# La regulación de los productos sanitarios con Inteligencia Artificial

Itziar Alkorta Idiakez

tirant 4.0

# LA REGULACIÓN DE LOS PRODUCTOS SANITARIOS CON INTELIGENCIA ARTIFICIAL

## COMITÉ CIENTÍFICO DE LA EDITORIAL TIRANT LO BLANCH



Procedimiento de selección de originales, ver página web:
www.tirant.net/index.php/editorial/procedimiento-de-seleccion-de-originales

# LA REGULACIÓN DE LOS PRODUCTOS SANITARIOS CON INTELIGENCIA ARTIFICIAL

**Itziar Alkorta Idiakez**
*Profesora Titular de Derecho Civil UPV/EHU*

tirant lo blanch
Valencia, 2025

La presente obra ha sido sometida a la revisión de pares ciegos según el protocolo de publicación de la editorial a efectos de ofrecer el rigor y calidad correspondiente tanto en su contenido como en su forma, aplicándose los criterios específicos aprobados por la Comisión Nacional E 016 (BOE núm. 286, de 26 de noviembre de 2016).



© TIRANT LO BLANCH
EDITA: TIRANT LO BLANCH
C/ Artes Gráficas, 14 - 46010 - Valencia
TELFS.: 96/361 00 48 - 50
FAX: 96/369 41 51
Email: tlb@tirant.com
www.tirant.com
Librería virtual: www.tirant.es
ISBN: 978-84-1095-863-0
MAQUETA: Innovatext

Si tiene alguna queja o sugerencia, envíenos un mail a: *atencioncliente@tirant.com*. En caso de no ser atendida su sugerencia, por favor, lea en *www.tirant.net/index.php/empresa/politicas-de-empresa* nuestro procedimiento de quejas.

Responsabilidad Social Corporativa: http://www.tirant.net/Docs/RSCTirant.pdf

# *Índice*

*Capítulo 4*

**EL REGLAMENTO GENERAL DE PROTECCIÓN DE DATOS (RGPD)**



*Capítulo 5*

**EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL: ASPECTOS GENERALES**



*Capítulo 6*

**LOS SISTEMAS DE ALTO RIESGO: REQUISITOS ESPECÍFICOS PARA SU AUTORIZACIÓN**

# *Introducción*[1]

La transformación digital de la medicina, impulsada por tecnologías basadas en inteligencia artificial (IA), promete mejorar radicalmente la calidad de los servicios sanitarios, desde el diagnóstico precoz hasta los tratamientos personalizados. Sin embargo, esta revolución también plantea retos éticos, técnicos y normativos que demandan una regulación integral y proactiva para garantizar el despliegue de una IA ética y fiable en la sanidad. Este estudio aborda de manera crítica y exhaustiva los elementos fundamentales que configuran el panorama regulatorio europeo para la IA en sanidad, poniendo especial énfasis en los riesgos inherentes a esta tecnología y las soluciones propuestas por la legislación vigente.

El primer capítulo introduce el impacto de la IA en la medicina, destacando su capacidad transformadora en áreas como el diagnóstico por imágenes, el monitoreo remoto de pacientes y la planificación de tratamientos. También se analizan los riesgos asociados, especialmente los sesgos algorítmicos, que pueden perpetuar desigualdades y comprometer la calidad del cuidado médico, afectando a poblaciones vulnerables.

El capítulo segundo amplía esta perspectiva al explorar la evolución histórica de la ética algorítmica. A través de un análisis que integra aportaciones académicas, computacionales y éticas, se pone en evidencia cómo los principios de transparencia y no discriminación se han convertido en pilares esenciales de los marcos regulatorios europeos. Para completar el estudio de la ética computacional, también se señalan las limitaciones de la solución técnica para abordar desigualdades estructurales, subrayando la necesidad de enfoques interdisciplinarios.

La obra avanza en el tercer capítulo hacia una evaluación crítica del derecho antidiscriminatorio europeo. Este marco, previsto para una época anterior a la irrupción de la IA, presenta limitaciones importantes a la hora de abordar eficazmente la discriminación algorítmica en el sector sanitario,

---

[1] Esta monografía es resultado del proyecto de investigación europeo SMARTBEAR, Smart Big Data Platform to Offer Evidence-based Personalised Support for Healthy and Independent Living at Home, Grant agreement ID: 857172, y se inscribe en el Grupo de Investigación consolidado del Gobierno Vasco/Eusko Jaurlaritza, "Persona, familia y patrimonio" (GIC IT1445-22).

debido a la opacidad y complejidad de los sistemas de IA. Aquí se destaca la importancia de reformar y adaptar las herramientas jurídicas tradicionales para responder a las particularidades de los algoritmos modernos.

En los capítulos cuarto y quinto, se analizan dos normativas clave: el Reglamento General de Protección de Datos (RGPD) y el Reglamento de Inteligencia Artificial (RIA). Mientras que el RGPD proporciona una base inicial para el tratamiento de los datos en la IA sanitaria, con principios como la calidad de datos y la protección frente a decisiones automatizadas el RIA introduce un marco específico para la gestión de riesgos en sistemas de alto impacto, integrando requisitos de transparencia y supervisión humana.

La regulación específica de los productos sanitarios con IA se aborda en los capítulos sexto y séptimo, donde se estudian las exigencias del Reglamento de Productos Sanitarios (RPS) y el propio RIA para dispositivos de alto riesgo. Estos capítulos ponen de relieve la complejidad de armonizar las normativas sectoriales con los principios generales del derecho europeo, resaltando la necesidad de marcos integrados y específicos para garantizar la seguridad y equidad en aplicaciones médicas.

Finalmente, los capítulos octavo y noveno ofrecen un análisis de herramientas normativas innovadoras, como el test de equidad para productos sanitarios, y evalúan la responsabilidad por daños causados por sistemas de IA en el ámbito sanitario. Se abordan cuestiones críticas como la adecuación de los marcos de responsabilidad objetiva y subjetiva, y se profundiza en los mecanismos previstos por la regulación propuesta para aliviar la carga probatoria en casos de discriminación algorítmica.

Este libro no solo revisa la normativa actual, sino que también propone mejoras y adapta el debate regulatorio a las necesidades emergentes del sector sanitario, convirtiéndose en una referencia esencial para legisladores, desarrolladores, profesionales de la salud y académicos interesados en el desarrollo ético y equitativo de la inteligencia artificial.

*Capítulo 1*

# *El uso de la Inteligencia Artificial (IA) en medicina y su impacto en los pacientes*

La integración de la inteligencia artificial en el ámbito sanitario ha demostrado su capacidad para transformar profundamente los procesos médicos, desde el diagnóstico hasta la gestión personalizada de tratamientos y medicamentos. Sin embargo, este avance tecnológico no está exento de tensiones. Los sistemas de IA, diseñados para mejorar la eficacia y la precisión de los productos sanitarios y fármacos, presentan limitaciones que pueden comprometer tanto la equidad como la calidad de la atención médica. En este contexto, el primer capítulo de esta obra se centra en los sesgos algorítmicos, una problemática que, lejos de ser meramente técnica, impacta directamente en la justicia distributiva de los recursos sanitarios y en la protección de derechos fundamentales en el acceso a la salud.

En concreto, el análisis aborda el origen del sesgo en las distintas etapas del ciclo de vida de un sistema de IA, desde la recolección de datos hasta su implementación, así como los efectos que puede tener en la amplificación de desigualdades preexistentes en la atención médica. Mediante ejemplos específicos, se destacan las repercusiones de estos sesgos en grupos vulnerables poniendo de relieve los vacíos regulatorios y éticos asociados a su mitigación. Además, se explora la conexión entre los principios normativos europeos y los instrumentos específicos destinados a garantizar un funcionamiento inclusivo y equitativo de estas tecnologías en el ámbito de la salud.

## 1.1. EVOLUCIÓN DE LA IA MÉDICA Y SITUACIÓN ACTUAL

La inteligencia artificial constituye una herramienta esencial en la medicina moderna, que está llamada a transformar la asistencia sanitaria de forma radical, desde el diagnóstico y el tratamiento, hasta la planificación de los servicios sanitarios. Los sistemas de IA han demostrado su capacidad

para procesar grandes volúmenes de datos de manera eficiente, hallando patrones y mejorando los procesos para que los médicos y las organizaciones puedan tomar decisiones más rápidas y fundamentadas (Alkorta Idiakez, 2022-b).

Un ejemplo destacado se encuentra en el diagnóstico por imagen en el análisis de la retina, donde herramientas basadas en IA, permiten auto-diagnosticar enfermedades como la retinopatía diabética con una precisión que supera la experiencia de los especialistas (Rajesh, 2023; Channa, 2021). Asistentes de diagnóstico para el cáncer de mama (Emiroglu et al., 2020) o de pulmón se emplean rutinariamente en los hospitales. Otro ámbito revolucionado por la IA es la asistencia al diagnóstico médico mediante aplicaciones capaces de evaluar síntomas y ofrecer recomendaciones fundamentadas en algoritmos avanzados. El monitoreo de pacientes crónicos a través de herramientas virtuales ha dado un gran paso adelante en los últimos años. La medicina cuenta con dispositivos inteligentes capaces de detectar síntomas en tiempo real, desde alteraciones en el ritmo cardíaco hasta patrones vitales anómalos. Estos avances han mejorado significativamente la prevención y el manejo de enfermedades crónicas en pacientes de edad avanzada. Las soluciones referidas están llamadas a optimizar los recursos al reducir la necesidad de consultas físicas para casos menos urgentes (Pecqueux, 2022).

La gestión personalizada de tratamientos también ha evolucionado gracias a plataformas que emplean análisis predictivos, permitiendo a los médicos planificar intervenciones ajustadas a las características individuales de cada paciente, especialmente en áreas complejas como la oncología (Wang et al., 2024). Asimismo, sistemas de triaje sanitario automatizado, como los utilizados en los entornos hospitalarios, optimizan la evaluación de la urgencia de los pacientes, reduciendo tiempos de espera y mejorando la eficiencia operativa (Tahernejad et al., 2024). Más allá de estas aplicaciones específicas, la IA ha tenido un impacto notable en áreas como el análisis epidemiológico, especialmente durante la pandemia de COVID-19, al prever la propagación del virus y diseñar estrategias de mitigación (Pereira et al., 2022). Finalmente, el diseño, administración y evaluación de los medicamentos, también constituye un campo prometedor en el que la IA se está aplicando con éxito (EMA 2024).

