tienen extraordinario valor de cara al ejercicio del derecho del paciente a decidir someterse o no al acto médico en cuestión, con la consiguiente facultad de rechazarlo y, en su caso, de revocar el consentimiento previamente prestado. Este haz de facultades es reconducible de forma genérica al derecho a la autonomía del paciente, cuya plasmación normativa se encuentra en la Ley 41/2002, de 14 de noviembre, básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica, que contempla todo un catálogo de derechos y obligaciones del paciente directamente aplicables al específico ámbito que ahora se trata<sup>30</sup>.

El respeto de este derecho básico pudiera resultar comprometido en el ejercicio de la medicina asistida por IA por una serie de aspectos que, básicamente, se reconducen a la incapacidad de estos sistemas para tener en cuenta la gama de componentes subjetivos que están presentes en una decisión médica. Esta, en efecto, no consiste solo en un aséptico resultado fruto de la atención a datos objetivos, sino que también ha de tener en cuenta las circunstancias que rodean al paciente, inclusiva de sus preferencias vitales. También sobre ellas habrá de ser oído para que encuentren reflejo en la terapia que recibe, lo que mal se concilia con decisiones basadas en un arsenal de datos objetivos.

---

su estricta función de asistentes y auxilio del profesional". Específicamente en relación con la incompatibilidad de las decisiones automatizadas con los esquemas propios de la medicina personalizada, pp. 36 ss.

<sup>30</sup> En su artículo 2, bajo la rúbrica *Principios generales*, reconoce a lo largo de sus distintos apartados, entre otros, el derecho al respeto a la autonomía de la voluntad, que reclama el consentimiento del paciente o usuario una vez cuente con la información adecuada para prestarlo, el derecho a decidir libremente después de contar con la información adecuada entre las opciones clínicas disponibles, así como el derecho a negarse al tratamiento. Por su parte, conforme al art. 4.1., "Los pacientes tienen derecho a conocer, con motivo de cualquier actuación en el ámbito de su salud, toda la información disponible sobre la misma, salvando los supuestos exceptuados por la Ley. Además, toda persona tiene derecho a que se respete su voluntad de no ser informada. La información, que como regla general se proporcionará verbalmente dejando constancia en la historia clínica, comprende, como mínimo, la finalidad y la naturaleza de cada intervención, sus riesgos y sus consecuencias". Sobre el alcance de estos derechos tuve ocasión de ocuparme en *La responsabilidad penal del médico*, 2ª ed., Valencia 2008, pp. 19 ss.



Con ser fundada esta amenaza denunciada por la literatura<sup>31</sup>, lo cierto es que la tantas veces repetida irrenunciabilidad de la intervención humana relativiza en buena medida su alcance, en tanto que habrá de ser el profesional quien en última instancia valide o no los resultados del sistema, decisión en la habrá de tener en cuenta no sólo las características personales del paciente, sino también sus preferencias y singularidades, tanto físicas como psicológicas. Por ello, en realidad, lo que reclama la garantía de la autonomía del paciente en el ámbito que interesa es la adaptación a este nuevo escenario de los clásicos aspectos sobre los que debe recaer la información como presupuesto del ejercicio de aquel derecho. En concreto, el paciente ha de ser informado ante todo del empleo de tales herramientas, del papel que desempeña el profesional en su uso, de las razones que aconsejan el recurso a dichos sistemas en el caso concreto, así como de las ventajas e inconvenientes asociadas a su empleo en comparación con los métodos tradicionales<sup>32</sup>. Más allá de estos extremos básicos, un aspecto importante sobre el que igualmente habrá de extenderse la información será el alcance de los posibles resultados que arroje la herramienta empleada con función de diagnóstico, especialmente cuando se nutra de datos genéticos. Los hallazgos inesperados a que puede dar paso motivan la conveniencia de extender la información a tal posibilidad, a efectos de recabar el consiguiente consentimiento del paciente, tanto respecto a su averiguación como, en su caso, a la comunicación de sus resultados.

Con todo, la adaptación de las exigencias del deber de información a las peculiaridades de la IA no despeja las dudas en torno a si su empleo permite ofrecer al paciente un grado de información equiparable al que recibe en el marco del ejercicio de la medicina tradicional. El germen del principal obstáculo para ese reconocimiento se encuentra ya en una de las características inherentes al funcionamiento del sistema; a saber, la que se ha dado en llamar como la opacidad del proceso por medio del cual llega a emitir un determinado resultado. En efecto, la falta de transparencia es en la actualidad un rasgo inevitable de las herramientas de IA, definida usualmente como una *black box* o caja negra<sup>33</sup>, en la que resulta imposible identificar los factores de entrada tomados en consideración y el modo en que son relacionados y procesados para llegar al resultado. De hecho, el proceso de razonamiento resulta inaccesible incluso para los programadores del sistema, entre otras razones, porque los algoritmos iniciales van creando a su vez otros de mayor complejidad hasta dar paso a una cadena en la que tan solo es posible leer una secuencia de algoritmos y operaciones algebraicas, sin traducción directa en los esquemas del razonamiento humano.

Así, por ejemplo, si aquel expresara la conveniencia de un ingreso hospitalario al paciente aquejado de una infección renal, resultaría imposible conocer si la decisión está basada en un cálculo probabilístico de la evolución de la enfermedad conforme a los datos obrantes en el sistema sobre su comportamiento estándar, si tiene en cuenta la previsible interacción de otros factores o complicaciones que pudiera influir en ella, si se asienta en exclusiva en las características del paciente o grupo de pacientes, o incluso si el resultado se basa en la predisposición genética de aquél. La emisión en tales condiciones del consentimiento

---

<sup>31</sup> De ella da cuenta, por ejemplo, ISLA GÓMEZ, Julián, "Aplicaciones de la inteligencia artificial en salud", *Inteligencia artificial y Bioética*, (ed. Amo Usanos, Rafael), Universidad Pontificia de Comillas, Madrid, 2023, pp. 37 s.

<sup>32</sup> En relación con el empleo de los llamados robots quirúrgicos, sobre este aspecto insisten distintos documentos, entre ellos el conocido como proyecto *RoboLaw*, cuyos resultados se publicaron en la "Guidelines for Regulating Robotics". Sobre él habrá ocasión de volver más adelante. Interesa de momento tan solo dar cuenta de que el documento subraya la importancia de que el consentimiento del paciente se emita desde el conocimiento del empleo de este tipo de técnicas, debiéndose garantizar la explicación del papel del robot, sus limitaciones y riesgos, así como las alternativas disponibles. Expresamente el paciente deberá conocer que el uso de la tecnología avanzada no elimina los riesgos de la intervención, y todo ello en un lenguaje asequible para él. Señala para ello la conveniencia de establecer protocolos estandarizados que garanticen todos estos aspectos, pp. 74 ss. Disponible en: [http://www.robolaw.eu/RoboLaw\\_files/documents/robolaw\\_d6.2\\_guidelinesregulatingrobotics\\_20140922.pdf](http://www.robolaw.eu/RoboLaw_files/documents/robolaw_d6.2_guidelinesregulatingrobotics_20140922.pdf), (última consulta: 1 de agosto de 2025).

<sup>33</sup> Entre otros, WATSON, David / KRUTZINNA, Jenny / BRUCE, Ian / GRIFFITHS, Christopher / MCINNES, Iain / BARNES, Michael / FLORIDI, Luciano, "Clinical applications of machine learning algorithms: beyond the black box", disponible en: <https://www.bmj.com/content/364/bmj.l886.long>, (última consulta: 17 de junio de 2025). Destacan los autores las distorsiones a que pueden conducir sus resultados sin una adecuada explicación del proceso por el que se llega al resultado. Ejemplifican con el caso del algoritmo que predice el riesgo de morir de una neumonía, que arroja resultados más bajos en el caso de los pacientes asmáticos, sin ofrecer la explicación de que ello no es debido a la envergadura de la enfermedad, sino al tratamiento intensivo que reciben estos pacientes, p. 4.

descansaría entonces necesariamente en una información previa incompleta, lo que a su vez comprometería seriamente la validez del consentimiento así prestado<sup>34</sup>. Las dudas se diversifican en varias preguntas que sugieren a su vez las distintas opciones posibles. Son las que, en un extremo, plantean si la aceptación de estas limitaciones de información debiera llevar a invalidar el recurso a aquellos sistemas; las que en el lado opuesto cuestionan la conciliación del clásico reconocimiento del derecho a la información con el nuevo modelo y, por último, las que en una vía intermedia pudieran plantear la conveniencia de una reformulación de aquellos derechos para armonizarlos con el nuevo modelo.

Precisamente en la línea de reformular las exigencias del derecho a la información se presenta el debate en torno a la existencia de un supuesto derecho a la *explicación* del paciente, no necesariamente coincidente con el derecho a la información. Se hace referencia con él a la hipotética pretensión del interesado de conocer, no sólo las características de su enfermedad y el tratamiento que reclama, así como las opciones o alternativas y, en general, los riesgos asociados a cada uno de estos extremos. Se reclamaría con él adicionalmente el conocimiento de las razones por las cuales el sistema alcanza la decisión médica de que se trate<sup>35</sup>, por mucho que ese derecho se presente necesariamente adaptado a la capacidad de comprensión del paciente. El asidero para su reconocimiento bien pudiera encontrarse en una serie de documentos<sup>36</sup>, entre los que destaca el Reglamento General de Protección de Datos (RGPD). Aun carente de valor vinculante, su Considerando 71 recalca el derecho de todo individuo a no ser objeto de una decisión que incida, entre otros aspectos, en la salud, y que se haya basado exclusivamente en el tratamiento automatizado. A partir de tal premisa contempla el derecho del interesado a “recibir una explicación de la decisión tomada después de tal evaluación”.

Sin desconocer el valor de tales previsiones en orden a fundamentar las bases para un supuesto derecho a la explicación, lo cierto es que la repetida comprensión de la IA como mera herramienta al servicio del médico, nunca como modelo sustitutivo de éste, hace decaer en buena medida la intensidad del debate en el específico ámbito que interesa. Así resulta, en efecto, en tanto que aquellas previsiones toman como referente las decisiones basadas exclusivamente en procesos

---

<sup>34</sup> Sólo en este sentido comparto la crítica que se ha formulado en la doctrina relativa a la eventual afectación de la IA a la autonomía del paciente. De modo más amplio véase ROMEO CASABONA, Carlos María/ LAZCOZ MORATINOS, Guillermo, “Inteligencia artificial aplicada a la salud: ¿qué marco jurídico?”, en *Revista de Derecho y Genoma Humano*, 52/2020, pp. 152 ss., quienes alertan del riesgo de que pueda llegar “a imponerse el imperativo tecnológico sobre la libertad del paciente de aceptar o de rechazar un tratamiento”, p. 153, planteándolo como un riesgo asociado a la elaboración de perfiles de los pacientes y su consiguiente clasificación o estratificación, arrojando decisiones rutinarias o automatizadas. Véase también GALVIN GORDILLO, M., “Análisis legal del uso de los robots en medicina”, *Ius et Scientia*, vol. 9, núm. 1, 2023, marg. 133.

<sup>35</sup> Cuestionan tal derecho WACHTER S./ MITTELSTADL B./ FLORIDI, I, “Why a right to explanation of automater decisionmaking does not exist in the general data protection regulation”, *International Data Privacy Law*, 2017, pp. 76 ss., disponible en: <https://academic.oup.com/idpl/article-abstract/7/2/76/3860948?redirectedFrom=fulltext>, (última consulta: 28 de mayo de 2025). Sostienen que el RGPD en su forma actual, no implementa un derecho a la explicación, sino un «derecho a ser informado» limitado, planteando la mayor protección que otorga la comprensión de una prohibición de decisiones opacas en lugar de la plasmación de un derecho a la explicación. Véase también FERRETI, A.M./SCHEIDER, M./ BLASIMME, A., “Machine Learning in Medicine. Opening the new data protection Black Box”, *European Data Protection Law Review*, núm. 4, vol. 3, 2018, pp. 320 ss. Tras detectar los distintos tipos de opacidad de los sistemas (falta de divulgación, opacidad epistémica y opacidad explicativa, pp. 325 ss.), considera que, si bien el RGPD contiene disposiciones ordenadas a la transparencia y la comunicación de información significativa sobre el procedimiento automatizado de datos, no consagra como tal un derecho a la explicación (pp. 330 ss).

<sup>36</sup> Las bases del reconocimiento del eventual derecho a la explicación pudieran igualmente fundamentarse en el apartado 1 del artículo 86 del Reglamento IA, que consagra el derecho del interesado a conocer el papel que ha tenido el sistema de IA en la toma de decisiones: “Toda persona que se vea afectada por una decisión que el responsable del despliegue adopte basándose en los resultados de salida de un sistema de IA de alto riesgo que figure en el anexo III, con excepción de los sistemas enumerados en su punto 2, y que produzca efectos jurídicos o le afecte considerablemente del mismo modo, de manera que considere que tiene un efecto perjudicial para su salud, su seguridad o sus derechos fundamentales, tendrá derecho a obtener del responsable del despliegue explicaciones claras y significativas acerca del papel que el sistema de IA ha tenido en el proceso de toma de decisiones y los principales elementos de la decisión adoptada”. Con el carácter de Recomendaciones, puede verse el documento de Directrices éticas para una IA confiable, disponible en <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (última consulta: 8 de julio 2025). Se trata de un documento elaborado por un grupo independiente de expertos de alto nivel en IA, creado por la Comisión Europea en 2018 para ofrecer unas guías éticas del uso de la IA. Entre los cuatro principios básicos para una IA fiable contempla la *explicabilidad*. Reconociendo la imposibilidad de explicar siempre por qué un modelo ha arrojado un determinado resultado, presenta aquel principio como esencial para mantener la confianza en el sistema y para poder impugnar adecuadamente una decisión (punto 53). Su garantía se contempla esencial “cuando un sistema de IA tenga un impacto significativo en la vida de las personas” (punto 77).

completamente automatizados, excluidos del ámbito de la salud<sup>37</sup>. Por ello, sin necesidad siquiera de entrar a buscar modelos de solución basados en paralelismos entre la oscuridad del sistema y las propias del razonamiento humano<sup>38</sup>, el espacio que viene a cubrir el supuesto reconocimiento del derecho a la explicación queda ya satisfecha con la debida información al paciente de las razones por las cuales el profesional, como centro último de decisión, se acoge o se aparta de los criterios o resultados que arroja el sistema. Dicho de otro modo, el cumplimiento del derecho a la información se garantiza en tanto se ponga directamente en relación con la explicación que ofrezca el profesional sobre las razones de su decisión apoyadas en el sistema, nunca en relación directa con el modo de razonar de éste, por definición ininteligible en la actualidad<sup>39</sup>.

Cuestión distinta será, obviamente, la carga y la consiguiente fuente de responsabilidad que en potencia puede suponer para el sanitario fundamentar un criterio distinto al que arrojen los algoritmos. Como habrá ocasión de tratar al abordar el diseño de los esquemas de responsabilidad del sanitario por un resultado lesivo en este nuevo marco de la IA, la conformidad o no a Derecho de su actuación requiere en última instancia el cotejo de los resultados del sistema con las reglas de experiencia propias de la medicina tradicional. Serán exclusivamente las razones que han motivado al médico a la toma de decisión conforme a aquellas reglas las que integren el ámbito del derecho a la información del paciente.

Sobre todo ello habrá ocasión de volver. De momento, y sin abandonar las peculiaridades del derecho a la información en el ámbito de la práctica de la medicina asistida por IA, alguna mención merece una cuestión que ya fue parcialmente anunciada y que no sólo es distinta de la anterior, sino que se proyecta en una dirección opuesta. Se trata de la posible conveniencia de articular mecanismos para evitar un exceso de información que pudiera ponerse a disposición del paciente, situación llamada a darse en la práctica con una frecuencia nada desdeñable a la vista del ingente volumen de resultados que pueden generar estas herramientas<sup>40</sup>, especialmente en el caso de enfermedades condicionadas por una predisposición genética. Surge entonces la duda en torno a si el paciente debe ser informado de todos los aspectos predecibles en relación con su estado de salud o si, por el contrario, sería conveniente su limitación a aquellos que específicamente motivaron el análisis, en aras a garantizar otros intereses, entre los que se cuenta su salud psíquica y consiguiente estabilidad emocional. La cuestión se revela especialmente importante en relación con el ámbito que aquí se trata, a la vista del ampliamente denunciado temor a derivar en una medicina despersonalizada, en la que el valor del acceso a la información prevalezca sobre la empatía y el trato humano que reclama el ejercicio de la medicina<sup>41</sup>.

---

<sup>37</sup> Salvo las excepciones contempladas en las letras a y g del art. 9.2 (consentimiento explícito o interés público importante). De acuerdo con las previsiones del art. 14, entre los aspectos sobre los que debe recaer la información cuando los datos personales no se han obtenido del interesado se cuenta la existencia de las decisiones automatizadas a que se refieren los apartados 1 y 4 del art. 22, relativos a las decisiones basadas *exclusivamente* en un proceso automatizado. En tal caso el citado art. 14 incluye el deber de facilitar la "información significativa sobre la lógica aplicada".

<sup>38</sup> No es infrecuente llamar la atención sobre el hecho de que la opacidad propia de los mecanismos de decisión de la IA no debiera verse como algo distinto a la que en general es propia de la decisión médica en la práctica tradicional, donde también las razones que llevan al profesional a adoptar la medida de que se trate sólo son relevantes para el paciente en tanto puedan incidir en la toma de una decisión, no desde luego a modo de una suerte de "información científica". Por todos, ROMEO CASABONA C.M., /LAZCOZ MORATINOS G., "Inteligencia artificial aplicada a la salud: ¿qué marco jurídico?", *op. cit.*, pp. 149 s.

<sup>39</sup> Si bien es cierto que se han desarrollado distintos sistemas de IA explicables, los conocidos como sistemas XAI. Es el caso, por ejemplo, del sistema *Merative*, (IBM Watson Health) que ofrece explicación de las decisiones. Así, en relación con el diagnóstico por imagen, ofrece explicación sobre las razones o características que llevaron al sistema a identificar la zona.

<sup>40</sup> ROMEO CASABONA, C.M., "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*, número extraordinario, 2019, p. 90.

<sup>41</sup> Sobre la necesidad de mantener estos valores, puede verse MELLADO ORELLANA, Raúl / ESCOBAR ARRIAGA, Elizabeth / DE LA MORA MOLINA, Héctor / DÍAZ GREENE, Enrique Juan / RODRÍGUEZ WEBER, Federico Leopoldo: "Estos sistemas de cognición deben ser creados con la meta de potenciar la inteligencia humana. Es por esto que las cualidades humanas como el razonamiento moral, la compasión y empatía permanecen como fundamentales e irremplazables a la hora de la toma de decisiones", en "Estudio de concordancia entre el sistema Watson for Oncology y la práctica clínica en pacientes con cáncer de mama dentro del Hospital Angeles Pedregal", *Acta Médica Grupo Angeles*, vol. 21, núm. 4, 2023; disponible en: <https://www.medigraphic.com/pdfs/actmed/am-2023/am234i.pdf> p. 341, (última consulta: 25 de agosto de 2025).

En realidad, la respuesta pasa por trasladar al ejercicio de la medicina asistida por IA las previsiones y pautas ensayadas en relación con la información obtenida con métodos tradicionales. Cumple en este sentido recordar que, para orientar la actuación del profesional en tales casos, nuestro Derecho positivo cuenta con alguna disposición al respecto. En concreto, el apartado 4 del art. 5 de la Ley 41/2002, básica reguladora de la Autonomía del Paciente y Derechos y Obligaciones en materia de información y Documentación clínica contiene una previsión que supone de algún modo la introducción en nuestro Derecho del llamado *privilegio terapéutico del médico*. Tras dejar sentado en su art. 4.2 que la información clínica tiene que ser verdadera, dispone: "El derecho a la información sanitaria de los pacientes puede limitarse por la existencia acreditada de un estado de necesidad terapéutica. Se entenderá por necesidad terapéutica la facultad del médico para actuar profesionalmente sin informar antes al paciente, cuando por razones objetivas el conocimiento de su propia situación pueda perjudicar su salud de manera grave. Llegado este caso, el médico dejará constancia razonada de las circunstancias en la historia clínica y comunicará su decisión a las personas vinculadas al paciente por razones familiares o de hecho".

Como tuve ocasión de sostener en otro lugar, el presupuesto que legitime tal opción no debe vincularse de modo exclusivo con el riesgo de alterar la tranquilidad o estabilidad emocional del paciente, sino con la utilidad misma de la información. Quiere esto decir que la omisión de información sólo será admisible allí donde no prive al paciente de la opción de consentir en un determinado tratamiento o medida paliativa. Resulta así que, dejando a salvo el caso en que el paciente hubiera ejercido su derecho a no saber<sup>42</sup>, la facultad del sanitario para acogerse a dicho privilegio terapéutico encontrará su límite allí donde existiera alguna terapia para aumentar la esperanza de vida o reducir los sufrimientos de la enfermedad, medida que reclamaría entonces el consentimiento y la consiguiente información del paciente<sup>43</sup>.

Resta todavía siquiera por mencionar una cuestión adicional que pudiera plantearse en la práctica motivada por el funcionamiento de estos sistemas. Se trata de la eventual procedencia de transmitir los resultados que arrojen, no ya al propio paciente, sino a terceras personas que pudieran estar interesadas en conocerlos, por incidir estos en la información relativa a su estado de salud. El supuesto remite a un escenario de conflicto entre los intereses del paciente y el de terceros en el caso de que aquéllos se hayan obtenido a partir del análisis del perfil genético de aquel. Como es sabido, los datos genéticos se caracterizan por su carácter predictivo y generacional, de tal modo que su acceso o revelación afecta a la vida futura del titular, así como a la de los miembros de su familia. Es lo que sucedería, por ejemplo, cuando el sistema arrojarase como resultado una mutación genética capaz de dar paso a una enfermedad grave cuyo riesgo de desarrollo sea compartido por todos los miembros de una misma familia, aun cuando con un índice de probabilidad distinto dependiendo del grado de parentesco. La conveniencia de facilitar la información obtenida se plantearía entonces no ya sólo en relación con el paciente, sino con esos otros familiares interesados en el acceso a aquellos resultados como medio de conocer, prever o completar los relativos a su propio estado de salud. En tal escenario, la línea de solución no debiera diferir de la que en general se ha ensayado para los supuestos en los que el sujeto se somete a análisis genético que revela descubrimientos inesperados, casos de los que tuve ocasión de ocuparme en otro lugar a cuyas conclusiones me remito en este punto<sup>44</sup>. Bastaría con concluir ahora que solo en situaciones muy excepcionales donde pudieran apreciarse los presupuestos de un estado de necesidad, procedería plantear la revelación a los familiares, en los términos previstos en los arts. 4 y 49.2 de la Ley de Investigación Biomédica<sup>45</sup>.

---

<sup>42</sup> Reconocido en el art. 9.1 de la Ley 41/2002: "Cuando el paciente manifieste expresamente su deseo de no ser informado, se respetará su voluntad haciendo constar su renuncia documentalmente, sin perjuicio de la obtención de su consentimiento previo para la intervención". Específicamente en relación con los análisis genéticos, el derecho se encuentra reconocido en el art. 49.2 de la Ley de Investigación Biomédica: "Cuando el sujeto fuente haya ejercido el derecho a no ser informado de los resultados de un análisis genético sólo se suministrará la información que sea necesaria para el seguimiento del tratamiento prescrito por el médico y aceptado por el paciente".

<sup>43</sup> De ello tuve ocasión de ocuparme en *La responsabilidad penal del médico*, op. cit., pp. 116 ss.

<sup>44</sup> GÓMEZ RIVERO, M.C., "Los límites del deber de sigilo del médico en las situaciones de riesgo: especial referencia al ámbito de la genética", *Revista de Derecho y Genoma Humano*, 2007, núm. 26, pp. 43 ss.

<sup>45</sup> Art. 4: "Se respetará el derecho de la persona a decidir que no se le comuniquen los datos a los que se refiere el apartado anterior, incluidos los descubrimientos inesperados que se pudieran producir. No obstante, cuando esta información, según criterio del médico responsable, sea necesaria para evitar un grave perjuicio para su salud o la de



## 2.- La eventual afectación de otros derechos del paciente

Con ser importante, no es el derecho a la autonomía del paciente el único que reclama una atención singular en el marco de la medicina asistida por la IA. Otro de los aspectos que merece alguna reflexión por el riesgo de suponer una quiebra de los derechos básicos de aquél encuentra su origen en el punto de partida del funcionamiento de los sistemas de IA; a saber, el llamado *big data* o el almacenamiento de grandes volúmenes de datos de los que se nutren aquellas herramientas<sup>46</sup> y a partir de cuyo procesamiento pueden arrojar sus resultados: desde los indicadores de frecuencia de determinadas patologías hasta su patrón de evolución o posibilidades de supervivencia, presentándose en la actualidad como una herramienta irrenunciable con los más variados fines<sup>47</sup>.

La recogida de esa información masiva permite a su vez el funcionamiento del llamado sistema de aprendizaje automático o *machine learning*, un método inductivo concebido como una forma de aprendizaje de la máquina, de modo que a partir de la información (datos) que recibe puede arrojar el trazo de patrones y, en definitiva, de nuevas reglas conforme a las que ofrecer ulteriores predicciones<sup>48</sup>. Las amenazas a los derechos del titular que plantea este modo de operar del sistema se proyectan, al menos, en tres frentes denunciados ampliamente en la literatura sobre la materia. El primero, el riesgo de que el manejo de esos datos suponga un trato discriminatorio para el paciente en la atención médica que recibe; el segundo, la eventual quiebra de la privacidad del titular, en tanto que los datos incorporados no sólo podrían permitir la asociación a su identidad, sino también inferir a partir de ellos información adicional sobre el sujeto de que se trate; el tercero, incidir en el riesgo ya referido de practicar una medicina por completo deshumanizada. Sobre este último potencial efecto perverso del sistema han advertido, entre otros autores<sup>49</sup>, ROMEO CASABONA y NICOLÁS JIMÉNEZ quienes atendiendo específicamente a las predicciones basadas en datos de carácter genético se plantean si el uso de herramientas basadas en el manejo de datos llevarían a justificar la exclusión de un paciente de un determinado tratamiento allí donde el sistema arroja una predicción de ineficacia de la terapia en cuestión, efecto que a su vez derivaría en última instancia en el riesgo de discriminación del sujeto de que se trate<sup>50</sup>.

---

sus familiares biológicos, se informará a un familiar próximo o a un representante, previa consulta del comité asistencial si lo hubiera. En todo caso, la comunicación se limitará exclusivamente a los datos necesarios para estas finalidades". Conforme al art. 49.2: "Cuando esta información sea necesaria para evitar un grave perjuicio para la salud de sus familiares biológicos, se podrá informar a los afectados o a su representante legalmente autorizado. En todo caso, la comunicación se limitará exclusivamente a los datos necesarios para estas finalidades".

Sobre el distinto presupuesto de aplicación de una y otra previsión y sus diferencias procedimentales puede verse ROMEO CASABONA, C.M.,/NICOLÁS, P., *Retos éticos y necesidades normativas en la actividad asistencial en medicina personalizada*, op. cit., pp. 51 ss.

<sup>46</sup> Definidos como "el tratamiento de grandes volúmenes de datos mediante algoritmos matemáticos con el fin de establecer correlaciones entre ellos, predecir tendencias y tomar decisiones". Es la definición que ofrece el Observatorio de Bioética y Derecho, *Documento sobre bioética y Big Data de salud: explotación y comercialización de los datos de los usuarios de la sanidad pública*, Universidad de Barcelona, 2015, p. 33, disponible en: <https://diposit.ub.edu/dspace/bitstream/2445/104585/1/08209cc.pdf>, (última consulta: 30 de mayo de 2025).

<sup>47</sup> Sobre ellos puede verse OLIVÁN MUZÁS, A./ALBUNDIO ESCALONA, S./ ARIAS BASTO, Y.C./ JIMÉNEZ HERNÁNDEZ, S., "El uso de Big Data en la investigación médica", *Revista Ocrono*, 2024, disponible en <https://revistamedica.com/uso-big-data-investigacion-medica/> (última consulta: 10 de abril de 2025), donde ofrecen un recorrido por el uso de la herramienta con fines de investigación farmacológica, en la medicina personalizada, en epidemiología y vigilancia de enfermedades, y la optimización de recursos y gestión de salud.

<sup>48</sup> Quizá la rama más avanzada del aprendizaje automático sea el llamado aprendizaje profundo, *Deep learning*, inspirado en el modo de funcionamiento del cerebro humano. Al respecto y en relación con el ámbito médico que aquí interesa, véase por ejemplo, ROMEO CASABONA, Carlos María /LACÓZ MORATINOS, Guillermo, "Inteligencia artificial aplicada a la salud: ¿qué marco jurídico?", op. cit., pp. 142 ss; NÚÑEZ REIZ, Antonio / ARMENGOL DE LA HOZ, Miguel Ángel / SÁNCHEZ GARCÍA, Miguel, "Big Data Analysis y Machine Learning en medicina intensiva", *Medicina Intensiva*, 2028, disponible en: <http://www.medicinaintensiva.org>, (última consulta: 1 de junio de 2025).

<sup>49</sup> CÁRCAR BENITO, Jesús Esteban, "La asistencia sanitaria dentro del ámbito de la inteligencia artificial (IA): el problema de los derechos de los pacientes", *El tiempo de los Derechos*, vol.1, 2020, p. 2018.

<sup>50</sup> ROMEO CASABONA, Carlos María /NICOLÁS JIMÉNEZ, Pilar, quienes denuncian "la posible utilización de la información genética para limitar el acceso a tratamientos médicos que se consideren muy probablemente fútiles en el caso de un paciente concreto. Imaginemos que un algoritmo de predicción augura (con un alto índice de certeza, pero no con certeza absoluta) que una terapia concreta será ineficaz en el caso de un paciente concreto. ¿Estaríamos legitimados para negarle ese tratamiento sobre la base de las conclusiones extraídas del uso de sus datos genéticos? ¿O hemos de entender que ello supondría una forma de exclusión o de discriminación?", *Retos éticos y necesidades normativas en la actividad asistencial en medicina personalizada*, op. cit., p. 71.



Precisamente enlazando con el peligro de que el uso de tales sistemas lleve a aceptar actuaciones discriminatorias, hay que reconocer que aquél se asocia en esencia a la fase de introducción o almacenamiento de datos, que de realizarse de forma sesgada, pone en entredicho la prohibición de cualquier práctica discriminatoria<sup>51</sup>. Sobre ella alerta no sólo una abundante doctrina especializada<sup>52</sup>, sino distintos documentos<sup>53</sup> que advierten del importante foco discriminatorio que puede suponer que los operadores del sistema lo nutran con datos que no sean inclusivos, esto es, representativos de todos los sectores de población.

Como se anunciaba, no son, desde luego, los anteriores los únicos riesgos asociados al almacenamiento de una ingente cantidad de datos. La misma preocupación se plantea en relación con la eventual quiebra de la privacidad del paciente, que puede ser comprometida no sólo desde el punto de vista de su clásica dimensión individual, sino también, como se ha advertido en la literatura, de su dimensión grupal<sup>54</sup>. Esta última sería, en realidad, una reinterpretación de aquella, desde el momento en que el almacenamiento masivo de datos que identifican a un grupo da paso a una nueva dimensión del clásico derecho a la privacidad, que tomaría entonces por base las características del colectivo. Bastaría pensar, por ejemplo, en el caso de que el desarrollo de una determinada enfermedad se asociara a un grupo de población, con su consiguiente exposición a las consecuencias que se derivasen del conocimiento del dato. Como sea, y dejando a un lado esta faceta, que en la actualidad no pasa de ser un concepto emergente sin respaldo normativo, resulta incuestionable la eventual vulnerabilidad en este ámbito del clásico derecho a la privacidad. Esta encuentra su germen en el riesgo de la identificación del paciente cuyos datos ha sido introducidos en el sistema a partir de un proceso de asociación. Ciertamente es que se trata de información tratada con garantías iniciales de preservación de la identidad del titular, destacando entre ellas la seudonimización, contemplada como medida garantizadora del anonimato en el RGPD, que permite una disociación reversible de los datos y su titular. Definida en su art. 4.5 como “el tratamiento de datos personales de manera tal que ya no puedan atribuirse a un interesado sin utilizar información adicional, siempre que dicha información adicional figure por separado y esté sujeta a medidas técnicas y organizativas destinadas a garantizar que los datos personales no se atribuyan a una persona física identificada o identificable”, el Reglamento la contempla como medio de protección de datos en la fase de diseño (art. 25), así como en lo relativo al tratamiento de los datos personales almacenados (art. 32).