Más allá de sus usos y aplicaciones en la medicina, resulta imprescindible analizar la propia evolución de la inteligencia artificial, que en los últimos años ha incorporado nuevos métodos predictivos basados en el aprendizaje automático. Esta evolución ha estado marcada por una trans-

formación profunda, desde los primeros modelos predictivos fundamentados en técnicas de regresión hasta la sofisticación actual impulsada por algoritmos de *machine learning*. En sus inicios, la construcción de modelos predictivos se apoyaba en métodos estadísticos tradicionales, como la regresión logística y lineal, cuyo poder predictivo dependía en gran medida de la capacidad humana para identificar y seleccionar variables relevantes. Aunque estos enfoques aportaron mejoras significativas, presentaban limitaciones inherentes para gestionar grandes volúmenes de datos heterogéneos y complejos.

El punto de inflexión se produjo con la irrupción del aprendizaje automático, el cual permitió trascender las restricciones de los modelos convencionales. A diferencia de sus predecesores, los modelos basados en aprendizaje automático (*machine learning*) poseen la capacidad de aprender patrones directamente a partir de los datos, sin necesidad de una programación explícita para cada regla de decisión. Este avance ha sido posible gracias al acceso a vastos repositorios de datos clínicos, el incremento exponencial de la capacidad computacional y el desarrollo de algoritmos cada vez más sofisticados, como los bosques aleatorios (*random forests*) y las redes neuronales profundas *(deep learning*).

La aplicación del aprendizaje automático en la medicina ha dado lugar a una proliferación de modelos predictivos capaces de abordar tanto diagnósticos como pronósticos con una precisión sin precedentes. Estos modelos no solo estiman la probabilidad de la presencia de una enfermedad o la ocurrencia de un evento clínico futuro, sino que también pueden identificar relaciones complejas entre variables que pasarían desapercibidas para el análisis humano tradicional. Así, han surgido herramientas predictivas que abarcan desde el riesgo cardiovascular, como el *Framingham Risk Score*, hasta modelos más recientes aplicados en oncología, enfermedades neurodegenerativas y, más recientemente, en la respuesta a pandemias globales, como se evidenció durante la crisis de la COVID-19.

Actualmente, la IA basada en aprendizaje automático no solo complementa el juicio clínico, sino que redefine la manera en que se conceptualiza el riesgo, se personaliza el tratamiento y se optimiza la atención al paciente. La creciente complejidad de estos modelos ha planteado problemas importantes. La opacidad de algunos algoritmos de aprendizaje automático, conocidos como "cajas negras", ha suscitado preocupaciones en torno a la transparencia, la reproducibilidad y la equidad en la toma de decisiones clínicas. El sesgo algorítmico, por su parte, constituye un problema generalizado en los sistemas de inteligencia artificial, que puede

llevar a perpetuar desigualdades en la atención médica, afectando especialmente a poblaciones minoritarias y a mujeres (Richter et al., 2020). A ello se suman cuestiones éticas relacionadas con la privacidad de los datos y la transparencia en los procesos de decisión, especialmente en sistemas complejos como las redes neuronales profundas a las que nos hemos referido.

En respuesta a estas inquietudes, se han desarrollado iniciativas corporativas y nuevos marcos normativos cuyo alcance y limitaciones serán objeto de estudio en el presente trabajo.

## 1.2. FIABILIDAD DE LOS RESULTADOS DE DIAGNÓSTICO DE LA IA

La utilización de sistemas de inteligencia artificial en el ámbito de la medicina ha ganado terreno en los últimos años, como acabamos de exponer, en particular para la generación de diagnósticos clínicos automatizados. No obstante, la fiabilidad de estos diagnósticos requiere análisis detallado. El primer problema al que se enfrenta la integración de estos dispositivos en la práctica clínica consiste en la calidad de los datos empleados en el entrenamiento y validación de estos sistemas. En segundo lugar, es preciso valorar las métricas de validación clínica como la sensibilidad y especificidad de los modelos de IA empleados. Un tercer factor relevante es el de la capacidad del personal sanitario que ha de emplear estos dispositivos para interpretar los resultados de salida y valorarlos de forma adecuada.

El primer aspecto fundamental a la hora de valorar la integración de diagnósticos basados en IA en la práctica clínica es el de la **calidad de los datos** utilizados para entrenar estos modelos. Los sistemas de IA dependen de la representatividad y calidad de los datos de entrada. Si los datos de entrenamiento no son diversos o no reflejan adecuadamente la población objetivo, las métricas de sensibilidad y especificidad reportadas pueden ser engañosas. Por ejemplo, un modelo entrenado exclusivamente con datos de una población homogénea podría mostrar altos niveles de especificidad en esa población, pero fallar al detectar enfermedades en poblaciones diferentes debido a la falta de generalización (Beam y Kohane, 2018). Los sesgos en los datos pueden dar lugar a tratamientos discriminatorios que generan daños físicos y morales en los pacientes (Parissi-Poumian, 2021). Esta cuestión será ampliamente desarrollada en este trabajo.

Un segundo aspecto relevante a la hora de emplear estos dispositivos en la práctica clínica rutinaria es del equilibrio entre la **sensibilidad y la especificidad de los modelos**.

La sensibilidad de un modelo de IA en medicina se refiere a su capacidad para identificar correctamente los casos positivos, es decir, aquellos en los que el paciente presenta la condición o patología que se quiere detectar. Un sistema con alta sensibilidad resulta adecuado en el cribado de patologías graves y tratables, como los cánceres en etapas iniciales o las enfermedades infecciosas de alto riesgo, donde el objetivo es detectar todos los casos que puedan presentar la enfermedad. En estos escenarios, un falso negativo puede tener consecuencias devastadoras para el paciente. Por ejemplo, en el cribado de cáncer de mama mediante mamografía, dar prioridad a la sensibilidad permite identificar la mayoría de los casos, aunque ello implique un aumento de falsos positivos. Esta solución ha demostrado ser efectiva para mejorar los resultados en salud pública a largo plazo (Elmore et al., 2005).

Por otro lado, la especificidad mide la capacidad del sistema para identificar correctamente los casos negativos. Los sistemas altamente específicos están recomendados en situaciones donde un diagnóstico erróneo podría llevar a intervenciones innecesarias o incluso dañinas, como pruebas invasivas o tratamientos agresivos. En enfermedades de baja prevalencia, donde los casos reales son escasos en relación con la población total evaluada, una baja especificidad puede llevar a que una gran parte de los resultados positivos sean en realidad falsos. Esto significa que, aunque el sistema diagnostique a muchas personas como afectadas, la mayoría de esos diagnósticos podrían ser incorrectos debido al reducido número de casos verdaderos en comparación con los errores generados (Grimes & Schulz, 2002).

El equilibrio entre sensibilidad y especificidad que deben reunir los sistemas de diagnóstico que incorporan IA depende de las características de la población objetivo y la prevalencia de la enfermedad. Por tanto, no existe una respuesta universal sobre si es mejor que un sistema diagnóstico sea más sensible o más específico. Esta decisión debe fundamentarse en el contexto clínico, considerando las consecuencias de los falsos negativos y positivos. En patologías graves y tratables, la sensibilidad suele ser prioritaria para asegurar que ningún caso pase desapercibido. En enfermedades con tratamientos invasivos o consecuencias importantes de un diagnóstico erróneo, la especificidad adquiere mayor importancia. La literatura destaca la importancia de adaptar estas decisiones al contexto clínico y poblacional específico para maximizar el impacto positivo de los sistemas diagnósticos basados en IA (Zhou et al., 2011).

Un tercer factor relevante es la **interpretabilidad** clínica de los resultados. Como acabamos de señalar, el médico debe saber que, pese a que

un sistema de IA puede alcanzar métricas aparentemente robustas en un entorno controlado de pruebas, su implementación en la práctica clínica puede arrojar falsos positivos y negativos. El personal clínico debe estar capacitado para interpretar métricas como sensibilidad y especificidad en el contexto de cada paciente y tomar decisiones integrales que combinen los resultados del sistema de IA con información clínica adicional, antecedentes médicos y preferencias del paciente.

La información destinada al médico debe ajustarse a los estándares profesionales y normativos que permitan un uso seguro y eficaz del producto. Como veremos más adelante, el Reglamento (UE) 2017/745 sobre Productos Sanitarios establece la necesidad de que los fabricantes proporcionen información técnica suficiente para que los profesionales sanitarios puedan comprender el funcionamiento de los productos, incluyendo aspectos relativos a la fiabilidad de los diagnósticos generados por la IA en función de las métricas de validación clínica como sensibilidad y especificidad de la IA, las posibles limitaciones de los algoritmos y los posibles sesgos inherentes al propio modelo. Este nivel de detalle técnico es indispensable para que el médico evalúe los riesgos asociados y tome decisiones clínicas basadas en los resultados proporcionados por el sistema de IA, integrándolos de manera adecuada en el contexto específico de cada paciente.

Además, la implementación de sistemas de IA debe ir acompañada de auditorías continuas y recalibraciones periódicas para garantizar que los modelos mantengan su precisión diagnóstica a lo largo del tiempo. La recalibración periódica puede ayudar a preservar la sensibilidad y especificidad del sistema, evitando degradaciones en su desempeño (Wiens et al., 2019).

Los factores descritos más arriba deberán ser tenidos en cuenta a la hora de evaluar riesgos y beneficios de las tecnologías sanitarias que incorporan IA, lo cual resulta especialmente relevante en un entorno clínico dinámico, donde los patrones de enfermedades, las tecnologías de diagnóstico y los tratamientos disponibles están en constante evolución.

## 1.3. EL SESGO EN LOS SISTEMAS DE IA

El sesgo en los sistemas de inteligencia artificial utilizados en productos sanitarios es una preocupación creciente debido a su potencial para perpetuar y amplificar desigualdades preexistentes en la atención médica. Este concepto hace referencia a patrones sistemáticos de error que gene-

ran resultados desproporcionados o injustos para determinados grupos de pacientes.