---

<sup>51</sup> La prohibición de cualquier práctica discriminatoria en la prestación sanitaria encuentra ya su reconocimiento normativo en el apartado 2 del art. 3 de la Ley 14/1986, de 25 de abril, Ley General de Sanidad cuando dispone: “La asistencia sanitaria pública se extenderá a toda la población española. El acceso y las prestaciones sanitarias se realizarán en condiciones de igualdad efectiva”.

<sup>52</sup> La existencia de sesgos en este tipo de sistemas se presenta como uno de sus mayores handicaps, habiéndose detectado tres tipos en atención a su origen: los preexistentes, los técnicos y los emergentes. Sobre ellos puede verse LACÓZ MORATINOS, Guillermo, “Inteligencia artificial, sesgos y salud: dilemas en torno a la propuesta de Reglamento AIA”, *Inteligencia artificial y Bioética* (ed. Lit. Amo Usanos, Rafael) Universidad Pontificia Comillas, Madrid, 2023, p. 184. Entre otros, sobre el riesgo de discriminación, DE MIGUEL BERRAIN, Íñigo, MURSULI YANES, Yenifer, “Sesgos, IA y Biomedicina: un comentario desde la ética y el Derecho”, *Revista de Derecho y Genoma Humano*, núm. 59, julio-diciembre 2023, pp. 129 ss; 134 ss.

<sup>53</sup> Es el caso de las Directrices éticas para una IA confiable, disponible en: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (última consulta: 8 de julio de 2025). Conforme a su punto 44: “En el contexto de la IA, la igualdad implica que el funcionamiento de este tipo de sistemas no debe generar resultados injustamente sesgados (por ejemplo, los datos utilizados para la formación de los sistemas de IA deben ser lo más inclusivos posibles, de forma que estén representados los diferentes grupos de población). Esto también requiere un adecuado respeto de las personas y grupos potencialmente vulnerables, como los trabajadores, las mujeres, las personas con discapacidad, las minorías étnicas, los niños, los consumidores u otras personas en riesgo de exclusión”. Véanse también al respecto las Recomendaciones Sectoriales sobre Inteligencia Artificial (IA) presentadas el 23 de julio de 2020 por el Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial (GEAB IA), disponible en: <https://futurium.ec.europa.eu/en/european-ai-alliance/document/ai-hleg-sectoral-considerations-policy-and-investment-recommendations-trustworthy-ai> (última consulta: 20 de agosto de 2025).

Véase también el Considerando 94 del Reglamento 2022/2065, del Parlamento Europeo y del Consejo, de 19 de octubre, relativo a un mercado único en Servicios Digitales, así como el *Libro Blanco sobre Inteligencia Artificial: un enfoque europeo de la excelencia y la confianza*, publicado el 19 de febrero de 2019.

<sup>54</sup> Sobre los aspectos de esta nueva dimensión véase, entre otros, LAUKYTE, Migle, “Privacidad e inteligencia artificial: cambios e incertidumbres”, *Inteligencia artificial y Bioética* (ed. Lit. Amo Usanos, Rafael), Universidad Pontificia de Comillas, Madrid, 2023, pp. 138 ss; COTINO HUESO, Lorenzo, “Nuevo paradigma en la garantía de los derechos fundamentales y una nueva protección de datos frente al impacto social y colectivo de la inteligencia artificial”, *Derechos y garantías ante la inteligencia artificial y las decisiones automatizadas* (dir. Cotino Hueso, Lorenzo), Thomson Reuters Aranzadi, Pamplona 2022, pp. 86 ss.

No obstante, pese a que el uso de tal técnica se presenta como una importante garantía de anonimato frente a la fragilidad del simple almacenamiento de datos<sup>55</sup>, la doctrina así como distintos documentos<sup>56</sup> han puesto sobre el tapete el riesgo nada despreciable que para la preservación de la privacidad supone el carácter potencialmente reversible de la disociación de los datos, esto es, la posibilidad de volver a identificar a su titular conforme a la referida técnica de *seudonimización*<sup>57</sup>, con la consecuencia adicional de que a él se asocie no sólo la información explícitamente contenida en los datos de que dispone el sistema, sino también otros deducibles de modo implícito de aquella información de partida<sup>58</sup>. De hecho, la Disposición Adicional Decimoséptima de la LO 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales, relativa al tratamiento de datos personales en el ámbito de la investigación en salud pública y biomédica, contempla la posibilidad de que pueda “procederse a la reidentificación de los datos en su origen, cuando con motivo de una investigación que utilice datos seudonimizados, se aprecie la existencia de un peligro real y concreto para la seguridad o salud de una persona o grupo de personas, o una amenaza grave para sus derechos o sea necesaria para garantizar una adecuada asistencia sanitaria”.

Se trata, en definitiva, de un riesgo que, al igual que sucede con los anteriormente descritos, representa un desafío consustancial al empleo de estos sistemas. No es por ello de extrañar que las únicas medidas que puedan proponerse no se orienten, por su imposibilidad, a la eliminación de tales focos de peligro como tal, sino al trazo de las condiciones bajo las cuales pueda considerarse admisible. Básicamente, se trata, además de garantizar la adopción de una serie de medidas organizativas y técnicas para garantizar la seguridad de los datos<sup>59</sup>, de la exigencia de contar con un consentimiento reforzado del paciente, comprensivo no sólo del asentimiento en el riesgo de la reidentificación, sino también de la eventual obtención de información adicional a la contenida en los datos<sup>60</sup>. Sólo rodeado de tales garantías, los riesgos descritos podrán considerarse permitidos.

---

<sup>55</sup> Un ejemplo práctico lo ofrece la plataforma de análisis genético *23andMe*, que en 2007 inició la venta de kits para análisis genéticos. Mediante una simple muestra de saliva es capaz de ofrecer información, entre otros aspectos, sobre los riesgos genéticos de padecer determinadas enfermedades. En 2023 fue objeto de un ataque informático que dejó al descubierto una importante brecha de seguridad del sistema. Los hackers accedieron a la información genética de los usuarios y en general a todos sus datos personales. Sobre las carencias detectadas en la seguridad del sistema, puede verse el Informe conjunto de los Comisionados de Privacidad de Canadá y Reino Unido sobre 23andMe, disponible en: <https://www.dlapiper.com/en-us/insights/publications/2025/06/compliance-recommendations-from-privacy-commissioners-report-on-23andme-data-breach> (última consulta: 12 de agosto de 2025).

<sup>56</sup> Entre ellos se cuenta el aprobado por la Agencia de la Unión Europea para la Ciberseguridad (ENISA) en marzo de 2022 *La adopción de técnicas de seudonimización. El caso del sector sanitario*, disponible en: <https://www.enisa.europa.eu/publications/deploying-pseudonymisation-techniques> (última consulta: 30 de julio de 2025).

<sup>57</sup> Sobre las diferencias con la anonimización, en la que no existen datos adicionales del interesado de modo que es un procedimiento “razonablemente irreversible”, se ocupa el considerando 26 del Reglamento, ciñendo el ámbito de aplicación del Reglamento a aquellos otros datos. Un estudio de estas previsiones puede verse en ROMEO CASABONA, Carlos María, quien se ocupa ampliamente de las posibilidades de reversibilidad conforme a una y otra técnica, “Revisión de las categorías jurídicas de la normativa europea ante la tecnología del *big data* aplicada a la salud”, *op. cit.*, pp. 108 ss. Sobre el riesgo de reidentificación véase también ROMEO CASABONA, Carlos María, /NICOLÁS JIMÉNEZ, Pilar, *Retos éticos y necesidades normativas en la actividad asistencial en medicina personalizada*, *op. cit.*, pp. 33 ss.

<sup>58</sup> De este riesgo advierten ROMEO CASABONA, Carlos María / LAZCOZ MORATINOS, Guillermo, en tanto que el titular de los datos pudiera quedar identificado a partir del cruce de datos anónimos o seudonimizados con otros personales, “Inteligencia artificial aplicada a la salud: ¿qué marco jurídico?”, *op. cit.*, pp. 146 ss.

<sup>59</sup> A la adopción de estas y otras medidas adicionales se refiere ya el Considerando 28 RGPD, cuando señala que la seudonimización es solo una medida para reducir los riesgos de quiebra de la privacidad, sin excluir la adopción de otras. Ya en el articulado del texto, la adopción de medidas técnicas y organizativas adicionales a la seudonimización se contempla en el art. 25. Entre las medidas técnicas pueden incluirse, entre otras, las relativas al control de acceso, el cifrado de extremo a extremo, o la adopción de técnicas de auditorías que evalúen el nivel de riesgo para garantizar la privacidad del sistema en cuestión. Entre las medidas organizativas podría mencionarse la creación de comités de ética y protección de la privacidad con funciones de supervisión del uso de datos almacenados con fines de IA aplicada a la salud o el diseño de auditorías externas e internas que identifiquen posibles riesgos o quiebras de la privacidad. Véase al respecto la Guía para profesionales del sector sanitario de la Agencia Española de Protección de Datos, revisión de 2024, disponible en: [www.aepd.es](http://www.aepd.es) (última consulta: 1 de septiembre de 2025), pp. 14 ss.

<sup>60</sup> Entre otros, BLÁZQUEZ RUIZ, Francisco Javier, “Riesgos para la privacidad en la aplicación de la inteligencia artificial al ámbito biosanitario. implicaciones éticas y legales”, *Anales de la Cátedra Francisco Suárez*, Vol. 56, 2022, pp. 253 ss.

### III.- Los retos del diseño de las bases de la responsabilidad penal por un resultado lesivo en el ejercicio de la medicina asistida por IA

La irrupción de herramientas basadas en la IA plantea la necesidad de llevar a cabo una revisión de los parámetros tradicionalmente elaborados en relación con la imprudencia, permitiendo con ello sentar las bases conforme a la cuales puede derivarse responsabilidad por un resultado lesivo para los distintos agentes implicados en su uso. En concreto, es común distinguir tres grandes grupos de sujetos que intervienen desde el diseño hasta el funcionamiento del sistema. El primero corresponde a los desarrolladores y fabricantes, ocupados de su creación y diseño; el segundo, señala a quienes en calidad de profesionales están llamados a manejarlo; por último, el tercero identifica a los destinatarios finales del sistema<sup>61</sup>, grupo éste que en el específico ámbito de la medicina estaría referido a los pacientes.

De estos tres posibles centros de imputación, las consideraciones que siguen se centran básicamente en la eventual responsabilidad de los profesionales sanitarios encargados del manejo de aquellas herramientas. En relación con los diseñadores o creadores del sistema, existe una abundante literatura que residencia las bases de su tratamiento en los esquemas generales diseñados para el trazo de la responsabilidad por el producto, y que en buena medida recurre como criterio para acotar la permisibilidad del riesgo por su puesta en el mercado a parámetros basados en el *compliance*, diseñado en su origen, como es sabido, con la finalidad de trazar los presupuestos de la responsabilidad de las personas jurídicas<sup>62</sup>. Con todo, allí donde se registre un daño, esa misma literatura coincide también a la hora de dar cuenta de las dificultades para derivar responsabilidad penal por la producción de un resultado lesivo. Se enuncian entre ellas desde la complejidad para identificar la relación de causalidad entre el daño y la actuación eventualmente defectuosa, hasta la dificultad que encierra verificar los criterios de imputación -debido a la posible interacción de los profesionales usuarios del sistema, que en última instancia deben alimentarlo con datos cuyo manejo puede haber estado en la base del fallo-. Por su parte, en relación con los usuarios finales -pacientes- es cierto que también pueden asumir una cuota de responsabilidad por el mal uso del sistema inteligente, si bien en la constelación de casos que aquí interesa -receptores o destinatarios de una diagnóstico o tratamiento pautado con ayuda de la IA-, en general su papel no pasa de ser el de meros sujetos pasivos de la aplicación de la herramienta.

Centrando, por tanto, las consideraciones que siguen en la posible responsabilidad de los profesionales de la salud y con carácter previo a abordar los distintos escenarios en que ésta puede surgir, resulta conveniente dejar sentado una serie de presupuestos que tracen el perímetro de los deberes que el uso de tales herramientas suponen para el sanitario y, en definitiva, que sirvan de presupuesto para concretar los estándares de actuación cuya inobservancia pudiera dar lugar a responsabilidad, en su caso, en el orden penal. Básicamente, son cinco las cuestiones que conviene dejar sentadas como punto de partida.

La primera, más que un presupuesto de responsabilidad supone en realidad un límite a ella. Se enuncian líneas más arriba los posibles centros de imputación por los resultados lesivos que

---

<sup>61</sup> Por todos, VALLS PRIETO, Javier, "Sobre la responsabilidad penal por la utilización de sistemas inteligentes", *Revista Electrónica de Ciencia Penal y Criminología*, 2022, pp. 24 ss.

<sup>62</sup> Entre otros, ROMEO CASABONA, Carlos María/ LAZCOZ MORATINOS, Guillermo, "Inteligencia artificial aplicada a la salud: ¿qué marco jurídico?", *op. cit.* p. 160. También en ROMEO CASABONA, Carlos María, "La atribución de responsabilidad penal por los hechos cometidos por sistemas autónomos inteligentes, robótica y tecnologías conexas", *Revista de Direito SA*, vol. 16. 2022, p. 9, 22: "Habría que trasladar al régimen de los sistemas inteligentes procedimientos de validación de su inocuidad antes de introducirlos en el mercado o de ponerlos a disposición de la persona que los encargó (probable usuario final)".

Sigue la propuesta de ROMEO CASABONA, Carlos María, por ejemplo, RUEDA MARTÍN, María Ángeles, "La concreción del deber objetivo de cuidado en el desarrollo de la actividad médico-quirúrgica curativa a la luz de los avances de los sistemas de Inteligencia Artificial", *InDret*, 1, 2025, pp. 217 s. De forma crítica sobre las dificultades de trasladar los esquemas del cumplimiento normativo al ámbito de los sistemas inteligentes, véase VALLS PRIETO, Javier, "Sobre la responsabilidad penal por la utilización de sistemas inteligentes", *op. cit.*, pp. 30 s.

podieran producirse. Junto con sus manipuladores o responsables del despliegue<sup>63</sup> -en el caso que interesa, los profesionales sanitarios-, se señalaba básicamente la eventual responsabilidad de los desarrolladores o diseñadores y programadores del sistema, que pudieran responder por los fallos en su diseño y funcionamiento conforme a los esquemas propios de la responsabilidad por el producto. Se subrayaban también entonces las dificultades que a menudo surgen en la práctica para determinar la responsabilidad de estos agentes. Como sea, lo que interesa destacar ahora es que la aceptación de las limitaciones propias de la derivación de responsabilidad a tales sujetos no puede llevar a buscar necesariamente responsables alternativos para satisfacer la necesidad social de identificar un culpable por la producción de resultados que afectan a bienes jurídicos fundamentales. La irrupción de nuevas tecnologías conlleva, ante todo, la aceptación de determinados niveles de permisibilidad de la conducta; esto es, la necesidad de tolerar el riesgo de producción de determinados resultados tanto en un juicio *ex ante*, a modo de balanza apriorística entre beneficios y riesgos del sistema, como tras su materialización en un daño concreto, que entonces deberá valorarse tan admisible, por inevitable, como los daños causados a consecuencia de cualquier intervención propia de la medicina tradicional que respeta los parámetros de lo permitido<sup>64</sup>. Obviamente, estando en juego el riesgo de lesión de bienes jurídicos del más alto nivel, como es el caso de la vida y la integridad física, la autorización misma del sistema requiere que, conforme al estado de la técnica, su uso comporte una probabilidad prácticamente despreciable del daño. Por otro lado, y en lo que concierne a la actuación de sus usuarios, reclama que los profesionales de la salud hayan seguido los correspondientes protocolos de actuación, entre los que se cuentan, como habrá ocasión de desarrollar más adelante, tanto su instrucción y capacidad para el manejo de tales herramientas como el juicio crítico sobre sus resultados.