El sesgo surge, fundamentalmente, del uso de datos históricos que reflejan desigualdades sociales existentes. En el ámbito médico, esto ocurre cuando los sistemas de IA se entrenan con conjuntos de datos que no representan adecuadamente a todas las poblaciones.

Las deficiencias y baja calidad en los datos afectan tanto a la precisión diagnóstica como a las recomendaciones terapéuticas. Un sistema entrenado con datos incompletos puede subestimar la gravedad de ciertas enfermedades en grupos infrarrepresentados, lo que resulta en diagnósticos incorrectos o en tratamientos que no se ajustan adecuadamente a las necesidades específicas de esos pacientes. Este fenómeno se debe a que los algoritmos aprenden patrones sesgados hacia la población mayoritaria representada en el conjunto de datos de entrenamiento.

Para mitigar este efecto es imprescindible contar con conjuntos de datos representativos que incluyan la diversidad de las poblaciones objetivo. Además, las estrategias para abordar el sesgo deben implementarse en todas las etapas del desarrollo del sistema, desde la recopilación de datos hasta la validación y la evaluación en entornos reales (Suresh y Guttag, 2021). Sin estas medidas, los sistemas de IA corren el riesgo de perpetuar y amplificar desigualdades existentes, en lugar de contribuir a una atención médica más inclusiva y equitativa.

La literatura destaca la necesidad de investigaciones continuas y de auditorías específicas que evalúen el impacto del sesgo en el desempeño clínico de los sistemas de IA, lo cual resulta especialmente relevante si se quiere que la calidad de la atención médica y la equidad se conviertan en prioridades fundamentales para los sistemas de salud (Rajkomar et al., 2018).

### *a) Tipos de sesgo en función del ciclo de vida de un sistema de IA*

El sesgo algorítmico en los sistemas de inteligencia artificial es una problemática multifacética que puede provenir de diferentes fuentes, como la calidad de los datos de entrenamiento, las decisiones de diseño de los modelos y las condiciones en las que se despliegan los sistemas (Mehrabi et al., 2021). Además, puede presentarse en diversas etapas del ciclo de vida de estos sistemas, desde la recolección de datos hasta su implementación y retroalimentación.

Un caso emblemático de sesgo relacionado con los datos históricos es la infra representación de minorías étnicas en los conjuntos de entrenamien-

to, lo que lleva a sistemas que ofrecen resultados significativamente menos precisos para estos grupos (Obermeyer et al., 2019; Suresh y Guttag, 2021). Este sesgo puede ocurrir incluso sin incluir explícitamente variables sensibles, como raza o género, ya que las discrepancias en la representación de los datos son suficientes para generar disparidades en el rendimiento de los modelos. Por ello, garantizar que los sistemas de IA sean desarrollados con criterios que promuevan la equidad es un paso esencial para evitar que reproduzcan las desigualdades inherentes a los datos con los que fueron entrenados.

¿Qué tipo de sesgos son los más habituales en cada etapa del ciclo de vida del sistema?

El ciclo de vida del sistema condiciona la aparición de los sesgos por lo que resulta imprescindible conocer las etapas de las que se compone dicho proceso, para lo cual seguimos el modelo de la Agencia Europea del Medicamento (EMA, 2024).

En la fase de recolección de datos, el sesgo aparece cuando los conjuntos de datos empleados para entrenar los modelos no reflejan adecuadamente la diversidad de la población objetivo, lo que puede ocurrir, por ejemplo, cuando los datos históricos están dominados por información de pacientes de ascendencia europea, lo que resulta en modelos menos precisos para otras etnias (Suresh y Guttag, 2021). Se ha dicho ya que se trata de un problema grave, puesto que la insuficiente representación de ciertos grupos no solo limita la equidad del sistema, sino que perpetúa desigualdades preexistentes, exacerbando con ello las disparidades en el acceso y la calidad de la atención médica.

Durante la etapa de diseño de un modelo de inteligencia artificial, el sesgo puede surgir debido a las decisiones tomadas en el proceso de desarrollo del algoritmo. Estas decisiones incluyen la selección de las variables que el sistema utilizará y los objetivos de inferencia que se optimizarán. En muchos casos, las variables elegidas no son características reales sino *proxies*, es decir, representaciones indirectas de características que, aunque no sean sensibles por sí mismas, están correlacionadas con factores como la raza, el género o el nivel socioeconómico.

Un ejemplo claro de esto es el uso de códigos postales para evaluar el riesgo médico. Aunque los códigos postales no son variables sensibles por naturaleza, están frecuentemente asociados con condiciones socioeconómicas y demográficas que reflejan desigualdades sociales. Como resultado, un sistema que los utilice podría introducir discriminación indirecta en sus decisiones, afectando negativamente a ciertos grupos (Barocas y Selbst, 2016).

Este tipo de problema resalta la necesidad de analizar cuidadosamente las variables seleccionadas para los modelos de inteligencia artificial. Es esencial identificar posibles correlaciones indirectas y adoptar estrategias que minimicen el impacto de estos sesgos en el desempeño del sistema. Sin este análisis crítico, existe el riesgo de que los modelos perpetúen desigualdades estructurales, en lugar de ayudar a mitigarlos. Investigaciones recientes han subrayado la importancia de abordar estos defectos desde el diseño de sistemas de IA, proponiendo métodos para detectar y corregir sesgos en las etapas iniciales de desarrollo (Mehrabi et al., 2021).

Durante la implementación del sistema, el contexto en el que se implementará constituye un elemento central a tener en cuenta. Los sesgos emergen cuando un modelo diseñado para un entorno específico se utiliza en otro contexto diferente, lo que compromete su precisión y relevancia. Un algoritmo desarrollado en un hospital urbano con recursos avanzados puede no ser aplicable a un centro rural con características demográficas y tecnológicas distintas (Mehrabi et al., 2021). Este desajuste entre el diseño y el uso real del sistema subraya la necesidad de adaptar los modelos al entorno en el que se aplicarán para evitar consecuencias negativas.

El sesgo de retroalimentación, finalmente, ocurre cuando las decisiones tomadas por un sistema de IA sirven para informar al propio sistema, creando un ciclo de auto-reforzamiento de patrones sesgados. Por ejemplo, un modelo sesgado de selección de candidatos que favorezca consistentemente a ciertos perfiles se utilizará en futuras iteraciones del sistema, perpetuando desigualdades estructurales (Barocas y Selbst, 2016). Esta dinámica puede ser especialmente problemática en entornos médicos, donde las decisiones basadas en datos sesgados afectan directamente a la calidad y equidad del cuidado de los pacientes.

### *b) Casos documentados de sesgo algorítmico*

El caso COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) ejemplifica cómo los algoritmos pueden perpetuar sesgos raciales en el sistema de justicia penal. Este sistema, utilizado en Estados Unidos para predecir la probabilidad de reincidencia de los presos, fue objeto de una investigación por parte de la agencia ProPublica en 2016. La citada agencia analizó las predicciones de riesgo que proponía COMPAS comparándolos con resultados reales de reincidencia del condado de Broward, Florida. Los hallazgos revelaron una discrepancia significativa: los internos negros tenían casi el doble de probabilidades que los blancos de ser clasificados erróneamente como de alto riesgo de reinciden-

cia, mientras que los acusados blancos, etiquetados como de bajo riesgo, reincidieron con mayor frecuencia (Angwin et al., 2016).

Para llegar a estas conclusiones, ProPublica planteó su análisis en varias etapas. Primero, calculó las tasas de falsos positivos (clasificaciones incorrectas de alto riesgo) y falsos negativos (clasificaciones incorrectas de bajo riesgo) por raza; y, a continuación, ajustó los modelos para considerar variables como la edad y los antecedentes penales. Este proceso reveló que, aunque COMPAS cumplía con la métrica de «paridad predictiva»—es decir, igualdad en la probabilidad de reincidencia para un mismo nivel de riesgo—, sus decisiones seguían mostrando disparidades raciales significativas debido a los datos históricos con los que se había alimentado el sistema (Angwin et al., 2016; Lagioia et al., 2022).[2]

Mediante este ejemplo pretendemos poner de manifiesto que la detección y corrección de sesgos algorítmicos no está al alcance de cualquiera. Identificar estos problemas requiere un conocimiento especializado en estadísticas, aprendizaje automático y una comprensión profunda de las implicaciones éticas y sociales de las decisiones algorítmicas. La complejidad técnica inherente a estos sistemas, junto con la necesidad de emitir juicios de valor en la implementación de criterios de equidad, convierte este proceso en una tarea extremadamente compleja y especializada.

Por ello, cuando los afectados son personas físicas, normalmente no podrán enfrentarse de manera aislada a la tarea de demostrar los sesgos y resultados discriminatorios de estos sistemas. La complejidad de los algoritmos y la falta de transparencia en sus procesos internos dificultan que una

---

[2] El trabajo de Lagioia et al. (2022) profundiza en la evaluación de la equidad algorítmica mediante criterios de paridad grupal, utilizando el sistema COMPAS y sus simplificaciones, SAPMOC I y SAPMOC II, como ejemplos. Los autores demuestran que un sistema que es igualmente preciso para diferentes grupos puede no cumplir con los estándares de paridad grupal debido a diferentes tasas base en la población. Argumentan que la igualación de clasificaciones o decisiones entre grupos puede lograrse mediante el ajuste de umbrales específicos para cada grupo, y cuestionan en qué contextos puede ser útil esta aproximación para alcanzar objetivos sociales. Concluyen que la implementación de estándares de paridad grupal debe ser responsabilidad de tomadores de decisiones humanos competentes, bajo una supervisión adecuada, ya que implica elecciones políticas basadas en valores discrecionales. Por lo tanto, los sistemas predictivos deben diseñarse de manera que los objetivos políticos relevantes puedan implementarse de forma transparente. Agradezco a los autores las explicaciones proporcionadas con ocasión del Seminario *Algorithmic Fairness in AI* celebrado en la EUI, Florence School of Transnational Governance, 9 a 12 de octubre de 2024.

persona no experta pueda identificar y sustentar, por sí sola, la existencia de discriminación algorítmica.

Para mitigar el sesgo es imprescindible contar con auditorías algorítmicas previas al despliegue de los sistemas, llevadas a cabo por equipos interdisciplinarios independientes. Estas auditorías deben evaluar tanto la equidad de los datos como la adherencia del sistema a estándares éticos.