La segunda premisa remite a una cuestión más amplia, relacionada con los presupuestos de la exigencia de responsabilidad por un resultado lesivo en el ejercicio de la medicina tradicional. Se trata de la relativa a la relación que media entre la necesaria información que ha de recibir el paciente y la eventual incidencia de sus quiebras en la responsabilidad por un resultado lesivo. Hubo ya ocasión de recordar líneas más arriba la importancia del deber de informar por parte del médico como presupuesto de la validez del consentimiento del paciente que, precisamente por ello, para que surta efectos habrá de adjetivarse como *informado*. También se dio cuenta entonces de que el traslado de tal exigencia al ejercicio de la medicina asistida por sistemas de IA exige extender aquella información, ante todo, al uso mismo de dichas tecnologías, de modo tal que el paciente habrá de ser informado no sólo de su empleo, sino también de las ventajas esperables y de las limitaciones y riesgos eventualmente asociados. Lo que interesa destacar ahora es que la omisión de dicha información habrá de discurrir, en su caso, por cauces ajenos a los que se derivan de un mal recurso de la técnica, siempre claro está que el acto médico con finalidad curativa se haya realizado conforme a las normas de cuidado. Lejos de ser éste un aspecto específico del ámbito que interesa, se trata tan sólo de la traslación a él de la premisa que también rige en relación con el ejercicio de la medicina tradicional. Si bien reconociendo que esta cuestión no siempre ha recibido respuesta unánime en la doctrina, se parte en estas líneas de que el consentimiento del paciente no forma parte de la *lex artis* médica, por no tener incidencia directa en el resultado lesivo<sup>65</sup>. Resulta de lo anterior que el escenario en que procederá plantear la eventual responsabilidad del profesional por aquel en el marco de una actividad curativa asistida por IA se acota a los casos en los que la producción del daño se asocie al mal uso o a fallos no detectados del sistema, nunca a irregularidades relativas a la información del paciente, que habrán de ventilarse -en la vía penal o en su caso civil- bajo la óptica de la afectación de la libertad del enfermo, un interés por completo ajeno a la lesión del bien jurídico vida o salud que toman por base los respectivos delitos de homicidio o lesiones.

---

<sup>63</sup> En terminología del hay citado el documento de Directrices éticas para una IA confiable, punto 56, disponible en: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (última consulta: 8 de julio de 2025).

<sup>64</sup> Sobre sus límites en la práctica médica tradicional. GÓMEZ RIVERO, María del Carmen, *La responsabilidad penal del médico*, op. cit., pp. 264 ss.

<sup>65</sup> Sobre los argumentos para llegar a esta conclusión *Ibidem.*, pp. 269 ss.



Como tercera premisa básica, conviene marcar las bases de la relación apriorística entre las herramientas que se tratan y la actuación de los profesionales de la sanidad. Expresada sintéticamente la cuestión, importa determinar si los resultados que ofrecen tales sistemas gozan de una suerte de presunción de corrección, con la consiguiente capacidad para reconocer anclajes sobre los que fundamentar la confianza del sanitario o si, por el contrario, la relación debe trazarse en términos de desconfianza, lo que sucedería si aquellos resultados no gozaran del beneficio de su fiabilidad *a priori*. La importancia de la premisa que se adopte se comprende sin dificultad debido a la repercusión inmediata que tiene en relación con una cuestión más amplia; a saber, la relativa a si el empleo de los sistemas de IA en medicina se traduce en la práctica en una elevación de las cotas de exigencias de cuidado, explicable por la diligencia adicional que requiere el manejo del sistema, o si, por el contrario, permite cierta relajación. Es esto último lo que sucedería si el recurso a tales herramientas hubiera de valorarse como una sólida guía de actuación para el sanitario, con la consecuencia entonces de que la confianza en el sistema le eximiera de realizar actuaciones adicionales.

Sobre este aspecto habrá ocasión también de volver al desarrollar los criterios conforme a los que derivar, en su caso, responsabilidad para los profesionales de la sanidad. Con la única pretensión de sentar de momento sus bases puede decirse que, si bien por razones obvias, la razón de ser del desarrollo de dichos sistemas es ofrecer un asidero fiable a la intervención humana, necesariamente ha de ser ésta el centro último de decisión. Quiere esto decir que, al menos en la mayoría de los casos<sup>66</sup>, los resultados que arrojen los sistemas de IA no debieran pasar de verse como pautas o propuestas de actuación sobre los que es posible fundamentar la apriorística presunción de su corrección. Lejos, sin embargo, del valor del conocido como principio de confianza que se desarrolló para las actividades protagonizadas por distintos participantes y conforme al cual es posible compartimentalizar tareas y, con ello, responsabilidades, el alcance de esa presunción de corrección de los resultados arrojados por el sistema es bastante más débil. Y todo ello porque no permite sectorializar funciones ni resultados. La máquina no es un interviniente más con autonomía equiparable a la del humano, que permitiera encontrar cierta similitud con la división -vertical- de funciones. La máquina es, tan solo, una herramienta al servicio del único centro decisorio final, el hombre<sup>67</sup>.

Precisamente estas consideraciones dan paso a una cuarta premisa. Es la que recuerda la vigencia en este ámbito como punto de partida de la clásica elaboración de la imprudencia, aplicable en relación con el ejercicio de la medicina tradicional a quien actúa con capacidad decisoria. No se pretende desde luego reproducir en estas líneas la consolidada elaboración de los criterios conforme a los cuales derivar responsabilidad por imprudencia médica en el orden penal, pero sí de recordar algunos aspectos básicos. Entre ellos, el reconocimiento generalizado de que el contenido de injusto del delito imprudente pivota sobre la *infracción del deber de cuidado*, y que sólo a partir de la ponderación de todas las circunstancias del caso concreto procede realizar un juicio en torno a la probabilidad objetiva de la lesión del bien jurídico como presupuesto del deber objetivo de cuidado, probabilidad que, como también suele admitirse sin dificultades, habrá de limitarse por la *cognoscibilidad* del autor<sup>68</sup>.

A la hora de fijar los parámetros que sirven de base a la comprobación del cumplimiento del deber objetivo de cuidado suele trazarse como criterio la atención a las reglas técnicas -allí donde se trate de una actividad reglada-, y a las que marca la experiencia -en otro caso-. En relación con lo primero, las normas positivizadas que guíen la actuación del sanitario a modo de protocolos, hay que reconocer que, si bien sus bases generales están sentadas en distintos

<sup>66</sup> Se exceptuarían aquellos en los que el dispositivo fuese programado para arrojar un resultado exacto como, por ejemplo, las dosis de insulina que debe recibir un paciente diabético en función de los datos que previamente se han introducido en el sistema.

<sup>67</sup> No obstante, no ha faltado algún defensor de buscar en el principio de confianza un parámetro para sentar las bases de la responsabilidad del sanitario en el marco del uso de sistemas de IA por la producción de resultados adversos motivados por las indicaciones del sistema. En la doctrina chilena es el caso de PERÍN, Andrea, quien propuso como paradigma las desviaciones respecto a las guías de actuación -*guidelines*-, y de modo singular la atención a los esquemas del trabajo en equipo y el principio de confianza por el que se rige, "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*, vol. 8, núm. 1, 2019, pp. 15 ss.

<sup>68</sup> GÓMEZ RIVERO, María del Carmen, *La responsabilidad penal del médico*, op. cit., pp. 337 ss.



documentos, empezando por la tantas veces referida ley de IA, la imposibilidad de pormenorizar las reglas de validez de sus resultados de modo específico en relación con el uso de cada herramienta determina que adquiera un protagonismo de primer orden la comprobación de los parámetros que definen la experiencia acumulada por el desarrollo de la actividad profesional. El traslado de este criterio a la asistencia sanitaria guiada por IA permite referenciarla a una doble realidad, y, con ello, manejar el término en un doble sentido. El primero toma como referente la experiencia que deriva del propio manejo de estos dispositivos. Ni que decir tiene que esta será más rica cuanto más implementada esté la herramienta en cuestión y, con ello, mayor sea la experiencia acumulada de su uso. El segundo se refiere a la que deriva del ejercicio de la medicina tradicional, cuyo acervo se erige en primer referente de comparación en orden a la validación de los resultados que arroje la IA. Este segundo parámetro se revela tanto o más decisivo que el primero desde el momento en que, como se ha venido subrayando de modo reiterado, no es función ni propósito de tales sistemas –al menos en su desarrollo actual- sustituir la labor del médico, sino quedar a su servicio para la toma de decisiones o la ejecución de las tareas que le siguen siendo propias. Se formula así un importante enunciado de cara a la delimitación y limitación de su eventual responsabilidad: la presunción de validez o certeza de los resultados arrojados por el sistema será tanto menor cuanto más se aparten de los que aconsejarían las reglas de la experiencia enunciadas en el ámbito de la medicina tradicional, definitorias de la actuación propia de un médico prudente. Resulta a su vez de lo anterior que la valoración de la decisión médica como conforme a las reglas de cuidado requerirá comprobar que *las pautas que arroja el sistema resulten razonables de acuerdo con los criterios o reglas de la experiencia*. Sobre todo ello también habrá ocasión de volver a tratar al valorar la aplicación de estos presupuestos a los escenarios concretos que pueden plantearse en la práctica.

La atención a este aspecto sirve a su vez para cerrar con una quinta premisa las coordenadas de la responsabilidad médica asistida por IA. Habrá de considerarse en todo caso imprudente la aceptación de los resultados del sistema sin que medie la supervisión humana significativa contemplada en el art. 22 del Reglamento IA, que expresamente veta la toma de decisiones basadas únicamente en el tratamiento automatizado de datos allí donde produzca los efectos que describe<sup>69</sup>. Conviene al respecto insistir en que por intervención humana significativa no sólo hay que entender la continua revisión y supervisión del correcto funcionamiento del sistema, sino sobre todo y ante todo, la revisión y supervisión de los resultados que aquel arroje, garantizando de ese modo que su valor no pase de ser lo que se ha dado en llamar *Sistema de Apoyo a las Decisiones (Decision Support Systems o DSS)*<sup>70</sup>, esto es, una herramienta de asistencia a la labor del sanitario<sup>71</sup>, nunca un reemplazo del mismo<sup>72</sup>. Cuestión distinta es la relativa al modo en que se entienda cumplida la exigencia de aquella intervención, concreción que, si bien omite el Reglamento, ha sido abordada por algún documento de carácter interpretativo en el sentido de requerir una intervención real y no meramente simbólica<sup>73</sup>.

---

<sup>69</sup> Conforme al apartado primero de dicho precepto, “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”. Ampliamente al respecto, así como sobre las exigencias que ha de tener la intervención humana y lo que se entienda por supervisión humana significativa, véase LAZCOZ MORATINOS, Guillermo, “Sistemas de inteligencia artificial en la asistencia sanitaria: cómo garantizar la supervisión humana desde la normativa de protección de datos”, *op. cit.*, pp. 10 ss.

<sup>70</sup> Destaca este aspecto, ROMEO CASABONA, Carlos María/LAZCOZ MORATINOS, Guillermo, “Inteligencia artificial aplicada a la salud: ¿qué marco jurídico?”, *op. cit.*, pp. 152 ss.

<sup>71</sup> Declaración del Comité Ejecutivo de la Asociación Médica alemana, adoptada en su reunión de 14 de enero de 2025, “Stellungnahme kunstliche Intelligenz in der Medizin”, *Deutsches Ärzteblatt* 4/2025, pp. 7 ss., 9. Disponible en: <https://www.aerzteblatt.de/archiv/stellungnahme-kuenstliche-intelligenz-in-der-medizin-1581a074-7da3-4724-875e-09f2be35a9a3> (última consulta: 23 de julio de 2025)

<sup>72</sup> Como se ha dicho, se trata de alcanzar los que pudieran llamarse “médicos mejorados por la IA”, no “médicos sustituidos por la IA”, GÓMEZ GONZÁLEZ, Emilio/ GÓMEZ Emilia/ MARQUEZ-RIVAS, Javier, et al. *Inteligencia Artificial en Medicina y Salud: revisión y clasificación de las aplicaciones actuales y del futuro cercano y su impacto ético y social*, *op. cit.*, p. 9.

<sup>73</sup> Véase al respecto el Grupo de Trabajo sobre Protección de Datos, artículo 29. Se trata de órgano consultivo independiente de la UE en materia de protección de datos e intimidad. En su documento sobre *Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679*, adoptadas el 3 de octubre de 2017: “Para ser considerada como participación humana, el responsable del tratamiento debe garantizar que cualquier supervisión de la decisión sea significativa, en vez de ser únicamente un gesto simbólico. Debe llevarse a cabo por parte de una persona autorizada y competente para modificar la decisión. Como parte del análisis, debe tener

La referida exigencia resulta crucial allí donde la herramienta no ofrezca un único diagnóstico posible o tratamiento, sino que plantee varias opciones o alternativas, de modo que necesariamente el profesional habrá de valorar cada una de ellas a la luz de las peculiaridades del caso en cuestión. Pero incluso si el sistema arroja un resultado único, su consideración como una herramienta en manos del actuar médico impediría igualmente su validación ciega o automática, reclamando en su lugar el acuñado “control humano significativo”<sup>74</sup>. En definitiva, en cualquiera de los casos, la irrenunciabilidad de la intervención humana en el marco de la medicina asistida por IA lleva inexorablemente a centrar la cuestión en dilucidar la eventual responsabilidad del médico que ha actuado secundando las indicaciones de la herramienta.

A partir de las premisas anteriores estamos en condiciones de exponer algunas propuestas de solución para los principales escenarios problemáticos que pueden presentarse en la práctica. Conviene advertir que se ciñen estas consideraciones a los supuestos de posibles resultados lesivos motivados estrictamente por la decisión médica de seguimiento o no de los resultados del sistema. Al margen de ellas quedan los casos en que el origen del daño se anude al desconocimiento o falta de instrucción acerca del mecanismo de su funcionamiento, aspecto que daría origen a la correspondiente responsabilidad por imprudencia debida al emprendimiento mismo de la actividad. Así habría de verse en tanto que presupuesto de su manejo es el adecuado conocimiento por el profesional sanitario de su funcionalidad y de sus posibilidades y limitaciones<sup>75</sup>. Por lo mismo, al margen de estas consideraciones quedan igualmente los supuestos inversos en que se plantee el papel que debiera atribuirse a los conocimientos superiores del profesional sanitario respecto al rendimiento del sistema, saberes que sin embargo no emplea. Su tratamiento habrá de discurrir conforme a los esquemas generales trazados con carácter general por la dogmática penal para determinar cuándo deben actualizarse tales conocimientos, a los que procede en este punto remitirse<sup>76</sup>.

Así acotado el objeto de este apartado, la peculiaridad de las distintas fases en que puede plantearse el error obliga a diferenciar, al menos, tres grupos de casos correspondientes, respectivamente, a la fase de diagnóstico, de toma de decisiones a partir del diagnóstico previamente aceptado y, por último, a la de ejecución del tratamiento en cuestión. Se tratan por separado cada uno de ellos.

### 1.- La responsabilidad por los daños derivados de una actuación asistida por IA en la fase de diagnóstico

Conforme se ha anticipado, la posible responsabilidad por un error de diagnóstico, positivo o negativo en una actuación médica guiada por IA puede surgir en dos posibles escenarios. El primero comprende los casos en que el profesional de la salud se aparte del resultado que arroje aquella herramienta, considerándose aquél *a posteriori* correcto. El segundo corresponde a los supuestos inversos, esto es, aquellos en los que el profesional adopta sin reservas el resultado del sistema, que se revela finalmente dañino para el paciente.

Tanto en uno como en otro caso interesa no perder de vista un aspecto ampliamente consolidado en el enjuiciamiento de errores en el marco de la medicina tradicional; a saber, el hecho de que la fase de diagnóstico se considera la más compleja de cuantas integran la intervención médica, debido a la multitud de factores que deben tomarse en consideración para obtener una diagnosis precisa, así como al hecho de que sean racionalmente admisibles variadas interpretaciones

---

en cuenta todos los datos pertinentes”, p. 23, disponible en <https://ec.europa.eu/newsroom/article29/items/612053/en> (última consulta: 3 de septiembre de 2025).

<sup>74</sup> BECK, Susanne / FABER, Michelle / GERNDT, Simon, “Rechtliche Aspekte des Einsatzes von KI und Robotik in Medizin und Pflege”, *Ethik in der Medizin*, núm. 35, 2023, pp. 255 ss.

<sup>75</sup> Señala en este sentido Declaración del Comité Ejecutivo de la Asociación Médica alemana, adoptada en su reunión de 14 de enero de 2025, que los sanitarios “tienen la obligación de garantizar que el dispositivo médico funciones y esté en buen estado, así como de seguir las instrucciones de mantenimiento del fabricante, por ejemplo, instalando las autorizaciones pertinentes”, “Stellungnahme künstliche Intelligenz in der Medizin”, *op. cit.*, p. 9, 11.

<sup>76</sup> De todo ello tuve ocasión de ocuparme en *La responsabilidad penal del médico*, *op. cit.*, pp. 381 ss.

válidas para unos mismos síntomas. No es por ello de extrañar que en relación con el ejercicio de la medicina tradicional la jurisprudencia se muestre proclive a una valoración más benévola de la imprudencia por errores de diagnóstico, de tal modo que sólo en los casos de imprudencia extraordinariamente burda haya apreciado las cotas más altas de negligencia y, específicamente, la imprudencia profesional<sup>77</sup>.

El ejercicio de la medicina asistida por IA no debiera llevar a baremar la negligencia en fase de diagnóstico desde una óptica distinta teniendo en cuenta el aspecto tantas veces repetido en torno a que aquellos sistemas son tan solo una herramienta en manos del médico, que en absoluto sustituyen su labor personal de análisis de la etiología de los síntomas. Cuestión distinta es que esta afirmación debe matizarse atendiendo a la mayor o menor fiabilidad de las aplicaciones asistenciales en fase de diagnóstico. A título de ejemplo, entre las que se presentan con un valor meramente orientativo del criterio médico puede citarse Babylon Health, una aplicación basada en el tratamiento de un ingente volumen de datos para facilitar el diagnóstico, cuyos resultados no pasan de ser meros criterios orientadores del juicio médico humano<sup>78</sup>. Muy superior es la fiabilidad que ofrece el diagnóstico por imágenes<sup>79</sup>, ámbito en el que aquellas herramientas aseguran en general una precisión y unas capacidades superiores a las del juicio humano<sup>80</sup> y, desde luego, a determinadas pruebas que en la medicina tradicional no pasan de tener un valor indiciario<sup>81</sup>.

El diferente grado de precisión de cada uno de estos sistemas impide una valoración unitaria de la confianza que el profesional puede depositar en ellas. Es claro que en el primer caso, el de herramientas menos precisas, será imprescindible la validación de sus resultados a la luz de los criterios propios de la medicina convencional, cobrando plena vigencia el enunciado propuesto líneas más arriba, en el sentido de que el seguimiento de las pautas que arroja habrá de valorarse como conforme a las reglas de cuidado allí donde sus resultados se consideren *razonables de acuerdo con los criterios o reglas de la experiencia*. Por su parte, allí donde se trate de sistemas con alto grado de precisión habrá de tenerse por prudente la aceptación de sus resultados, con la única salvedad de que estos colisionaran abiertamente con los que ofrecen las reglas de la experiencia. Sirve de nuevo como paradigma el diagnóstico basado en la imagen. En el caso de que revelase un diagnóstico positivo, la decisión del médico de ignorarlo y atender en exclusiva al juicio clásico de detección de la enfermedad habría de valorarse a todas luces como contraria al estándar de cuidado, y con ello, podría fundamentar una posible responsabilidad en comisión por omisión. Estaría, en definitiva, omitiendo la actuación que demanda un diagnóstico con alta presunción de su fiabilidad. Distinta habría de ser la valoración si, atendiendo precisamente a las mayores posibilidades de detección de la enfermedad que ofrecen esos sistemas, el profesional descartase la patología atendiendo a sus resultados. Por las razones tantas veces reiteradas, no se trata, desde luego, de reconocer a la herramienta un valor absoluto de fiabilidad, ni tampoco

---

<sup>77</sup> GÓMEZ RIVERO, María del Carmen, *La responsabilidad penal del médico*, op. cit., p. 379.

<sup>78</sup> Sobre el sistema, véase por ejemplo, PELCYGER, Ben, *Artificial Intelligence in healthcare: Babylon Health & IBM Watson take the lead*, Prescouter, 2017, disponible en: <https://www.prescouter.com/2017/12/artificial-intelligence-healthcare/> (última consulta: 12 de junio de 2025).

<sup>79</sup> Entre otros ámbitos, el de la dermatología para la detección del melanoma. Sobre la capacidad de estos sistemas véase por ejemplo CODELLA, Noel/ CAI, Junjie/ ABEDINI, Mani/ GARNAVI, Rahil/ HALPERN, Alan, SMITH, John., "Deep Learning, Sparse Coding, and SVM for Melanoma Recognition in Dermoscopy Images", en: Zhou, Luping., Wang, Li, Wang, Qian, Shi, Yinghuan. (eds.) *Machine Learning in Medical Imaging*, 6 th International Workshop, MLMI 2015, Held in Conjunction with MICCAI 2015, Munich, Germany, October, 2015, Proceedings, 2015, pp. 118 ss., donde se destaca el alto rendimiento del sistema.

<sup>80</sup> Sirva de ejemplo un algoritmo desarrollado por Google Health, capaz de detectar el cáncer de pulmón con una precisión superior a la de los radiólogos, noticia disponible en: <https://www.redaccionmedica.com/secciones/oncologia-medica/un-algoritmo-de-google-detecta-el-cancer-de-pulmon-mas-rapido-que-un-medico-6366> (última consulta: 20 de agosto de 2025). Entre otras aplicaciones con vocación de superar en precisión al criterio humano puede verse el sistema desarrollado por IBM para una identificación precisa del cáncer de mama, puesto a prueba en una muestra de 71 casos en que los radiólogos descartaron la enfermedad, siendo así que la aplicación identificó un 48% de falsos negativos en dicha muestra. Al respecto, ÁVILA-TOMÁS, José Francisco/ MAYER-PUJADAS, Miguel Angel/ QUESADA VARELA Víctor Julio, "La inteligencia artificial y sus aplicaciones en medicina II: Importancia actual y aplicaciones prácticas", op. cit., p. 83.

<sup>81</sup> Bastaría recordar, entre otros tantos ejemplos que pudieran venir a colación, que los marcadores de sangre son un indicio, pero no prueba definitiva del padecimiento de un tumor; que los niveles de hormonas tiroideas pueden variar por diferentes causas sin que ello signifique una patología del tiroides, o que las pruebas de detección de anticuerpos pueden ser indicadores de una infección, pero nada más.

de ignorar los posibles errores que puede arrojar en fase de diagnóstico<sup>82</sup>, pero sí de presumir la validez de sus indicaciones de acuerdo con las mayores posibilidades de certeza que ofrece. La única salvedad habría de verse, una vez más, allí donde aquellos resultados colisionaran abiertamente con la actuación que demandarían las reglas de la experiencia. Se incluyen aquí tanto las propias del ejercicio tradicional de la medicina como la acumulada por el propio uso de los sistemas. En este segundo sentido, la experiencia que vaya generando su progresiva implementación será esencial para determinar la eventual responsabilidad del sanitario por el seguimiento o abandono de sus patrones.

## 2.- Determinación de la responsabilidad por el uso de la IA en fase de toma de decisiones sobre el tratamiento más adecuado para un determinado diagnóstico

Singular complejidad presenta la resolución de los casos en los que se ventila la corrección o no de la actuación médica cuando ya ha sido diagnosticada la enfermedad. Se trata, en definitiva, de decidir el valor que cobren las directrices proporcionadas por la IA; esto es, el peso que deba adquirir en la decisión final para considerar la actuación médica conforme al cuidado debido. Debe ante todo reconocerse que, al menos en el estado actual del desarrollo de estos sistemas, la cuestión no debe plantearse en términos de cuál deba ser el juicio prevalente en el caso concreto, esto es, el criterio humano o el artificial. La rotundidad del debate así planteado sólo cobraría sentido en los ámbitos en que los sistemas artificiales diseñados para pautar un tratamiento tuvieran reconocida una alta fiabilidad, lo que desde luego no es hasta la fecha el caso de los disponibles para todos los tipos de patologías<sup>83</sup>. Es más, incluso allí donde el grado de fiabilidad o precisión de la herramienta se considere elevado, también la comunidad científica reconoce la improcedencia de trasladar automáticamente sus resultados a la decisión final<sup>84</sup>, entre otras razones, debido a la ya referida opacidad de estos sistemas y a la consiguiente

<sup>82</sup> Sobre ellos advierte el ya citado Informe del Parlamento Europeo “Artificial intelligence in healthcare Applications, risks, and ethical and societal impacts”, 2022, ya sea en la forma de falsos negativos o falsos positivos, con la consecuencia entonces de tratamientos innecesarios, pp. 31 ss.

<sup>83</sup> Interesante es al respecto la referencia al conocido como programa Watson for Oncology (WFO), un sistema de toma de decisiones con IA desarrollado por IBM con la colaboración de oncólogos del Memorial Sloan Ketterin Cancer Center, y que comenzó a usarse en el año 2015. Se nutre de una ingente cantidad de datos de pacientes oncológicos a partir de los cuales arroja el diseño de pautas de tratamiento personalizadas; específicamente, el tipo de quimioterapia más indicada al caso. Los estudios comparativos entre esos resultados y los propuestos por equipos multidisciplinares arrojan resultados muy distintos dependiendo del tipo de cáncer. Así, por ejemplo, si bien el índice de coincidencia es alto en tipos como el de mama que alcanza más de un 90%, la coincidencia no supera el 54% en otros tipos de cáncer, como es el caso del gástrico. Por ello, en la comunidad científica se ha llegado a definir el sistema como una “herramienta”, no como una muleta. Así lo hace HAMILTON, Jada, et al. “A Tool, Not a Crutch”: Patient Perspectives About IBM Watson for Oncology Trained by Memorial Sloan Kettering, *Journal of Oncology Practice* Volume 15, Number 4, 2019, disponible en: <https://ascopubs.org/doi/10.1200/JOP.18.00417> (última consulta: 28 de julio de 2025). Sobre la coincidencia de resultados entre esta herramienta y las recomendaciones de los equipos multidisciplinares según los distintos tipos de cáncer, véase JIE Zou / ZHIYING, Zeng / LI, Li, “A meta-analysis of Watson for Oncology in clinical application”, *Scientific Report*, núm. 11, 2021, disponible en: <https://doi.org/10.1038/s41598-021-84973-5> (última consulta: 20 de julio de 2025). El estudio ofrece una relación de ventajas y desventajas del uso de la herramienta. Si bien entre las primeras se cuenta que puede prevenir errores de cálculo humano, destacando entre sus inconvenientes las limitaciones actuales del sistema debido a una serie de factores que explicarían la falta de consistencia de criterios en determinados tipos de cáncer. Véase también TIAN Yulong / LIU Xiaodong / WANG Zixuan / CAO Shougen / LIU Zimin / JI Qinglian / LI Zequn /, SUN Yuqi / ZHOU Xin / WANG Daosheng / ZHOU Yanbing, “Concordance Between Watson for Oncology and a Multidisciplinary Clinical Decision-Making Team for Gastric Cancer and the Prognostic Implications: Retrospective Study”, *Journal of Medical Internet Research*, 2020, disponible en: <https://pubmed.ncbi.nlm.nih.gov/32130123/> (última consulta: 12 de agosto de 2025), donde destacan el rendimiento del sistema en el caso del cáncer de mama y pulmón, no así en el gástrico.

<sup>84</sup> En relación con el programa WFO ya referido y respecto al alto rendimiento que ofrece el sistema en relación con el cáncer de mama, también se recomienda su uso como mero complemento del juicio médico. Véase MELLADO ORELLANA, Raúl / ESCOBAR ARRIAGA, Elizabeth / DE LA MORA MOLINA, Héctor / DÍAZ GREENE, Enrique Juan / RODRÍGUEZ WEBER, “Estudio de concordancia entre el sistema Watson for Oncology y la práctica clínica en pacientes con cáncer de mama dentro del Hospital Ángeles Pedregal”, *op. cit.*, quienes concluyen que “el sistema WFO fortalece la toma de decisiones por médicos oncólogos, pero no reemplaza la capacidad o decisión humana”, p. 341. Otro tanto puede decirse de otros sistemas, como Microsoft’s InnerEye, en relación con el tratamiento de tumores que, pese a su alta fiabilidad, reclama también la intervención y supervisión humana. Puede consultarse al respecto la información publicada en <https://news.microsoft.com/es-xl/una-herramienta-de-ia-de-microsoft-ayuda-a-acelerar-el-tratamiento-contra-el-cancer-y-addenbrookes-sera-el-primer-hospital-del-mundo-en-usarla/>



dificultad de validar automáticamente sus resultados<sup>85</sup>. Este factor impide ya de por sí considerar acorde con los estándares de prudencia su aceptación como infalibles o definitivos, de tal modo que su valor no pasará de ser un indicador o apoyo adicional a la decisión del médico<sup>86</sup>.

Por otra parte, como se ha repetido ya líneas más arriba, una de las principales características de los sistemas comunes de IA aplicados a la medicina es que su toma de decisiones se basa usualmente en los resultados inferidos a partir del análisis de una ingente cantidad de datos, no en un análisis estrictamente individual del paciente atento a sus circunstancias singulares. El modelo descansa, en esencia, en la toma en consideración de promedios sobre determinados grupos de individuos; esto es, en el análisis de los patrones que generalmente se asocian a ese grupo de población. Este método, que puede ofrecer resultados altamente fiables en comportamientos de datos objetivables, como los económicos, encuentra como principal objeción en el ámbito de la medicina la necesidad de tener en cuenta las características singulares del enfermo, lo que, entre otros extremos, como los ambientales o la interacción de condiciones personales, pasa por atender a sus peculiaridades genéticas. De hecho, la toma en consideración de estos aspectos es uno de los pilares de la tan pregonada en la actualidad *medicina personalizada*, que lejos de guiarse conforme a modelos estandarizados, se basa en el diseño de patrones de actuación individualizados conforme a aquellos factores personales tanto en relación con el diagnóstico, pronóstico como tratamiento de una enfermedad<sup>87</sup>, siendo su máxima aspiración la creación de los llamados *gemelos digitales*<sup>88</sup>.

El reconocimiento de todas estas peculiaridades requiere adaptar al específico ámbito de la toma de decisiones médicas asistidas por IA la clásica construcción de la imprudencia desarrollada para el ejercicio de la actividad sanitaria tradicional, lo que supone, ante todo, la búsqueda de parámetros conforme a los cuales determinar a la luz de las normas positivizadas y de experiencia si la elección del tratamiento guiada por tales sistemas infringió el deber objetivo y subjetivo de cuidado.