Cuando se producen daños derivados de decisiones sesgadas, las acciones colectivas se vuelven esenciales para garantizar la rendición de cuentas y la reparación de los perjuicios. Estas acciones colectivas no solo permiten amplificar el alcance de las denuncias, sino que también facilitan una supervisión más rigurosa y eficaz de los sistemas de inteligencia artificial. En este contexto, los marcos éticos y regulatorios juegan un papel central para promover la equidad y la responsabilidad en el desarrollo y uso de estos sistemas, subrayando la importancia de la transparencia y la supervisión humana en todas las etapas del ciclo de vida de la inteligencia artificial (Barocas y Selbst, 2016). De ambas cuestiones nos ocuparemos extensamente en los capítulos siguientes.

En el ámbito de la medicina, los sesgos algorítmicos han demostrado tener impactos negativos tanto en la equidad del sistema como en la calidad de la atención médica. Un caso paradigmático es el de un algoritmo ampliamente utilizado en Estados Unidos para priorizar pacientes en programas de atención a pacientes de alto riesgo. Este sistema empleaba el costo económico de la atención prestada como un indicador de la necesidad médica, bajo la premisa de que los pacientes que habían generado mayores costos requerían mayor atención. Sin embargo, debido a las disparidades históricas en el acceso a la atención médica, los pacientes negros, que tradicionalmente habían recibido menos servicios y, por tanto, habían generado menores costos, fueron sistemáticamente subestimados por el algoritmo, incluso cuando sus condiciones de salud eran igual de graves que las de otros grupos. Este error implicó que muchos pacientes negros quedaran fuera de programas que podían mejorar significativamente su salud, perpetuando así las desigualdades raciales existentes (Obermeyer et al., 2019).

Otro ejemplo significativo es el de los algoritmos diseñados para la detección de cáncer de piel mediante el análisis de imágenes. Estos sistemas han mostrado una notable falta de precisión en pacientes con piel más oscura. La razón subyacente es que los conjuntos de datos utilizados para entrenar los modelos estaban predominantemente compuestos por imágenes de personas con piel clara, lo que limitó la capacidad del sistema

para identificar correctamente lesiones en tonos de piel más oscuros. Este sesgo no solo afecta la eficacia del diagnóstico, sino que también tiene consecuencias graves para la salud pública, al retrasar el tratamiento o llevar a diagnósticos erróneos en poblaciones ya vulnerables (Mehrabi et al., 2021).

En genética médica la IA se utiliza para analizar datos genómicos y proponer tratamientos personalizados. En este campo los algoritmos suelen estar entrenados con datos de personas de ascendencia europea, dejando de lado la diversidad genética de otras poblaciones, lo cual genera diagnósticos menos precisos y tratamientos subóptimos para personas de otras etnias (Popejoy y Fullerton, 2016).

En áreas como la cardiología, los algoritmos predictivos diseñados para evaluar riesgos de enfermedades cardíacas tienden a ignorar los síntomas que son más comunes en mujeres, ya que estos sistemas han sido entrenados mayoritariamente con datos de hombres como demostró el estudio *Framingham*. Dichos sesgos han llevado a tasas más altas de diagnósticos erróneos o tardíos en mujeres, perpetuando desigualdades de género en la atención médica (Vasan, 2018).

Durante la pandemia de COVID-19, los sistemas de triaje basados en inteligencia artificial también suscitaron críticas debidas a los sesgos contenidos en los modelos o en los datos. Algunos algoritmos priorizaban el acceso a camas de cuidados intensivos basándose en métricas como la esperanza de vida o la fragilidad general del paciente, lo que excluía a menudo a personas con discapacidades o condiciones preexistentes, violando principios de equidad y derechos humanos (Wynants et al, 2020)

Estos casos ponen de manifiesto no solo la magnitud de los problemas asociados con el sesgo algorítmico en la medicina, sino también la necesidad urgente de abordarlos de manera integral. Los sesgos en los datos de entrenamiento, las decisiones de diseño y las métricas de evaluación pueden perpetuar desigualdades estructurales y socavar la confianza en las tecnologías médicas. Por ello, garantizar que los sistemas de inteligencia artificial sean justos, transparentes y equitativos no es solo un desafío técnico, sino también un imperativo ético y social que requiere una colaboración interdisciplinaria entre científicos, médicos, reguladores y comunidades afectadas. Este fenómeno exige un marco normativo sólido que asegure la equidad y no discriminación en el uso de IA en la sanidad.

*Capítulo 2*

# *Perspectiva computacional y académica de la ética algorítmica*

Desde mediados del siglo XX, el desarrollo de la inteligencia artificial ha suscitado intensos debates sobre sus implicaciones éticas en la atención médica, la justicia penal, el crédito financiero o el empleo. La historia de la equidad computacional se entrelaza con los primeros pasos de la informática y la cibernética, evolucionando de forma paralela como una disciplina que plantea mecanismos para que los sistemas de IA sean justos y libres de discriminación.

La idea de "ética algorítmica" emergió de la confluencia entre la ética computacional, el aprendizaje automático, la ciencia de datos y las ciencias sociales, y pretendía dar respuesta a preocupaciones sobre los sesgos y desigualdades que pueden surgir cuando los algoritmos se aplican en ámbitos sensibles a los que nos hemos referido en el capítulo anterior. El enorme impacto de la IA en el conjunto de la sociedad refleja la necesidad de abordar estos riesgos inherentes a los algoritmos y garantizar un control social de los mismos.

En los primeros años de la década de 2020, la equidad de la inteligencia artificial adquirió un lugar destacado en las agendas regulatorias de la Unión Europea. Se trataba de que la toma de decisiones automatizadas debía ser intrínsecamente justa y estar exenta de sesgos a todos los niveles: datos de entrada, modelos de algoritmos, procesos empleados y resultados generados.

Sin embargo, este principio, aunque ampliamente invocado, no cuenta con una definición explícita en los nuevos marcos legislativos europeos. Pese a los avances en la regulación de la IA en la Unión Europea a los que nos referiremos en los capítulos siguientes, persisten importantes lagunas, especialmente en lo referente a la prohibición explícita de la discriminación algorítmica. En su lugar, la equidad algorítmica se integra en una compleja red de valores y principios —como la no discriminación, la transparencia y la igualdad de derechos—, que se reflejan en instrumentos

normativos de diversa naturaleza y alcance. Esta indefinición ha generado numerosas incertidumbres, en particular sobre cómo distinguirla de valores próximos como la igualdad y la no discriminación. El problema se agrava en áreas como la atención sanitaria, donde estos sistemas tienen el potencial de afectar de manera negativa a individuos o poblaciones vulnerables, perpetuando patrones históricos de discriminación.

Por todo ello, este capítulo explora la evolución histórica del concepto de equidad algorítmica, examinando sus fundamentos académicos, computacionales y éticos, así como las soluciones tecnológicas, normativas y de gobernanza desarrolladas a lo largo de los años para afrontar los problemas éticos que plantea la inteligencia artificial.

A través del análisis de las diversas perspectivas que han surgido en la historia de esta disciplina desde los ámbitos académico, profesional y científico, se sientan las bases para abordar, en el siguiente capítulo, la regulación vigente en Europa, fundamentada en los principios éticos y computacionales aquí desarrollados.

## 2.1. EL PLANTEAMIENTO DEL PROBLEMA: BREVE HISTORIA DE LA EQUIDAD ALGORÍTMICA

La equidad en la inteligencia artificial encuentra sus primeros antecedentes en los debates académicos y en las reflexiones iniciales del ámbito de la informática y de la ciencia de datos. Walter Maner acuñó el término «ética computacional» en 1978 (Manner, 1978, 1993), destacando la necesidad de una disciplina ética específica para abordar los problemas derivados del uso de computadoras y algoritmos en relación con los derechos fundamentales. En esta misma época, la Asociación de Maquinaria Computacional (ACM, por sus siglas en inglés) estableció el primer código de ética (1972), marcando un hito en la formalización de principios éticos aplicados a la tecnología.

Durante las décadas de 1980 y 1990, la ética computacional se consolidó con contribuciones significativas como las de James Moor y Deborah Johnson. Moor, en su influyente artículo «*What is Computer Ethics?*» (1985), identificó vacíos normativos derivados del desarrollo de la informática y propuso políticas para el uso responsable de la tecnología. El mismo año, Johnson publicó el primer libro de texto relevante sobre ética computacional, listando los problemas éticos más críticos de esta disciplina.

El interés por la equidad algorítmica creció de manera exponencial en la década de 2010, cuando comenzaron a documentarse casos de sesgo en sistemas de inteligencia artificial. Investigaciones como las de Latanya Sweeney (2013) demostraron cómo los algoritmos de anuncios en línea podían reproducir prejuicios sociales. Cathy O'Neil, en su libro *Weapons of Math Destruction* (2016), denunció que algoritmos aparentemente neutrales podían discriminar a grupos vulnerables.

Paralelamente, estudios pioneros como los de Cynthia Dwork et al. (2012) introdujeron el concepto de «justicia individual», abogando por la necesidad de tratar de manera equitativa a individuos en circunstancias similares. Asimismo, Buolamwini y Gebru (2018) revelaron sesgos raciales y de género en sistemas de reconocimiento facial, evidenciando tasas de error notablemente mayores en personas de piel oscura y en mujeres. Estos hallazgos impulsaron una oleada de investigaciones y debates críticos sobre el impacto social de los sistemas algorítmicos.

El informe *Machine Bias* (2016), publicado por Angwin junto con la citada agencia ProPublica, profundizó en el uso de COMPAS, la herramienta algorítmica para evaluar el riesgo de reincidencia en el ámbito judicial, al que nos hemos referido en el capítulo anterior. Este trabajo sirvió como catalizador para la incorporación de la equidad algorítmica en la agenda de numerosos especialistas.

Aunque las soluciones propuestas para mitigar la discriminación algorítmica han tendido hacia soluciones tecnológicos, estas propuestas no están exentos de limitaciones (*). La mayor parte de los científicos de datos consideran la discriminación algorítmica como un problema técnico vinculado a la calidad de los datos y al diseño de los modelos. Este paradigma ha dado lugar a herramientas mecánicas, conocidas como mecanismos de «anti-sesgo» o «*debiasing*», que pretenden corregir sesgos mediante métricas como la paridad estadística, la igualdad de oportunidades y la equidad predictiva. Sin embargo, como argumentaremos más adelante, esta perspectiva reduccionista no aborda las desigualdades estructurales subyacentes.

Friedler y Scheidegger (2016) han desarrollado métodos para medir y reducir los sesgos en algoritmos aplicados en contextos sensibles como la justicia y el empleo, con el fin de implementar métricas que garanticen una equidad real en los sistemas algorítmicos. Ben Shneiderman, en su obra *Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy* (2022), subraya la necesidad de diseñar sistemas fiables y seguros, proponiendo un marco de inteligencia artificial centrada en el ser humano (HCAI, 2022).

El modelo se basa en la supervisión humana y la transparencia como elementos esenciales para garantizar que la IA sea una herramienta digna de confianza.

Finalmente es preciso traer a colación el trabajo de Karl Stöger et al. (2020, 2021), que ha contribuido significativamente al debate desde el contexto médico, destacando en su art. «*Legal Aspects of Data Cleansing in Medical AI*» la importancia de la calidad de los datos en aplicaciones de IA y la necesidad de marcos regulatorios claros para garantizar integridad y precisión. Sus hallazgos serán comentados en mayor detalle en el capítulo 8 de esta obra.

## 2.2. DECLARACIONES INTERNACIONALES PARA FOMENTAR LA EQUIDAD ALGORÍTMICA

Los primeros debates académicos en torno a la equidad algorítmica sentaron las bases para que organismos internacionales y asociaciones profesionales adoptaran los primeros principios éticos orientados al desarrollo y al uso responsable de la inteligencia artificial.

Estas iniciativas pretendían asegurar que los sistemas respeten derechos fundamentales como la no discriminación, la transparencia y la autonomía de los afectados por el tratamiento de los datos. Un hito inicial lo marcó el «Diseño Éticamente Alineado» del *Institute of Electrical and Electronics Engineers* (IEEE) en 2017, que propuso recomendaciones detalladas para garantizar la equidad y la responsabilidad en los sistemas autónomos.

En esa misma línea, la Declaración de Montreal sobre principios de IA responsable y los Principios de Asilomar elaborados *por Future of Life Institute*, ambos de 2017, abogaron por una IA que priorizara los derechos humanos y el bienestar social, promoviendo una visión humanista de la tecnología. Dos años después, el gobierno japonés publicó los *Principios Sociales de la IA centrada en el ser humano* (2019), en concordancia con lo propuesto por la OCDE para la regulación de la IA respetando los principios básicos de dignidad, diversidad, inclusión y sostenibilidad (OCDE, 2019). Asimismo, la federación sindical UNI Global Union (2020) estableció un decálogo ético que reforzó estos principios con una influencia notable en el ámbito de la discriminación laboral.

Durante más de un lustro, la *Global Partnership on Artificial Intelligence* (GPAI, 2020, 2023-a, 2023-b), formada en 2020 por países como Canadá, Francia, Alemania, Japón y Estados Unidos, impulsó la colaboración entre

académicos, sociedad civil y el sector privado para desarrollar tecnologías alineadas con principios éticos.

Más recientemente, la equidad algorítmica ha encontrado paralelismos con la bioética, en particular en la integración de tecnologías de IA en la atención sanitaria. Ya que los principios bioéticos establecidos por Beauchamps y Childress en 1979 —como la autonomía, la beneficencia, la no maleficencia y la justicia—, ofrecen un marco sólido para abordar problemas éticos en la IA. Así, la UNESCO publicó en 2021 una «Recomendación sobre la Ética de la Inteligencia Artificial», en la que estableció la equidad, la justicia y la transparencia como valores esenciales para el desarrollo responsable de la IA. En ella se dice que el principio bioético de la autonomía, referido al consentimiento informado y a la capacidad de decisión, presenta semejanzas con la necesidad de transparencia en los sistemas de IA. Por su parte, la beneficencia y la no maleficencia subrayan la importancia de minimizar los daños potenciales y maximizar los beneficios; mientras que la justicia pone de relieve la necesidad de garantizar que los algoritmos no reproduzcan desigualdades estructurales y promuevan un trato equitativo.

Otras organizaciones internacionales, como la Organización Mundial de la Salud, también han desarrollado directrices basadas en dichos principios bioéticos para regular la IA (2021). La propuesta de la OMS se alinea con el programa «AI for Good» de la ONU, promovido por la Unión Internacional de Telecomunicaciones en colaboración con diversas agencias de las Naciones Unidas. Más recientemente, en 2024, la Cumbre Internacional de Normas de IA, celebrada en Nueva Delhi, propuso nuevas directrices inspiradas en la bioética para resolver los dilemas éticos que plantea la inteligencia artificial.

Desde un punto de vista confesional, la *Pontifical Academy for Life*, a través de su «Rome Call for AI Ethics» (2020), ha desempeñado un papel relevante en la formulación de estándares éticos globales.

Finalmente, es preciso hacer referencia a los debates éticos que ha generado el desarrollo y el propio uso de la IA por parte de grandes empresas como Google y Amazon, IBM y otras firmas tecnológicas de gran tamaño.

Uno de los casos más notables de sesgo algorítmico está relacionado con el uso de algoritmos para automatizar procesos de contratación. IBM, al igual que muchas empresas tecnológicas, implementó sistemas diseñados para optimizar estas tareas, pero pronto surgieron preocupaciones sobre cómo estos algoritmos replicaban y amplificaban los sesgos contenidos en los datos históricos de contratación. Si los datos históricos reflejaban una preferencia por candidatos masculinos o provenientes de ciertas uni-

versidades, el algoritmo podía perpetuar estas tendencias, descartando automáticamente perfiles más diversos. Para atajar las críticas, IBM comenzó a ajustar sus sistemas y a promover prácticas más transparentes en la creación y uso de algoritmos con el fin de minimizar los sesgos y fomentar la equidad en los procesos automatizados.

En el caso de Google, uno de los incidentes más conocidos ocurrió con su motor de búsqueda, que en ocasiones asociaba términos raciales o de género con imágenes o información estereotipada (Buolamwini y Gebru, 2018). Por ejemplo, búsquedas como «CEO» arrojaban principalmente imágenes de hombres blancos, mientras que «trabajadoras domésticas» mostraba mujeres de ciertas etnias, lo cual evidenció cómo los datos culturales sesgados usados para entrenar algoritmos pueden influir negativamente en los resultados.

A raíz del despido de Timnit Gebru, autora del citado artículo de denuncia, se desató un debate global sobre la responsabilidad ética de Google y su falta de compromiso con prácticas inclusivas.

Además, los sistemas publicitarios de Google también han sido señalados por mostrar anuncios dirigidos de manera desigual. Investigaciones de la época revelaron que ofertas de empleos altamente remunerados se presentaban más frecuentemente a hombres que a mujeres, y que ciertos anuncios se asociaban desproporcionadamente con términos raciales, lo cual reflejaba no solo los sesgos en los datos de entrada, sino también las dinámicas económicas que priorizaban la maximización de ingresos sobre la equidad (Buolamwini y Gebru, 2018).

Aunque Google ha tomado medidas para abordar estas críticas, como realizar auditorías internas de sus algoritmos, desarrollar principios éticos para la IA y mejorar la representatividad en sus conjuntos de datos, muchos son los autores que consideran que estos esfuerzos son insuficientes (Buolamwini, 2019, Gebru et al., 2020; EFF, 2020).

El conjunto de iniciativas descritas en este apartado ha contribuido a posicionar la equidad como un principio central en el desarrollo de la IA basado en una métrica adecuada para medir su impacto social. El concepto ha servido de inspiración a regulaciones como el Reglamento de Inteligencia Artificial de la Unión Europea —al que nos referiremos en los capítulos 5 y 6 de esta obra—, y a otras propuestas como el proyecto de ley *Algorithmic Fairness Act* de 2020 en los Estados Unidos, orientada a reforzar la transparencia y la justicia en áreas críticas como la educación, el empleo, la salud y el crédito.

## 2.3. RECOMENDACIONES ESPECÍFICAS SOBRE LA IA APLICADA A LA MEDICINA PREDICTIVA

En el ámbito de la salud, existen guías de notable relevancia, tales como CONSORT-AI, promovida por Liu et al. (2020), y SPIRIT-AI, obra de Rivera et al. (2024), las cuales ofrecen directrices para ensayos clínicos que incorporan componentes de inteligencia artificial. Asimismo, la guía CLAIM (2024) se centra en estudios de imágenes médicas basados en IA. Por su parte, el *Good Machine Learning Practice (GMLP) for Medical Devices* (IMDRF, 2025) propone principios orientadores para asegurar que los dispositivos médicos que incorporan IA sean seguros, efectivos y de alta calidad a lo largo de su ciclo de vida.

Es imprescindible referirse también a la iniciativa TRIPOD+AI Statement (Collins et al., 2024) sobre IA de aprendizaje profundo aplicada a la medicina, por su carácter referencial en este campo. Esta guía asiste en la elaboración de informes sobre modelos de predicción clínica, reflejando los avances metodológicos en el uso de la inteligencia artificial y el aprendizaje automático. Promovida por un consorcio internacional de expertos liderado por Gary S. Collins y Karel G. M. Moons (2024), entre otros, la citada declaración TRIPOD+AI ha sido concebida bajo el auspicio de la red EQUATOR (*Enhancing the Quality and Transparency of Health Research*), que vela por la calidad en la comunicación de la investigación en salud. La motivación principal radica en la creciente preocupación por la falta de transparencia y la escasa reproducibilidad de los estudios que emplean IA de aprendizaje profundo, lo que puede comprometer la seguridad y la eficacia de los modelos predictivos cuando se implementan en la práctica clínica. Mientras que las anteriores versionas de esta declaración se orientan hacia intervenciones específicas o contextos particulares, TRIPOD+AI propone una armonización integral que abarque tanto el desarrollo como la evaluación del rendimiento de estos modelos, proporcionando una lista de 27 ítems con subapartados que detallan exhaustivamente los aspectos metodológicos y éticos a reportar para garantizar la reproducibilidad y la equidad de los prototipos.