Hay que reconocer que, en realidad, el modo en que se lleve a cabo tal adaptación viene en gran medida condicionado por una cuestión más amplia que previamente debiera resolverse y que ya fue anticipada líneas más arriba. Es la relativa a si la introducción de tales herramientas supone incrementar los estándares del deber de cuidado objetivo y subjetivo o si, por el contrario, permite cierta relajación de ellos. Sucedería esto último si los resultados que arroja la herramienta gozaran de una presunción de fiabilidad, de modo tal que bastase para el sanitario respetarlos, con la consiguiente simplificación de otros deberes de cuidado. Por las razones tantas veces reiteradas, el desarrollo actual de dichos sistemas en lo que se refiere a la toma de decisiones sobre la elección de un tratamiento no permite tal conclusión. Al contrario, habría que decir, su irrupción en la práctica médica supone una elevación de los deberes de cuidado exigibles en un doble sentido. En primer lugar, porque, al igual que sucede en relación con las herramientas de diagnóstico, ya de entrada añade nuevos deberes relativos a las condiciones que permitan su empleo, en tanto que a los saberes y competencias en general requeridos para el ejercicio de la actividad se suman los que reclama el manejo del sistema, inclusivos no sólo de su funcionamiento, sino también de sus propias limitaciones<sup>89</sup>. En segundo lugar, la elevación de

---

<sup>85</sup> SCHWELKART, Scott, *Who will be liable for Medical Malpractice in the Future? How the Use of Artificial Intelligence in Medicine Will Shape Medical Tort Law*, *Minnesota Journal of Law, Science & Technology*, núm. 22, 2021, pp. 3 ss., disponible en: <https://scholarship.law.umn.edu/mjlst/vol22/iss2/2/> (última consulta: 2 de junio de 2025). En nuestra doctrina puede verse, por ejemplo, TORRALBO MUÑOZ Bartolomé /MORALES MORENO, África María, "Medicina, inteligencia artificial y Derecho penal: una relación compleja", en *Foderics 10.0. Estudios sobre Derecho Digital* (coords., Bueno de la Mata Federico, González Pulido, I.), Ed. Comares, Granada, 2022, p. 664.

<sup>86</sup> Así lo recuerda el ya citado Informe del Parlamento Europeo "Artificial intelligence in healthcare Applications, risks, and ethical and societal impacts", 2022, p. 22.

<sup>87</sup> Ampliamente sobre su concepto, presupuestos y retos, ROMEO CASABONA, Carlos María, /NICOLÁS JIMÉNEZ, Pilar, *Retos éticos y necesidades normativas en la actividad asistencial en medicina personalizada*, op. cit., pp. 21 ss.

<sup>88</sup> Se entiende por tal la representación virtual del paciente, que al modo de un avatar supone una réplica del mismo continente de todos sus datos, patrones y características, lo que permite hacer una simulación de la evolución de su enfermedad en tiempo real

Al respecto REMENTERÍA, María José / PROL CASTELO, Guillermo/ URDA Beatriz/ NÚÑEZ CARPINTERO, Iker / CIRILLO, Davide/ VALENCIA, Alfonso, "Inteligencia artificial y medicina personalizada" (ed. lit. AMO USANOS, Rafael), *Inteligencia artificial y Bioética*, Universidad Pontificia de Comillas, Madrid, 2023, pp. 49 ss.

<sup>89</sup> A ello se refiere el art. 14 del Reglamento IA, que establece una serie de obligaciones para las personas a las que se encomienda su vigilancia. En concreto, deben "entender adecuadamente las capacidades y limitaciones pertinentes del



los niveles de cuidado se comprende a la vista de la tantas veces repetida imposibilidad de fundamentar una confianza ciega en los resultados del sistema, de tal modo que éstos, en lugar de simplificar la labor humana, se convierten en un referente adicional dentro del catálogo de los que ha de tener en cuenta el profesional a la hora de emitir su juicio y pautar el consiguiente tratamiento.

De acuerdo con lo anterior, habrá de apreciarse un comportamiento imprudente, por ejemplo, en el caso de que el sanitario llegase a aceptar las recomendaciones que arroja el sistema sin tener en cuenta su corrección a la vista de las peculiaridades del cuadro clínico del paciente, que sin embargo pudiera aconsejar conforme a los conocimientos médicos tradicionales la adopción de una terapia distinta. Otro tanto habría que decir, tomando ahora como referente los conocimientos específicos del profesional, allí donde su experiencia aportase elementos de juicio adicionales para apartarse de las indicaciones del sistema o, por el contrario, para secundarlas. En definitiva, cobra de nuevo plena vigencia la pauta señalada líneas más arriba en torno a que sólo podrá valorarse como prudente el seguimiento de los resultados de los sistemas de IA allí donde resulten *razonables conforme a las reglas de la experiencia*, erigidas en inevitable referente de cotejo para presumir su acierto.

Las consideraciones anteriores han tratado de pergeñar los presupuestos conforme a los cuales apreciar una actuación contraria a cuidado, esto es, el desvalor de acción de la imprudencia. Ni que decir tiene que la exigencia de responsabilidad penal, en cualquiera de los ámbitos en que se plantee, reclama adicionalmente la comprobación de la concurrencia de una determinada relación entre el resultado producido y aquella, de modo que pueda decirse que el primero sea imputable en términos normativos a la infracción del cuidado esperado. La plasmación de los criterios desarrollados por la dogmática penal para condensar aquella exigencia encuentra plena aplicación en este ámbito. Ciertamente es que por las razones tantas veces repetidas, la cuestión no se planteará en relación con la puesta en funcionamiento del sistema, en tanto que su autorización administrativa supone ya la aceptación, no sólo de que cumple con los estándares del riesgo permitido, sino de que no supone un incremento relevante del riesgo de producción del resultado respecto al que sería propio de la actuación realizada conforme a los cánones de la medicina tradicional. La cuestión que usualmente se planteará en la práctica será si la actuación del sanitario que de modo negligente confía o se aparta de los resultados del sistema en cualquiera de los escenarios referidos cumple con aquellos otros criterios de imputación; específicamente con el que reclama que la conducta suponga un incremento relevante del riesgo de producción del resultado. Es cierto que su constatación no resulta dudosa allí donde la falta de diagnóstico de la enfermedad en cuestión haya impedido un tratamiento con posibilidades de curación o, al menos, de mejoría. Las dudas surgirán inevitablemente cuando el estado avanzado de la enfermedad haga cuestionable la utilidad del tratamiento que se debiera haber sido suministrado, esto es, cuando en todo caso la conducta alternativa ajustada al deber de cuidado hubiera tenido escasas posibilidades de éxito. La solución de estos casos no habrá de ser distinta a la que inspira en general el tratamiento de los que se plantean en el ámbito de la medicina tradicional, de modo tal que la aceptación en el caso concreto de que la decisión adoptada no influye en el resultado final supondría la exclusión de aquella relación de imputación.

---

sistema de IA de alto riesgo y poder vigilar debidamente su funcionamiento, por ejemplo, con listas a detectar y resolver anomalías, problemas de funcionamiento y comportamientos inesperados; b) ser conscientes de la posible tendencia a confiar automáticamente o en exceso en los resultados de salida generados por un sistema de IA de alto riesgo («sesgo de automatización»), en particular con aquellos sistemas que se utilizan para aportar información o recomendaciones con el fin de que personas físicas adopten una decisión; interpretar correctamente los resultados de salida del sistema de IA de alto riesgo, teniendo en cuenta, por ejemplo, los métodos y herramientas de interpretación disponibles; decidir, en cualquier situación concreta, no utilizar el sistema de IA de alto riesgo o descartar, invalidar o revertir los resultados de salida que este genere; intervenir en el funcionamiento del sistema de IA de alto riesgo o interrumpir el sistema pulsando un botón de parada o mediante un procedimiento similar que permita que el sistema se detenga de forma segura.

### 3.- Bases para la apreciación de una eventual responsabilidad penal en la fase de ejecución del tratamiento: el específico problema de los robots sanitarios

Se tratan en lo que sigue los casos en los que el daño sufrido por el paciente tenga su origen, no ya en el proceso decisional del sistema, sino en la fase ulterior de ejecución de la decisión previamente tomada. Son, en definitiva, los supuestos en los que la actuación médica sea llevada a cabo a partir de herramientas inteligentes, esto es, robots, programados para ejecutar un determinado acto médico. Entre el vasto catálogo de aplicaciones que puede encontrar en medicina<sup>90</sup>, el paradigma se encuentra en el empleo de los llamados robots quirúrgicos, genuinos asistentes de primer orden del cirujano cuyas ventajas son ampliamente reconocidas<sup>91</sup>. Allí donde su empleo produzca un resultado lesivo se plantea tanto el trazo de los presupuestos de la exigencia de responsabilidad como la identificación del agente o agentes que debieran responder por él.

Conocido es el debate que ocupa actualmente a la doctrina penal en torno a la posibilidad de elevar a sujeto responsable a aquellas máquinas inteligentes o robots, o incluso en torno a la conveniencia de crear un *Derecho de robots* y, específicamente, un *Derecho penal de robots*<sup>92</sup>. No necesariamente acotado el problema al ámbito médico sino de modo más amplio a cualquiera en que la actividad sea ejecutada por un robot, han sido varios los criterios ensayados para el diseño de las bases de la responsabilidad por la producción de un daño. Las propuestas oscilan desde una reformulación de las clásicas categorías de la dogmática penal hasta el diseño de nuevos modelos al modo de las previsiones relativas a la responsabilidad de las personas jurídicas<sup>93</sup>, pasando por las que simplemente proponen, todo lo más, una ampliación de la Parte Especial para incorporar modernas figuras delictivas que tengan en cuenta la nueva realidad, pero desde la renuncia a contemplar los robots como centros de imputación, condición que solo correspondería a quienes intervienen en las distintas secuencias de su funcionamiento<sup>94</sup>.

Sin perjuicio de hacer referencia más adelante a la posible responsabilidad como tal de los sistemas inteligentes, interesa de momento centrar la atención en el diseño de las bases de la eventual responsabilidad del sanitario que actúa en colaboración con el robot. La cuestión se traduce, en definitiva, en determinar los parámetros conforme a los cuales deba calificarse su actuación como respetuosa o no con los deberes de cuidado que pesan sobre él.

Para empezar, habría que diferenciar dentro de la amplia gama de robots quirúrgicos distintos tipos atendiendo a su grado de autonomía respecto de la intervención humana<sup>95</sup>. A los efectos

<sup>90</sup> Se cuentan, entre otras tantas, las prótesis para recuperar la locomoción, incluyendo los exoesqueletos, los asistentes personales, los robots de laboratorio, los destinados a la dispensación de medicamentos, o los de radioterapia. Entre estos últimos, interesante resulta la cita de *Cyberknife*, que administra dosis precisas de radiación para el tratamiento de tumores. Entre ellos destacan las posibilidades que ofrecen sistemas como *Da Vinci*, que no solo permite una visión tridimensional ampliada de la zona a operar, sino que está dotado de unos brazos robóticos con mayor precisión y posibilidades de movimiento superiores a la mano del cirujano. Puede verse un recorrido por los distintos tipos de robots y sus aplicaciones en "La inteligencia artificial y sus aplicaciones ÁVILA-TOMÁS, José Francisco/ MAYER-PUJADAS, Miguel Angel/ QUESADA VARELA Víctor Julio, "La inteligencia artificial y sus aplicaciones en medicina I: Introducción antecedentes a la IA y robótica", *Atención Primaria*, núm. 52, 2020, p. 782 ss.

<sup>91</sup> Las ventajas que se esperan de ellos van desde la menor invasión de las intervenciones hasta la mayor precisión de la cirugía, pasando por la posibilidad de aumentar el número de pacientes tratados. Sobre sus ventajas e inconvenientes puede verse ÁVILA-TOMÁS, José Francisco/ MAYER-PUJADAS, Miguel Angel/ QUESADA VARELA Víctor Julio, "La inteligencia artificial y sus aplicaciones en medicina I: Introducción antecedentes a la IA y robótica", *op. cit.*, p. 782.

<sup>92</sup> Por todos, QUINTERO OLIVARES, Gonzalo, "La robótica ante el derecho penal: El vacío de respuesta jurídica a las desviaciones incontroladas", *Revista Electrónica de Estudios Penales y de la Seguridad*, núm. 1, 2017, pp. 7 ss.

<sup>93</sup> Contemplando la posibilidad sobre la base del modelo de la responsabilidad de las personas jurídicas, ROMEO CASABONA, Carlos María, "La atribución de responsabilidad penal por los hechos cometidos por sistemas autónomos inteligentes, robótica y tecnologías conexas", e *Estudios político-criminales, jurídico-penales y criminológicos: libro homenaje al profesor José Luis Díez Ripollés* (dirs. Muñoz Sánchez, Juan/ García Pérez, Ocatvio,/ Cerezo Domínguez, Ana Isabel, I/ García España, Elisa), Tirant lo Blanch, Valencia, 2023, pp. 721 ss. También publicado en *Revista de Direito da ULP*, vol. 16, 2022, p. 14 s.

Sobre las distintas opciones referidas en el texto puede verse, entre otros, DOMÍNGUEZ PECO, E., "Los robots en Derecho penal", *Derecho de los robots*, 2ª ed., Madrid, 2019, pp. 171 ss.

<sup>94</sup> LLEDÓ BENITO, Ignacio, *El Derecho penal, robots, IA y cibercriminalidad: desafíos éticos y jurídicos*, Madrid, 2022, p. 75, 83 ss.

<sup>95</sup> Suelen identificarse tres categorías de robots: los llamados pasivos o de bajo riesgo, con funciones muy limitadas; los restringidos, que si bien tienen mayor autonomía y comportan mayores riesgos, tienen una función restringida, y los

que aquí se tratan, resulta interesante la distinción entre aquellos diseñados con el propósito de completar los recursos actuariales y las fuentes de decisión dentro del equipo sanitario, esto es, los llamados robots colaborativos, frente a aquellos otros en los que el robot puede actuar con mayor grado de autonomía. El primero de los modelos es el que en la actualidad ha encontrado aplicación práctica en el ejercicio de la medicina. Tal vez la mejor muestra la ofrezca el conocido como robot Da Vinci<sup>96</sup>, un sistema empleado en intervenciones variadas, como ginecología, urología u oncología, que en absoluto sustituye la labor del cirujano, sino que mejora su precisión gracias a una mayor visión -tridimensional- de la zona intervenida así como a un elevado grado de control del movimiento, al estar dotado de unos brazos robóticos que ofrecen un superior grado de precisión, firmeza y posibilidades de movimiento respecto a la mano del cirujano<sup>97</sup>.

En tanto que en esta tipología de robots sigue siendo el profesional quien realiza la intervención con la asistencia del sistema, conforme al modelo conocido como *Human-in-the-Loop*<sup>98</sup>, la eventual responsabilidad por un resultado lesivo se derivará conforme a los patrones generales de la imprudencia. Esta surgirá, en primer lugar, allí donde aquél procediera a su manejo sin contar con la necesaria formación y habilidades, casos en los que cabría hablar de responsabilidad por el emprendimiento mismo de la actividad<sup>99</sup>. En segundo lugar, y es el ámbito más complejo, podrá derivarse responsabilidad para el profesional por su actuación negligente ante los fallos del sistema frente a los que no reaccionase con la capacidad y destreza esperable para continuar la intervención recurriendo a los procedimientos clásicos<sup>100</sup>. Se trataría, en definitiva, de tomar en cuenta el deber de cuidado esperable de un profesional con la cualificación y capacidades exigibles y cotejarlo con la actuación llevada a cabo por el actor en cuestión a la vista de todas las circunstancias concurrentes, entre las que necesariamente habrán de contarse los medios concretos a su alcance.

Diferente y más compleja es la situación que se plantea allí donde el sistema no se limite a ser una herramienta en manos del médico, sino que goce de plena autonomía, como es propio del modelo *Human-out-of-the-Loop*. Ciertamente es que este tipo de robots se encuentra en la actualidad en fase de experimentación, de modo que la cuestión no tiene hoy en día traducción práctica, aunque sí es de esperar que la tenga en un futuro. El ejemplo más conocido hasta la fecha es el llamado *Smart Tissue Autonomous Robots* (STAR), ensayado de momento con animales, y que

---

activos, que comportan mayor riesgo y se caracterizan por el mayor protagonismo del robot, BERRY, Marcos, "Medicina y Robótica", *Revista Médica Clínica Las Condes*, Vol. 16, núm. 3, 2005, p.160 pp. 158 ss., donde ofrece un detallado cuadro de la aplicación práctica de cada uno de estos tipos de robots.