Entre las propuestas clave de TRIPOD+AI destaca la incorporación de directrices para garantizar la equidad o *fairness* en los modelos predictivos. Este concepto se refiere a la capacidad de los modelos para evitar sesgos que puedan derivar en discriminación contra individuos o grupos en función de atributos como la edad, el género, la etnia o el estatus socioeconómico. Las recomendaciones en este ámbito no se limitan a un apartado aislado, sino que están integradas de forma transversal en el conjunto de ítems

que contempla la guía. Se insta, por ejemplo, a describir explícitamente cómo se ha abordado la representatividad de los datos de entrenamiento y validación, a informar sobre posibles desigualdades en los resultados del modelo entre diferentes subgrupos poblacionales, y a detallar las estrategias empleadas para mitigar sesgos durante el desarrollo y la evaluación del modelo.

Por otra parte, la Guía TRIPOD-LLM (Gallifant et al., 2024) constituye una extensión de la anterior que aborda las cuestiones regulatorias específicas que presentan los grandes modelos de lenguaje (LLM por sus siglas en inglés) en el ámbito médico. Esta guía consiste en 19 items principales desarrollados en 50 subítems que permiten a los promotores y desarrolladores tener en cuenta los elementos evaluables para desarrollar estos sistemas mitigando los posibles sesgos.

## 2.4. LAS SOLUCIONES TECNOLÓGICAS

La identificación y mitigación de sesgos en los sistemas de inteligencia artificial requiere un abordaje que combine herramientas técnicas con principios jurídicos sólidos. Este apartado da cuenta de las herramientas más destacadas desarrolladas para auditar algoritmos y evitar la discriminación algorítmica, organizándolas según su relevancia y ámbito de aplicación.

Entre las iniciativas más influyentes se encuentra el marco *Fairness, Accountability, and Transparency in Machine Learning* (FAT/ML), que establece principios y prácticas para asegurar que los sistemas de aprendizaje automático sean justos, responsables y transparentes. Lo cierto es que FAT/ML ha sido un pilar conceptual que ha guiado el desarrollo de numerosas herramientas específicas.

También el kit de herramientas *AI Fairness 360* (AIF360), desarrollado por IBM, representa un avance significativo al proporcionar un conjunto de métodos para identificar y mitigar sesgos en los modelos algorítmicos. Esta recomendación incluye métricas y técnicas para auditar tanto los datos como los algoritmos, ayudando a detectar disparidades en las decisiones automatizadas y a implementar correcciones. Complementariamente, la *AI Alliance*, también promovida por IBM, fomenta la cooperación internacional para establecer estándares de equidad algorítmica mediante el diálogo entre múltiples actores, incluidos desarrolladores, reguladores y académicos.

Otra herramienta notable es *Fairlearn*, creada por Microsoft, que permite a los desarrolladores evaluar y mejorar la equidad de sus modelos algorítmicos mediante análisis detallados de las disparidades en los resultados. *Aequitas*, del *Center for Data Science and Public Policy,* desarrollada por la Universidad de Chicago, amplía estas capacidades al ofrecer una evaluación exhaustiva de los impactos diferenciales de los algoritmos en diferentes grupos demográficos, facilitando la adopción de medidas correctivas.

En el contexto sanitario, cabe citar también la herramienta *FUTURE-AI* descrita por Lekadir et al. (2024), que destaca por su aproximación holística, abordando la equidad en todas las etapas del ciclo de vida del producto, tal como señalamos en el capítulo anterior. Entre las recomendaciones de *FUTURE-AI* sobresale la identificación temprana de posibles fuentes de sesgo, lo que permite a los desarrolladores anticipar y mitigar problemas antes de que afecten a los resultados. Además, se subraya la importancia de recopilar datos detallados sobre los atributos de los individuos, como género, etnia o edad, con el debido consentimiento informado. Uno de los problemas más interesantes que se aborda en esta herramienta es el equilibrio entre los beneficios de la no discriminación y los riesgos asociados a la reidentificación de datos sensibles.

El marco *FUTURE-AI* también se refiere a la gestión de riesgos, promoviendo la supervisión continua y la auditoría de las herramientas de IA, lo cual incluye la definición clara de los entornos clínicos previstos y la utilización de estándares científicos para garantizar la interoperabilidad y la calidad de los sistemas.

Esta herramienta recomienda el uso de métricas específicas para medir la equidad algorítmica. La tasa de paridad compara las decisiones positivas entre diferentes grupos, mientras que la equidad de oportunidades asegura que individuos con características similares reciban probabilidades iguales de obtener resultados positivos. Estas métricas se aplican en auditorías algorítmicas, que evalúan las disparidades en las decisiones generadas por los sistemas y guían las intervenciones que se deben realizar en los datos o en el diseño de los algoritmos.

A las herramientas tecnológicas como la citada *FUTURE-AI* se suman los estándares internacionales emitidos por organismos como ISO (ISO/IEC TS 12791:2024) y la CEN-CENELEC, a la que aludiremos más adelante, las cuales proporcionan marcos técnicos para garantizar la calidad y las buenas prácticas en la industria.

Finalmente, la Agencia Europea de Derechos Fundamentales y el Consejo de Europa (2021) señalan que la propia inteligencia artificial puede

desempeñar un papel activo en la supervisión de su propio uso, ayudando a construir evidencias sobre desigualdad y discriminación, e incluso revelando sesgos humanos ocultos en los procesos de decisión.

## 2.5. CRÍTICA DEL "SOLUCIONISMO" TECNOLÓGICO

A pesar de los avances producidos para tratar la discriminación algorítmica en la inteligencia artificial, la perspectiva tecno-céntrica a la que nos hemos referido en el epígrafe anterior, ha sido duramente criticada por su limitada efectividad ante la magnitud del problema que se presenta a la hora de corregir modelos que han sido construidos con millones de datos de internet sin ningún tipo de filtros. Morondo et al. (2022) han señalado al respecto que, aunque estas soluciones tecnológicas intentan corregir los sesgos mediante herramientas como métricas estadísticas y auditorías internas, suelen carecer de una conexión directa con los marcos jurídicos y, en muchos casos, resultan demasiado ambiguas para garantizar una protección efectiva del derecho a la igualdad y a la no discriminación.

El sistema de evaluación recomendado por las citadas herramientas técnicas, ha sido señalado por su carácter autorreferencial. Según Morondo et al. (2022) este modelo delega la evaluación de la equidad casi exclusivamente en los desarrolladores y diseñadores de los sistemas, sin incorporar una validación externa significativa de otros puntos de vista ni adaptarse a las complejidades de los entornos sociales donde operan estos algoritmos.

Así, la autoevaluación tiende a priorizar la percepción apriorística de justicia algorítmica sobre el impacto real en los colectivos afectados, perpetuando la tendencia al «solucionismo tecnológico». Esta perspectiva solipsista no solo ignora las desigualdades estructurales que existen en la sociedad y se reflejan en la información que esta genera, sino que plantea el riesgo de reforzar una visión reduccionista de la equidad, orillando la importancia de la supervisión independiente y de la responsabilidad social (Bosoer, Cantero Gamito, Rubio-Marin, 2024).

La práctica de limitarse a cumplir con listas de verificación internas que incluyen preguntas relacionadas con la diversidad y la no discriminación, puede derivar en un cumplimiento meramente superficial. Esta forma de proceder permite a muchas empresas satisfacer los requisitos técnicos sin lograr un impacto tangible en la reducción del sesgo o en el avance hacia una mayor equidad. Por ejemplo, los sistemas de evaluación

de currículums suelen reproducir desigualdades de género preexistentes, demostrando que los algoritmos pueden cumplir métricas estadísticas de paridad sin atender de manera efectiva las raíces de la discriminación estructural.

Además, como hemos visto más arriba, las soluciones tecnológicas anti-sesgo recomendadas por grandes empresas tecnológicas plantean dudas sobre si las herramientas diseñadas por ellas mismas para detectar y mitigar sesgos no perpetúan, en realidad, un análisis superficial del problema de fondo. Aunque investigaciones como la de Barocas et al. (2019) reconocen que los mecanismos anti-sesgo pueden aliviar el problema, persiste la tendencia a sobrevalorar su eficacia en los documentos regulatorios europeos.

Siguiendo a Floridi (2023), la equidad algorítmica debe trascender los planteamientos reduccionistas que se limitan a parámetros técnicos y estadísticos, adoptando una perspectiva interdisciplinaria que integre la ética, el Derecho y las ciencias sociales. Esta visión más amplia no solo facilita un análisis más profundo de los efectos de los algoritmos, sino que también permite identificar y abordar las formas de discriminación sistémica que estos inadvertidamente tienden a replicar o amplificar. Este enfoque es especialmente relevante en contextos donde las decisiones algorítmicas tienen implicaciones directas sobre valores democráticos y derechos ciudadanos, como señala Innerarity (2024, 2025) en su análisis de la gobernanza algorítmica.[3]

De acuerdo con Innerarity (2025), la relación entre gobernanza algorítmica y democracia representa tanto una continuidad como una ruptura con las clásicas formas de administración burocrática. Examinar la compatibilidad de estas nuevas formas de gobernanza con los valores democráticos requiere no solo considerar sus promesas de mayor objetividad en las decisiones políticas, sino también incorporar la perspectiva interdisciplinaria propuesta por Floridi (2022, 2023). La integración de la subjetividad ciudadana y la inevitabilidad de la política, incluso en entornos tecnológicos, subraya la necesidad de un diseño algorítmico que sea no solo técnicamente eficiente, sino también éticamente justo e inclusivo, capaz de responder a las expectativas democráticas sin amplificar discriminaciones sistémicas preexistentes.

---

[3] Agradezco al autor que me haya facilitado la consulta del borrador de su obra "Crítica de la razón algorítmica", actualmente en imprenta.

## 2.6. LA ÉTICA DE LA IA EN EUROPA, ¿HACIA UN MARCO REGULATORIO CONTRA LA DISCRIMINACIÓN ALGORÍTMICA?

A la hora de afrontar los sesgos y la discriminación algorítmica, la Unión Europea parece haber adoptado una postura que procura trascender las soluciones meramente tecnológicas citadas, mediante un marco regulatorio que priorice los derechos fundamentales. Este planteamiento representa un esfuerzo consciente por superar las limitaciones del «solucionismo tecnológico», reconociendo que las inequidades derivadas del uso de la inteligencia artificial no pueden resolverse únicamente desde la perspectiva computacional.