<sup>96</sup> Sobre las ventajas que ofrece este sistema puede verse el ya citado proyecto *RoboLaw*, una iniciativa de la Comisión Europea para el estudio de los desafíos éticos y legales que plantean los robots, cuyas labores concluyeron en mayo de 2012. Los resultados del proyecto se publicaron en un documento titulado "Guidelines for Regulating Robotics", pp. 29 ss. Disponible en: [http://www.robotlaw.eu/RoboLaw\\_files/documents/robotlaw\\_d6.2\\_guidelinesregulatingrobotics\\_20140922.pdf](http://www.robotlaw.eu/RoboLaw_files/documents/robotlaw_d6.2_guidelinesregulatingrobotics_20140922.pdf) (última consulta: 12 de junio de 2025).

Entre otros robots que comparten igualmente la característica de demandar supervisión médica puede citarse el Stryker Mako, utilizado principalmente en reemplazos de rodilla o cadera (Puede obtenerse información al respecto en [rthoracle.com/es/library/sustitucion-total-de-cadera-stryker-mako-asistida-por-robot-con-implantes-accolade-ii-y-tritanium/](http://rthoracle.com/es/library/sustitucion-total-de-cadera-stryker-mako-asistida-por-robot-con-implantes-accolade-ii-y-tritanium/), última consulta 1 de agosto de 2025), o ROSA, empleado entre otras aplicaciones en neurocirugía y ortopedia, para el reemplazo de rodilla, información disponible en: <https://www.raaot.org.ar/index.php/AAOTMAG/article/download/1579/5025?inline=1> (última consulta: 1 de agosto de 2025).

<sup>97</sup> El cirujano realiza la intervención controlando los brazos del robot mediante una serie de mandos mientras visualiza de forma nítida y ampliada la zona intervenida. Los brazos del robot portan el instrumental (bisturíes, tijeras, grapadoras, etc.), aportando como beneficio no sólo la ya referida precisión, sino el hecho de necesitar incisiones más pequeñas, permitir una recuperación más rápida y ofrecer menores riesgos de complicaciones.

<sup>98</sup> Suelen distinguirse hasta cuatro modelos en la interacción de la máquina con los humanos atendiendo al grado de intervención de este. El control humano sobre la máquina corresponde al citado modelo *Human-in-the-Loop*. Controles más débiles se refieren con los modelos *Human-in-the-Loop-form-Exceptions* y *Human-on-the-Loop*, correspondiendo la autonomía plena al modelo *Human-out-of-the-Loop*. Al respecto, véase por todos LLANO ALONSO, Fernando Higinio, *Homo ex machina. Ética de la inteligencia artificial y Derecho digital ante el horizonte de la singularidad tecnológica*, Tirant lo Blanch, Valencia, 2024 pp. 138 ss.

<sup>99</sup> En ello insiste la Propuesta de Resolución del Parlamento Europeo con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103(INL)). Entre los Principios generales relativos al desarrollo de la robótica y la inteligencia artificial para uso civil, el 17 subraya "la importancia de una formación y una preparación adecuadas de los médicos y auxiliares sanitarios", "la necesidad de definir los requisitos profesionales mínimos que deberá cumplir un cirujano para poder utilizar los robots quirúrgicos"; y la "especial importancia que reviste la formación de los usuarios para que puedan familiarizarse con los requisitos tecnológicos en este ámbito".

<sup>100</sup> Si bien desde el punto de vista de la responsabilidad civil, véase GARCÍA-MICÓ, Tomás Gabriel, *Robótica quirúrgica y derecho de daños*, Marcial Pons, Barcelona, 2024, pp. 194 ss.

ha demostrado su capacidad para tomar decisiones en tiempo real y adaptar el curso de la cirugía a las necesidades de su evolución<sup>101</sup>. Caso de que el sistema llegara a validarse para intervenciones en humanos, su empleo no sólo permitiría su práctica en lugares remotos sin la presencia física de un médico o posibilitar que un cirujano controle varios robots e intervenciones desde un mismo quirófano<sup>102</sup>, sino aumentar la precisión de las intervenciones respecto a la realizada por un humano. Obviamente, la ausencia de la presencia física de un médico durante la intervención no significa que su uso no requiera intervención humana y, con ello, que no se plantee la consabida cuestión en torno a la identificación del responsable por un fallo. Por ello, y dejando de momento al margen el debate en torno a la posibilidad de considerar a los robots como centros de responsabilidad penal, la mirada se vuelve a los diseñadores, fabricantes, programadores y distribuidores del sistema, a los responsables de la creación de las condiciones de la intervención, así como a los supervisores del proceso.

Focalizando estas consideraciones en la eventual responsabilidad del profesional sanitario y orillando por ello la relativa a la que puedan incurrir los participantes en la fase de programación y puesta en práctica del sistema, se plantearía, en primer lugar, la del médico que autoriza el uso del robot en el caso concreto sin comprobar la concurrencia de las condiciones que lo hacen aconsejable, supuesto en el que podría apreciarse ya una actuación negligente por la realización misma de la actividad. En segundo lugar, y es lo que plantea mayor dificultad, la mirada se vuelve a la figura del supervisor, encargado en tiempo real del seguimiento de la actuación de otro, el robot. La determinación de las bases de su responsabilidad requiere tener en cuenta las ya mencionadas previsiones del Reglamento IA. Conforme a cuanto hubo ya ocasión de exponer, los sistemas de alto riesgo, como sería el caso del uso de robots en medicina, reclaman conforme al art. 14 de dicha norma una supervisión humana significativa, esto es, que sean “vigilados de manera efectiva por personas físicas durante el período que estén en uso”.

El deber de supervisión se convierte precisamente en base para fundamentar la eventual responsabilidad de dichos agentes, bien sea en comisión por omisión -allí donde no despliegan el control necesario-, bien por acción -allí donde realizaran de modo deficiente esa labor-. Para determinar cuándo sea este el caso, bien pudiera recurrirse a un paralelismo con la elaboración teórica de la imprudencia específicamente diseñada para los supuestos de actuación en equipo por parte de distintos profesionales, ordenada a identificar los centros de imputación en el caso de que se produzca un resultado lesivo y que, como es sabido, básicamente distingue entre una distribución vertical u horizontal de funciones entre los distintos participantes en el equipo<sup>103</sup>.

Cierto es que existen diferencias perceptibles entre uno y otro caso, en tanto que cualquier tarea realizada por la máquina reclama supervisión humana y de ningún modo puede considerarse como un centro autónomo de responsabilidad. No se trata, por ello, de trasladar sin más aquella construcción al terreno que ahora interesa, sino tan sólo de extraer de ella una serie de pautas o guías conforme a las que depurar la eventual responsabilidad humana por los fallos ocurridos durante la actuación del sistema. En efecto, a diferencia de cuanto se sostuvo más arriba en relación con las aplicaciones de la IA que solo sirven de apoyo al diagnóstico o a la elección del tratamiento más apropiado, en el caso de los robots quirúrgicos autónomos es posible hablar de una parcelación de funciones, en tanto que irrumpen en la práctica médica con la vocación de asumir competencias ejecutivas, si bien en el marco de una labor que necesariamente debe dirigir un superior, un ser humano. Por ello, si realmente se tomase como inspiración de su tratamiento aquel modelo, su similitud habría de encontrarse con el diseño de las relaciones verticales. Se insiste en la idea de mera inspiración en ese diseño, entre otras cosas, porque, por definición, la actuación del sistema debe ser permanentemente supervisada por un humano, principio ajeno a la dinámica propia de la división de funciones. El valor de aquel modelo habría de verse entonces en descartar la infracción del deber de cuidado, y con ello, considerar la conducta como permitida, allí donde resultase razonable confiar en el funcionamiento correcto

<sup>101</sup> Como se indica en el texto, se trata aún de un sistema experimental, habiéndose realizado la primera intervención en cerdos, consistente en la extracción de la vesícula biliar. STAR realizó la intervención de forma totalmente autónoma, desde la identificación de los órganos hasta el cierre de la herida.

<sup>102</sup> Sobre las ventajas de este tipo de robots, así como de los que permiten cirugía asistida, véase por todos, BERRY, Marcos, “Medicina y Robótica”, *op. cit.*, p.160.

<sup>103</sup> De ellas tuve ocasión de ocuparme ampliamente en *La responsabilidad penal del médico*, *op. cit.*, pp. 398 ss.



del robot, por no detectarse tras aquellas tareas de supervisión y dirección irregularidad alguna, ya sea *de facto* o incluso a la vista de indicios de fallos de funcionamiento.

Queda en el aire una cuestión de futuro: la relativa a la eventual responsabilidad penal del robot como tal, si es que el estado de la técnica permitiera llegar a alcanzar su ansiada autonomía y en el plano jurídico se le llegase a reconocer la debatida *personalidad electrónica*<sup>104</sup>. Los posibles modelos de imputación que entonces vendrían en consideración han sido y son objeto de un amplio debate doctrinal<sup>105</sup>, cuyos respectivos argumentos descansan en la posibilidad o no de identificar los presupuestos de la responsabilidad penal en este tipo de entes<sup>106</sup>. Sin poder entrar ahora en los pormenores de la discusión, baste con señalar que el argumento que encuentra mayor predicamento entre los partidarios de su exigencia es la comparación de su problemática con el reconocimiento de responsabilidad penal de las personas jurídicas, construcción que igualmente descansa en una ficción que, según sus defensores, no parece muy diferente a la que necesitaría la atribución de responsabilidad a los sistemas inteligentes<sup>107</sup>.

Claro es que, puestos a pensar, todo es posible. Y si posible parece alcanzar en el futuro un estado de la técnica que permita conferir a los robots plena autonomía de actuación -por ejemplo, realizando de una forma completamente autónoma una intervención quirúrgica-, posible será también diseñar un sistema penal que, inspirado con más o menos fidelidad en el propio de los entes sociales y salvando sus distancias<sup>108</sup>, permita imponerle directamente consecuencias

<sup>104</sup> Sobre el debate al respecto puede verse, entre otros, GARCÍA SÁNCHEZ, María Dolores, "Inteligencia artificial y oportunidad de creación de una personalidad electrónica" *Ius et Scientia*, vol 6, núm. 2, 2020, pág. 83 ss., disponible en: <https://revistascientificas.us.es/index.php/ies/article/view/> (última consulta: 1 de julio de 2025); LLANO ALONSO, Fernando Higinio, *Homo ex machina. Ética de la inteligencia artificial y Derecho digital ante el horizonte de la singularidad tecnológica*, op. cit., pp. 125 ss.

A favor de ese reconocimiento se pronunció la Propuesta de Resolución del Parlamento Europeo con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103(INL)). En el punto 59 f) de las Recomendaciones a la Comisión Europea, contemplaba la conveniencia de "crear una personalidad jurídica específica para los robots, de modo que al menos los robots autónomos más complejos puedan ser considerados personas electrónicas con derechos y obligaciones específicos, incluida la obligación de reparar los daños que puedan causar; la personalidad electrónico se aplicaría a los supuestos en que los robots puedan tomar decisiones autónomas inteligentes o interactuar con terceros de forma independiente". No obstante, como respuesta a esa previsión, 285 expertos de diferentes áreas y países suscribieron una carta al presidente de la Comisión Europea y otros dirigentes con competencia en la materia en la que mostraban su oposición al reconocimiento de una "personalidad electrónica". Se trata del documento *Robotics, Open Letter*, Disponible en: <http://www.robotics-openletter.eu> (última consulta: 12 de agosto de 2025). Conforme a ese documento, tal pretensión "presenta numerosos sesgos basados en una sobrevaloración de las capacidades reales, incluso de los robots más avanzados, una comprensión superficial de la impredecibilidad y la capacidad de autoaprendizaje, y una percepción del robot distorsionada por la ciencia ficción y algunos anuncios de prensa sensacionalistas recientes".

<sup>105</sup> Al respecto puede verse, por ejemplo, FERNÁNDEZ GODINHO, Inés, "La autonomía de la IA como argumento dogmático para una reflexión sobre el modelo antropocéntrico de responsabilidad penal", *La atribución de una responsabilidad jurídico penal e internacional de la Inteligencia artificial*, (Coord. García Sánchez, B., Jiménez García, F), Iustel, Madrid, 2023, donde sintetiza los distintos modelos de imputación a dichos sistemas, pp. 79 ss.

<sup>106</sup> Entre otros, CÁMARA ARROYO, Sergio, "¿Sueñan los andróides delincuentes con penas electrónicas? Sistemas de armamento autónomo y responsabilidad penal: nuevos desafíos para el Derecho penal", en *La atribución de una responsabilidad jurídico penal e internacional de la Inteligencia artificial*, (Coord. García Sánchez, B., Jiménez García, F), Iustel, Madrid, 2023, pp. 248 ss.

<sup>107</sup> Por todos, ROMEO CASABONA, Carlos María, "La atribución de responsabilidad penal por los hechos cometidos por sistemas autónomos inteligentes, robótica y tecnologías conexas", en *Estudios político-criminales, jurídico-penales y criminológicos: libro homenaje al profesor José Luis Díez Ripollés*, op. cit., pp. 721 ss.

En realidad, el debate está presente no solo en la literatura, sino en algunos documentos que precisamente lo incentivan, como es el caso de la ya citada Propuesta de Resolución del Parlamento Europeo con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103(INL)). En el punto S de las consideraciones introductorias se lee que es "cada vez más urgente abordar la cuestión fundamental de si los robots deben tener personalidad jurídica", en tanto que "cuanto más autónomos sean los robots, menos se los podrá considerar simples instrumentos en manos de otros agentes". La ya citada carta abierta *Robotics, Open Letter*, mostraba igualmente su oposición a la equiparación de los robots a las personas jurídicas, ya que ello "implica la existencia de personas humanas detrás de la persona jurídica para representarla y dirigirla. Y esto no ocurre con un robot".

<sup>108</sup> Las diferencias son evidentes, no sólo en cuanto a las diferentes necesidades de castigo, sino por sus respectivas dinámicas. Primero, porque el robot como tal no deja de ser el producto de un humano creador, un producto defectuoso que nada tiene que ver con el entramado de relaciones propias del ente colectivo guiado directamente por la actuación humana y con capacidad para dar cobertura a quienes actúan a su amparo. Es por ello imposible aislar sin fisuras el alcance de la responsabilidad de tal ente para quien ha sido su mano creadora o programadora, lo que no ocurre cuando se ventila la responsabilidad de una persona jurídica, cuya sanción no tiene filtraciones de ningún tipo hacia otros implicados. Al contrario, el ciclo va en una dirección distinta: es ella quien recibe su responsabilidad a partir de la actuación de las personas físicas que por ella intervienen, con independencia de que, conforme al modelo que se siga, se la haga responsable directamente por ella o su responsabilidad sea vicaria. Segundo, porque si se toma como referente el modelo de autorresponsabilidad de la persona jurídica conforme al cual a ésta se la hace responsable del delito por un defecto



penales, como pueda ser la inhabilitación para ser usados, esto es, su desprogramación, o incluso su destrucción, que habría de plantearse entonces, no sólo en relación al concreto robot, sino a todo la serie o modelo<sup>109</sup>. Otra cosa es que realmente esas consecuencias merezcan ser calificadas como genuinas penas aplicables a artilugios<sup>110</sup> o si, en el fondo y se las llame como se las llame, no pasen de ser efectos que habrían de padecer en última instancia sus creadores o programadores, quienes verían entonces frustradas todas las ventajas derivadas de su puesta en el mercado en el caso de la destrucción, la retirada del mercado o su reprogramación.