El **Reglamento General de Protección de Datos** (RGPD) de 2016 marcó un precedente importante, al introducir principios como la licitud, la equidad y la transparencia en el procesamiento de datos, además de reconocer a las personas el derecho a no ser objeto de decisiones exclusivamente automatizadas. Sin embargo, como veremos en el capítulo siguiente, sus previsiones resultaban insuficiente para dar respuesta a algunos de los problemas específicos que plantean los algoritmos de IA. En consecuencia, la Unión Europea y el Consejo de Europa iniciaron una reflexión más amplia que culminaría en una serie de iniciativas dirigidas a garantizar que estas tecnologías respeten los derechos humanos, la democracia y los valores de la Carta de los Derechos Fundamentales de la UE (CDFUE) de 2000.

El año 2018, el **Grupo Europeo de Ética de la Ciencia y de las Nuevas Tecnologías** (EGE) presentó una declaración que subrayaba la necesidad de un marco ético y jurídico internacional para regular el diseño y uso de la inteligencia artificial. Este documento destacaba principios esenciales como el respeto a la dignidad humana, la autonomía, la prevención del daño y la equidad. En particular, la declaración hacía hincapié en la necesidad de establecer mecanismos claros de rendición de cuentas, asegurando que desarrolladores, fabricantes y usuarios fueran responsables de las consecuencias derivadas de estas tecnologías. También insistía en la transparencia y la supervisión humana como salvaguardas para evitar usos indebidos y garantizar que la IA sirviera al bien común.

La Comisión Europea, siguiendo las recomendaciones del Grupo Europeo de Ética de la Ciencia y de las Nuevas Tecnologías (EGE, 2018), encargó al **Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial** (EGTAI) la elaboración de un marco ético más detallado. En 2019, este grupo publicó las *Directrices para una IA digna de confianza*, (EGTAI, 2019)

un documento basado en tres principios fundamentales: legalidad, robustez técnica y ética. Estos principios se desarrollaron en directrices específicas, como el respeto por la autonomía humana, la prevención del daño, la equidad y la explicabilidad.

En el ámbito de la equidad, las directrices diferenciaban entre justicia procedimental, que se centra en garantizar la transparencia y la corrección de decisiones erróneas, y justicia sustantiva, orientada a evitar sesgos discriminatorios en los resultados. Este planteamiento implicaba la adopción de auditorías rigurosas, trazabilidad de los datos y diversidad en los equipos humanos de desarrollo, con el objetivo de abordar la discriminación algorítmica de manera integral.

Para implementar estos principios, en julio de 2020, la Comisión Europea introdujo el marco de auditoría algorítmica ***Assessment List for Trustworthy Artificial Intelligence*** (ALTAI, 2020). Este sistema ofrece una metodología para evaluar y garantizar que los sistemas de IA cumplan con los principios éticos establecidos. ALTAI permite identificar y corregir sesgos en los datos y algoritmos, al tiempo que promueve la validación continua y las auditorías independientes. El objetivo consiste en prevenir desigualdades y garantizar que los sistemas de IA sean inclusivos, accesibles y no perpetúen discriminaciones estructurales.

Por su parte, el ***Libro Blanco sobre Inteligencia Artificial,*** presentado en el mismo año, consolidó la estrategia regulatoria de la Unión Europea. Este documento proponía un marco normativo basado en el riesgo, clasificando los sistemas de IA en función de su impacto potencial y estableciendo requisitos más estrictos para aplicaciones de alto riesgo. Estas medidas incluyen la evaluación de riesgos, la supervisión humana y la trazabilidad de los resultados, además de la creación de la Oficina Europea de IA para garantizar el cumplimiento de las normas. El *Libro Blanco* (Comisión Europea, 2020-b) señalaba la discriminación algorítmica como uno de los riesgos más graves asociados a la IA, subrayando la necesidad de adaptar las normativas existentes para abordar este problema de manera efectiva.

En su *Recomendación sobre los impactos humanos y éticos de la IA* (2020) y la *Declaración Europea sobre Derechos y Principios Digitales* (2023), la Unión Europea reafirmó su compromiso con la transparencia, la equidad y la seguridad en el desarrollo de la inteligencia artificial, las cuales deben asegurarse en un marco en el que los algoritmos estén sujetos a un escrutinio riguroso.

De forma complementaria, el Convenio del Consejo de Europa para la Protección de las Personas con respecto al Tratamiento Automático de Datos Personales (COE, Convenio 108) ha sido el único instrumento in-

ternacional jurídicamente vinculante sobre la protección de la intimidad y los datos personales. Su versión modernizada, la Convención 108+, promulgada en 2022, ayuda a preservar la dignidad humana, al tiempo que proporciona un marco propicio y seguro para la libre circulación de datos.

Cabe citar también en este marco centrado en la defensa de los derechos humanos, la **Carta de Derechos Digitales** promovida por el Gobierno de España en julio de 2021. Dirigida a los ciudadanos, administraciones públicas y empresas, la Carta establece principios para proteger derechos como la privacidad, la seguridad, la igualdad, la accesibilidad y la libertad de expresión en el ámbito digital. Aborda cuestiones específicas como la desconexión digital, la educación tecnológica y la protección frente a sesgos algorítmicos, poniendo énfasis en la supervisión y transparencia de las tecnologías emergentes, incluida la inteligencia artificial. .

Finalmente, traemos a colación los **Principios Europeos para la Ética en la Salud Digital**, adoptados en enero de 2022 por la Red de Sanidad Electrónica (2022), los cuales establecen un marco ético para guiar el desarrollo y uso de herramientas digitales en el sector sanitario, con referencias a valores humanistas, inclusión, sostenibilidad y el respeto por los derechos de los usuarios. Esta iniciativa fue promovida por el Estado francés y avalada posteriormente por la Comisión Europea.

El documento resalta que la salud digital debe complementar y optimizar la atención sanitaria presencial, garantizando que los usuarios estén informados sobre los beneficios y limitaciones de estas herramientas, así como sobre su funcionamiento. Además, declara que su intención es promover que las personas puedan personalizar sus interacciones con estas tecnologías, promoviendo su participación activa en la construcción de los marcos normativos nacionales y europeos.

En relación con los sesgos algorítmicos, el documento señala que, cuando se implemente inteligencia artificial en el ámbito de la salud digital, se deben tomar todas las medidas razonables para garantizar que los algoritmos sean explicables y estén libres de discriminación. Es necesario desarrollar sistemas que no reproduzcan desigualdades existentes en los datos ni perpetúen discriminaciones basadas en factores sensibles como género, etnia o nivel socioeconómico. El principio bioético de no maleficencia, junto con la transparencia y la trazabilidad de los procesos, se refleja en la necesidad de auditar y validar estos sistemas antes de su implementación. El citado principio debe servir para que las decisiones algorítmicas sean comprensibles tanto para los usuarios como para los profesionales de la salud, promoviendo un uso fiable y justo de estas tecnologías. El citado

marco también destaca la importancia de hacer que la salud digital sea inclusiva, lo que se traduce en herramientas accesibles para personas con discapacidades o bajos niveles de alfabetización digital, interfaces intuitivas, formación para usuarios y soporte humano cuando sea necesario. Asimismo, se subraya la necesidad de medir y mitigar los impactos ambientales de las tecnologías de salud digital, promoviendo la reutilización y el reciclaje de los equipos.

El impacto normativo del conjunto de estas declaraciones y principios ha sido significativo, ya que han influido en regulaciones europeas como el del Espacio Europeo de Datos Sanitarios y el Reglamento de IA, integrando medidas específicas para mitigar los sesgos y asegurar la accesibilidad.

Como veremos en los capítulos siguientes, el marco regulatorio de la UE, en el que se integran los citados Reglamentos, pretende ir más allá de las soluciones técnicas para abordar la discriminación algorítmica desde una perspectiva ética y jurídica. La transparencia, la trazabilidad y la supervisión independiente no solo persiguen prevenir desigualdades, sino también fomentar un debate público sobre los impactos sociales de estas tecnologías.

En este trabajo, examinaremos las medidas regulatorias adoptadas en base a los citados documentos previos para determinar si resultan suficientes en la prevención de la discriminación algorítmica en el ámbito de la salud, y valoraremos qué elementos podrían mejorarse para asegurar una protección efectiva y sostenible de los derechos fundamentales.

*Capítulo 3*

# *El Derecho antidiscriminatorio frente a la discriminación algorítmica*

La integración creciente de sistemas automatizados de inteligencia artificial en la toma de decisiones, particularmente en el ámbito sanitario, plantea retos significativos para el derecho antidiscriminatorio. Este capítulo analiza el marco normativo vigente en materia de igualdad y no discriminación, valorando su capacidad para responder a los problemas específicos que surgen con la discriminación algorítmica. Tomando como referencia el marco normativo europeo y español, se examinan tanto los avances alcanzados como las limitaciones que presentan frente a las nuevas dinámicas generadas por los algoritmos.

Entre las limitaciones más acusadas destacan los obstáculos para acceder a pruebas relacionadas con decisiones automatizadas y la aparición de factores de riesgo discriminatorio que no se ajustan fácilmente a las categorías tradicionales. Aunque el Derecho antidiscriminatorio ha evolucionado para abarcar nuevos contextos y categorías protegidas, las particularidades de los sistemas algorítmicos —opacidad, sesgos en los datos y la interrelación de múltiples factores— requieren una adaptación normativa que trascienda los enfoques tradicionales.

El análisis crítico del marco actual evidencia la necesidad de revisar profundamente el Derecho antidiscriminatorio, con el objetivo de abordar los problemas prácticos de la discriminación algorítmica y anticipar los efectos sociales de un uso creciente de la inteligencia artificial en sectores esenciales como el sanitario.

## 3.1. NORMATIVA ANTIDISCRIMINATORIA EUROPEA

La prohibición de la discriminación se erige en pilar fundamental de la Unión Europea, como derecho consagrado en el art. 3 del Tratado de la Unión Europea (TUE) y en el art. 8 del Tratado de Funcionamiento de la Unión Europea (TFUE). De manera más específica, el art. 21 de la

Carta de los Derechos Fundamentales de la UE establece una cláusula general que prohíbe toda forma de discriminación basada en motivos como el sexo, la raza, el origen étnico, la religión o las convicciones ideológicas, reforzando así el compromiso de la Unión con la igualdad.