La respuesta depende de hasta dónde llegue el avance de la tecnología y hasta donde queramos llegar en lo jurídico. Porque por ver está la autonomía futura y desarrollo de los robots y hasta dónde puede, debe o conviene llegar con el Derecho penal para su sanción directa y, sobre todo, con qué etiquetas. Desde luego, como quedó dicho, límites a la construcción jurídica no hay, entre otras cosas porque también el sistema penal, al igual que los robots, es obra del ingenio humano y este puede llegar hasta donde llegue lo posible y el consenso lo admita, borrando de un plumazo si fuera preciso las ataduras dogmáticas actuales para su responsabilidad. Nadie debiera espantarse de ello, aunque sí debiera ser motivo de reflexión que la única razón tuviera que ver con consideraciones meramente utilitaristas, llamadas a calmar la demanda de sanción por el daño allí donde no se alcanzase a determinar un humano responsable. El tiempo dirá esto y todo lo demás. A fecha de hoy, lo único cierto es que elevar los robots a centros de imputación es demasiado lejano<sup>111</sup>; tan lejano como la distancia que media entre las bases del Derecho penal que aprendimos y la idea de sentar en un banquillo a entes con alma de software.

---

de organización, tal defecto resulta difícil de atribuir al ente artificial –léase un defecto de programación o desarrollo de su actuación conforme a lo programado-. En última instancia el responsable de todo ello es el creador del producto, no el producto creado, por mucho que la creación tenga entre sus funciones “pensar”. Siempre lo hará conforme al programa o software que le dio “vida”.

<sup>109</sup> “La atribución de responsabilidad penal por los hechos cometidos por sistemas autónomos inteligentes, robótica y tecnologías conexas”, en *Estudios político-criminales, jurídico-penales y criminológicos: libro homenaje al profesor José Luis Díez Ripollés*, op. cit., pp. 721 ss.

<sup>110</sup> Tanto por la dificultad de identificar los presupuestos que reclama su imposición, como por los fines que reclama su castigo. La literatura es muy abundante al respecto, pudiendo verse, entre otros, BLANCO CORDERO, Isidoro, “Un derecho penal de humanos para entidades artificiales inteligentes. ¿machina delinquere et puniri potest?”, *Manifestaciones de desigualdad en el sistema de justicia penal* (Dir. Doval País, A., Gutiérrez Pérez, E., vol. 1, Thomson Reuters Aranzadi, Pamplona, 2023, pp. 371 ss.

<sup>111</sup> ROMEO CASABONA, Carlos María, “La atribución de responsabilidad penal por los hechos cometidos por sistemas autónomos inteligentes, robótica y tecnologías conexas”, en *Estudios político-criminales, jurídico-penales y criminológicos: libro homenaje al profesor José Luis Díez Ripollés*, op. cit., pp. 721 ss.

## BIBLIOGRAFÍA CITADA

AMUNÁTEGUI PERELLÓ, Carlos, *Arcana Technicae. El Derecho y la inteligencia artificial*, Arcana Technicae. El Derecho y la inteligencia artificial, Valencia, Tirant lo Blanch, 2020.

ARMENGOL DE LA HOZ, Miguel Ángel/ SÁNCHEZ GARCÍA, Miguel, "Big Data Analysis y Machine Learning en medicina intensiva", *Medicina Intensiva*, 2018.

ÁVILA-TOMÁS, José Francisco/ MAYER-PUJADAS, Miguel Angel/ QUESADA VARELA Víctor Julio, "La inteligencia artificial y sus aplicaciones en medicina I: Introducción antecedentes a la IA y robótica", *Atención Primaria*, núm. 52, 2020.

- "La inteligencia artificial y sus aplicaciones en medicina II: Importancia actual y aplicaciones prácticas", *Atención Primaria*, núm. 53, 2020.

BECK, Susanne / FABER, Michelle /GERNDT, Simon, "Rechtliche Aspekte des Einsatzes von KI und Robotik in Medizin und Pflege", *Ethik in der Medizin*, núm. 35, 2023

BERRY, Marcos, "Medicina y Robótica", *Revista Médica Clínica Las Condes*, vol. 16, núm. 3, 2005.

BLANCO CORDERO, I., "Un derecho penal de humanos para entidades artificiales inteligentes. ¿machina delinquere et puniri potest?", *Manifestaciones de desigualdad en el sistema de justicia penal* (Dir. Doval País, A., Gutiérrez Pérez, E., vol. 1, Thomson Reuters Aranzadi, Pamplona, 2023.

BLÁZQUEZ RUIZ, Francisco Javier, "Riesgos para la privacidad en la aplicación de la inteligencia artificial al ámbito biosanitario. implicaciones éticas y legales", *Anales de la Cátedra Francisco Suárez*, vol. 56, 2022.

CÁRCAR BENITO, Jesús Esteban, "La asistencia sanitaria dentro del ámbito de la inteligencia artificial (IA): el problema de los derechos de los pacientes", *El tiempo de los Derechos*, vol.1, 2020.

CODELLA, Noel/ CAI, Junjie/ ABEDINI, Mani/ GARNAVI, Rahil/ HALPERN, Alan, SMITH, John., "Deep Learning, Sparse Coding, and SVM for Melanoma Recognition in Dermoscopy Images", en: Zhou, Luping., Wang, Li, Wang, Qian, Shi, Yinghuan. (eds,) *Machine Learning in Medical Imaging*, 6 th International Workshop, MLMI 2015, Held in Conjunction with MICCAI 2015, Munich, Germany, October, 2015, Proceedings, 2015.

COTINO HUESO, Lorenzo, "Nuevo paradigma en la garantía de los derechos fundamentales y una nueva protección de datos frente al impacto social y colectivo de la inteligencia artificial", en *Derechos y garantías ante la inteligencia artificial y las decisiones automatizadas* (dir. Cotino Hueso, Lorenzo), Thomson Reuters Aranzadi, Pamplona 2022.

DE LA CUESTA AGUADO, P., "Inteligencia artificial y responsabilidad penal", *Revista Penal México*, núm. 16-17, 2019.

DE MIGUEL BERIAIN, Íñigo., MURSULI YANES, Yenifer, "Sesgos, IA y Biomedicina: un comentario desde la ética y el Derecho", *Revista de Derecho y Genoma Humano*, núm. 59, julio-diciembre 2023.

DEL RÍO SOLÁ, María Lourdes/ LÓPEZ SANTOS, José María/VAQUERO PUERTA, Carlos, "La inteligencia artificial en el ámbito médico", *Revista Española de Investigaciones Quirúrgicas*, vol. XXI, núm. 3, 2018.

DÍEZ, Juan José, "Aplicaciones médico-sanitarias de la inteligencia artificial: una visión desde la práctica clínica", *Anales de la Real Academia de Doctores de España*, vol. 8, núm. 4, 2023.

DOMÍNGUEZ PECO, E., "Los robots en Derecho penal", *Derecho de los robots*, 2ª ed., Madrid, 2019.

FERNÁNDEZ GODINHO, Inés, "La autonomía de la IA como argumento dogmático para una reflexión sobre el modelo antropocéntrico de responsabilidad penal", *La atribución de una responsabilidad jurídico penal e internacional de la Inteligencia artificial*, (Coord. García Sánchez, B., Jiménez García, F), Iustel, Madrid, 2023.

FERRETI, A.M./ SCHEIDER, M./BLASIMME, A., "Machine Learning in Medicine. Opening the new data protection Black Box", *European Data Protection Law Review*, Núm. 4, vol. 3, 2018.

GALVIN GORDILLO, M., "Análisis legal del uso de los robots en medicina", *Ius et Scientia*, vol. 9, núm. 1, 2023.

GARCÍA-MICÓ, Tomás Gabriel, "Robótica quirúrgica y derecho de daños", Marcial Pons, Barcelona, 2024.

GARCÍA SÁNCHEZ, María Dolores, "Inteligencia artificial y oportunidad de creación de una personalidad electrónica", *Ius et Scientia*, vol. 6, núm. 2, 2020.

GÓMEZ GONZÁLEZ, Emilio/ GÓMEZ Emilia/ MARQUEZ-RIVAS, Javier, et al. *Inteligencia Artificial en Medicina y Salud: revisión y clasificación de las aplicaciones actuales y del futuro cercano y su impacto ético y social (traducción del original en inglés)*, arXiv: 2001.09778, 2020.

GÓMEZ RIVERO, M.C., "Los límites del deber de sigilo del médico en las situaciones de riesgo: especial referencia al ámbito de la genética", *Revista de Derecho y Genoma Humano*, núm. 26, 2007.

- *La responsabilidad penal del médico*, 2ª ed., Tirant lo Blanch, Valencia 2008.

GONZÁLEZ CABANES, Francisco/DÍAZ DÍAZ, Norberto, "¿Qué es la inteligencia artificial? En Gamero Casado, Francisco Luis (ed.), *Inteligencia artificial y sector público: Retos, límites y medios*, Tirant lo Blanch, Valencia, 2023.

GRUBE, Karl, "Is the future of medical diagnosis in computer algorithms?", *The Lancet Digital Health*, vol. 1, m. "A Tool, Not a Crutch": Patient Perspectives About IBM Watson for Oncology Trained by Memorial Sloan Kettering *Journal of Oncology Practice*, vol. 12, núm. 4, 2019.

HAMILTON, Jada, et al. "A Tool, Not a Crutch": Patient Perspectives About IBM Watson for Oncology Trained by Memorial Sloan Kettering", *Journal of Oncology Practice* Volume 15, Number 4, 2019.

HE Jianxing./ BAXTER Sally/ XU Jie/ XU Jiming/ ZHOU Xingtao/ ZHANG Kang, "The practical implementation of artificial intelligence technologies in medicine", *Naturemedicine*, núm 24, 2019.

ISLA GÓMEZ, Julián, "Aplicaciones de la inteligencia artificial en salud", *Inteligencia artificial y Bioética*, (ed. Amo Usanos, Rafael), Universidad Pontificia de Comillas, Madrid, 2023.

JIE Zou /ZHIYING, Zeng / LI, Li, "A meta-analysis of Watson for Oncology in clinical application", *Scientific Report*, núm.11, 2021.

LACOS MORATINOS, Guillermo, "Inteligencia artificial, sesgos y salud: dilemas en torno a la propuesta de Reglamento AIA", *Inteligencia artificial y Bioética* (ed. Lit. Amo Usanos, Rafael), Universidad Pontificia Comillas, Madrid, 2023.

LAUKYTE, Migle, "Privacidad e inteligencia artificial: cambios e incertidumbres", en *Inteligencia artificial y Bioética* (ed. lit. Amo Usanos, Rafael), Universidad Pontificia de Comillas, Madrid, 2023.

LLANO ALONSO, F.H., *Homo ex machina. Ética de la inteligencia artificial y Derecho digital ante el horizonte de la singularidad tecnológica*, Tirant lo Blanch, Valencia, 2024.

LLEDÓ BENITO, I., *El Derecho penal, robots, IA y cibercriminalidad: desafíos éticos y jurídicos*, Madrid, 2022.

MELLADO ORELLANA, Raúl / ESCOBAR ARRIAGA, Elizabeth / DE LA MORA MOLINA, Héctor / DÍAZ GREENE, Enrique Juan / RODRÍGUEZ WEBER, “Estudio de concordancia entre el sistema Watson for Oncology y la práctica clínica en pacientes con cáncer de mama dentro del Hospital Ángeles Pedregal”, *Acta Médica Grupo Ángeles*, vol. 21, núm. 4, 2023.

MIRÓ LLINARES, F., “Inteligencia artificial y justicia penal: Más allá de los resultados lesivos causados por robots”, *Revista de Derecho Penal y Criminología*, núm. 20, 2018.

NIETO MENGOTTI, *El Derecho penal frente a los robots*, Desafíos legales. Juristas con futuro, Madrid, 2016.

NÚÑEZ REIZ, Antonio / ARMENGOL DE LA HOZ, Miguel Ángel / SÁNCHEZ GARCÍA, Miguel, “Big Data Analysis y Machine Learning en medicina intensiva”, *Medicina Intensiva*, 2018.

OLIVÁN MUZÁS, A./ ALBUNDIO ESCALONA, S./ARIAS BASTO, Y.C./ JIMÉNEZ HERNÁNDEZ, S., “El uso de Big Data en la investigación médica”, *Revista Ocrono*, 2024.

PELCYGER, Ben, Artificial Intelligence in healthcare: Babylon Health & IBM Watson take the lead, Prescouter, 2017.

PERÍN, 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*, vol. 8, núm 1, 2019.

QUINTERO OLIVARES, Gonzalo, “La robótica ante el derecho penal: El vacío de respuesta jurídica a las desviaciones incontroladas”, *Revista Electrónica de Estudios Penales y de la Seguridad*, núm. 1, 2017.

REMENTERÍA, María José / PROL CASTELO, Guillermo/ URDA Beatriz/ NÚÑEZ CARPINTERO, Iker / CIRILLO, Davide/ VALENCIA, Alfonso, “Inteligencia artificial y medicina personalizada” (ed. lit. AMO USANOS, Rafael), *Inteligencia artificial y Bioética*, Universidad Pontificia de Comillas, Madrid, 2023.

ROMEO CASABONA, C.M., “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*, número extraordinario, 2019.

- “La atribución de responsabilidad penal por los hechos cometidos por sistemas autónomos inteligentes, robótica y tecnologías conexas”, *Revista de Direito SA*, vol. 16, 2022.
- “La atribución de responsabilidad penal por los hechos cometidos por sistemas autónomos inteligentes, robótica y tecnologías conexas”, en *Estudios político-criminales, jurídico-penales y criminológicos: libro homenaje al profesor José Luis Díez Ripollés* (dirs. Muñoz Sánchez, Juan/ García Pérez, Ocatvio,/ Cerezo Domínguez, Ana Isabel, I/ García España, Elisa), Tirant lo Blanch, Valencia, 2023.

ROMEO CASABONA, C.M., /NICOLÁS JIMÉNEZ, P., *Retos éticos y necesidades normativas en la actividad asistencial en medicina personalizada*, Fundación Instituto Roche, Madrid, 2018.

ROMEO CASABONA, C.M./ LAZCOZ MORATINOS, G., “Inteligencia artificial aplicada a la salud: ¿qué marco jurídico?”, *Revista de Derecho y Genoma Humano*, 52/2020.

RUEDA MARTÍN, M.A., “La concreción del deber objetivo de cuidado en el desarrollo de la actividad médico-quirúrgica curativa a la luz de los avances de los sistemas de Inteligencia Artificial”, *InDret*, 1, 2025.

SCHWELKART, Scott, "Who will be liable for Medical Malpractice in the Future? How the Use of Artificial Intelligence in Medicine Will Shape Medical Tort Law", *Minnesota Journal o Law, Science & Technology*, núm. 22, 2021.

TIAN Yulong /LIU Xiaodong / WANG Zixuan/ CAO Shougen / LIU Zimin / JI Qinglian/ LI Zequn./, SUN Yuqi / ZHOU Xin / WANG Daosheng / ZHOU Yanbing "Concordance Between Watson for Oncology and a Multidisciplinary Clinical Decision-Making Team for Gastric Cancer and the Prognostic Implications: Retrospective Study", *Journal of Medical Internet Reseach* 2020.

TORRALBO MUÑOZ Bartolomé /MORALES MORENO, África María, "Medicina, inteligencia artificial y Derecho penal: una relación compleja", en *Fodertics 10.0. Estudios sobre Derecho Digital* (coords., Bueno de la Mata Federico, González Pulido, I.), Comares, Granada, 2022.

VALLS PRIETO, Javier, "Inteligencia artificial, derechos humanos y bienes jurídicos", Aranzadi, Pamplona, 2021.

VALLS PRIETO, Javier, "Sobre la responsabilidad penal por la utilización de sistemas inteligentes", *Revista Electrónica de Ciencia Penal y Criminología*, 2022.

WACHTER S./MITTELSTADL B./ FLORIDI, I, "Why a right to explanation of automater decisionmaking does not exist in the general data protection regulation, *International Data Privacy Law*, 2017.

WATSON, David / KRUTZINNA, Jenny / BRUCE, Ian /GRIFFITHS, Christopher /MCINNES, Iain/ BARNES, Michael / FLORIDI, Luciano, "Clinical applications of machine learning algorithms: beyond the black box".