Este principio encuentra desarrollo legislativo en un conjunto de Directivas que delimitan el marco antidiscriminatorio en diferentes ámbitos. La Directiva 2000/43/CE aborda la aplicación del principio de igualdad de trato entre las personas independientemente de su origen racial o étnico. Por su parte, la Directiva 2000/78/CE establece un marco general de igualdad en el empleo y la ocupación, cubriendo motivos como la religión, las convicciones, la discapacidad, la edad o la orientación sexual. Finalmente, la Directiva 2004/113/CE extiende el principio de igualdad de trato entre hombres y mujeres al acceso y suministro de bienes y servicios; una disposición especialmente relevante en el ámbito de la salud. Estas Directivas pretenden abordar de manera integral la discriminación en sus diversas formas, aunque, como se ha dicho más arriba, presentan limitaciones importantes a la hora de abordar nuevas formas de discriminación como es la discriminación algorítmica.

En términos de acceso a la justicia, la Directiva 97/80/CE, sobre la carga de la prueba en los casos de discriminación por razón de sexo, establece una inversión parcial de la carga de la prueba en el ámbito del Derecho antidiscriminatorio. Esta norma fue derogada y sustituida por la Directiva 2006/54/CE, pero, el privilegio procesal se mantiene como un referente normativo aplicable a otros tipos de discriminación más allá del ámbito de la igualdad de género. El art. 8 de la Directiva 2000/43/CE, el art. 10 de la Directiva 2000/78/CE, el art. 19 de la Directiva 2006/54/CE y el art. 9 de la Directiva 2004/113/CE, obligan al demandado a demostrar que no ha habido discriminación cuando el demandante presenta indicios razonables de su existencia, equilibrando así la asimetría informativa que frecuentemente beneficia a quienes ejercen poder en contextos discriminatorios.

Este beneficio parte de la idea de que, en situaciones de desigualdad estructural, exigir a la víctima una prueba concluyente puede ser desproporcionado y frustrar la efectividad de los derechos fundamentales (De Schutter, 2010), aunque la persona demandante debe aportar indicios suficientes para establecer una presunción razonable de discriminación (Fredman, 2011). Lo anterior implica aportar indicios de que, de ser ciertos, revelarían una diferencia de trato no justificada. Una vez cumplido este requisito, la carga de la prueba se transfiere al demandado, quien debe acreditar que la decisión o conducta controvertida se basó en criterios legítimos y no discriminatorios.

## 3.2. EL DERECHO ANTIDISCRIMINATORIO EN ESPAÑA

El marco normativo español en materia de igualdad y no discriminación se desarrolla a partir de las citadas Directivas europeas, incorporando al ordenamiento interno disposiciones específicas que refuerzan los derechos de las víctimas y adaptan las medidas a la realidad nacional.

En primer lugar, el art. 30 de la Ley 15/2022, integral para la igualdad de trato y la no discriminación, señala que cuando la parte actora aporte indicios fundados de discriminación, corresponderá al demandado justificar objetiva y razonablemente sus decisiones y demostrar su proporcionalidad. La Ley antidiscriminatoria se fundamenta en que las pruebas suelen estar en manos del presunto discriminador, que a menudo opera de manera indirecta, lo que dificulta que las víctimas demuestren un trato injusto basado en un motivo protegido. Además, la norma amplía los motivos de discriminación previstos en las Directivas europeas, incorporando categorías como la orientación e identidad sexual, la expresión de género, la enfermedad, el estado serológico, la predisposición genética, la lengua, la situación socioeconómica y cualquier otra circunstancia personal o social (art. 2.1). Asimismo, incluye conceptos innovadores como la discriminación múltiple e interseccional, así como la inducción, la orden o la instrucción de discriminar. Las incorporaciones referidas muestran una evolución del Derecho antidiscriminatorio español hacia una perspectiva más inclusiva y ajustada a las necesidades y realidades actuales.

Además, la ley aborda de manera específica la discriminación en el ámbito sanitario, obligando a las administraciones sanitarias a garantizar la igualdad de trato y la no discriminación en la atención sanitaria, incluso cuando se empleen sistemas automatizados o de inteligencia artificial en la toma de decisiones. En concreto, el art. 15, titulado «Derecho a la igualdad de trato y no discriminación en la atención sanitaria», dispone que las administraciones sanitarias, en el ámbito de sus competencias, garantizarán la ausencia de cualquier forma de discriminación en el acceso a los servicios y en las prestaciones sanitarias por razón de cualquiera de las causas previstas en esta ley.

> *Artículo 15. Derecho a la igualdad de trato y no discriminación en la atención sanitaria.*
>
> *1. Las administraciones sanitarias, en el ámbito de sus competencias, garantizarán la ausencia de cualquier forma de discriminación en el acceso a los servicios y en las prestaciones sanitarias por razón de cualquiera de las causas previstas en esta ley.*
>
> *2. Nadie podrá ser excluido de un tratamiento sanitario o protocolo de actuación sanitaria por la concurrencia de una discapacidad, por encontrarse en*

*situación de sinhogarismo, por la edad, por sexo o por enfermedades preexistentes o intercurrentes, salvo que razones médicas debidamente acreditadas así lo justifiquen.*

*3. Las administraciones sanitarias promoverán acciones destinadas a aquellos grupos de población que presenten necesidades sanitarias específicas, tales como las personas mayores, menores de edad, con discapacidad, pertenecientes al colectivo LGTBI, que padezcan enfermedades mentales, crónicas, raras, degenerativas o en fase terminal, síndromes incapacitantes, portadoras de virus, víctimas de maltrato, personas en situación de sinhogarismo, con problemas de drogodependencia, minorías étnicas, entre otros, y, en general, personas pertenecientes a grupos en riesgo de exclusión y situación de sinhogarismo con el fin de asegurar un efectivo acceso y disfrute de los servicios sanitarios de acuerdo con sus necesidades.*

*4. Las administraciones sanitarias, en el ámbito de sus competencias, desarrollarán acciones para la igualdad de trato y la prevención de la discriminación, que podrán consistir en el desarrollo de planes y programas de adecuación sanitarios.*

*5. En los planes y programas a los que hace referencia el apartado anterior se pondrá especial énfasis en las necesidades en materia de salud específicas de las mujeres, como la salud sexual y reproductiva, entre otras.*

*6. Nadie podrá ser apartado o suspendido de su turno de atención sanitaria básica o especializada en condiciones de igualdad, ni ser excluido de un tratamiento sanitario por ausencia de acreditación documental o de tiempo mínimo de estancia demostrable.*

Por su parte, el art. 23 de la citada norma, cuya rúbrica reza: «Inteligencia Artificial y mecanismos de toma de decisión automatizados», establece específicamente que las administraciones públicas y las entidades privadas que utilicen inteligencia artificial o mecanismos automatizados para la toma de decisiones que afecten de manera significativa a las personas deberán garantizar que dichos sistemas no generen discriminación.

La Ley 15/2022 prohíbe expresamente que el acceso a los servicios sanitarios o las prestaciones se vea condicionado por características como la discapacidad, el estado socioeconómico, el sexo o las enfermedades preexistentes, una obligación que se extiende incluso al uso de inteligencia artificial en la toma de decisiones. Este aspecto resulta especialmente relevante en un contexto donde los algoritmos, como apuntamos más arriba, tienden a reproducir los sesgos inherentes a los datos utilizados en su entrenamiento. La necesidad de garantizar la igualdad material en el acceso a los servicios sanitarios exige que estos sistemas estén diseñados para mitigar tales sesgos, en lugar de perpetuarlos.

Asimismo, la normativa subraya la importancia de atender las necesidades concretas de grupos vulnerables. El mandato legal de promover acciones específicas para colectivos como las personas mayores, las minorías

étnicas o quienes padecen enfermedades raras, obligando a una actuación que trasciende la igualdad formal. La norma insta a que los sistemas automatizados sirvan como herramientas para reducir desigualdades existentes, alineándose con un modelo más inclusivo que priorice las necesidades particulares de estos grupos (Morondo et al., 2022). En este sentido, las administraciones sanitarias deben diseñar sistemas automatizados capaces de identificar estas particularidades y responder adecuadamente a ellas, lo que requiere que los algoritmos sean entrenados con conjuntos de datos representativos y desagregados, evitando generalizaciones que ignoren las particularidades de estos colectivos.

En esta línea, la supervisión de los sistemas de inteligencia artificial mencionada en el art. 23 de la Ley 15/2022 establece un estándar que demanda evaluaciones rigurosas de impacto algorítmico antes de que los sistemas sean implementados. La obligación de diseñar planes y programas destinados a prevenir la discriminación exige no solo un monitoreo constante de los sistemas empleados, sino también la implementación de medidas correctivas cuando sea necesario. Estas evaluaciones deben dirigirse a identificar posibles sesgos con antelación, reforzando la legitimidad de las decisiones tomadas, garantizando que estas sean justificables en términos éticos y legales.

La normativa también incide en la responsabilidad de las administraciones sanitarias de rendir cuentas, lo cual se traduce en la necesidad de actuar con transparencia y de establecer vías claras y accesibles de reclamación para los usuarios que consideren haber sido discriminados y ofrecen a las personas afectadas una oportunidad real de obtener reparación frente a decisiones injustas.

A anotar que la Ley 19/2013, de acceso a la información pública y buen gobierno, garantiza el acceso a los datos generados por la administración, incluidos aquellos relacionados con decisiones automatizadas que afectan a los derechos de la ciudadanía. Según el art. 13 de la Ley 19/2013, la información sobre algoritmos públicos, en la mayoría de los casos, se considera información pública ("contenidos […] cualquiera que sea su formato o soporte […] elaborados o adquiridos en el ejercicio de sus funciones"), a la que es posible acceder a través del derecho de acceso a la información pública. No obstante, Cotino Hueso (2024-a, 2024-b) señala que el acceso a los algoritmos públicos es desigual y, en muchos casos, depende de factores arbitrarios.

A nivel autonómico, la Ley 1/2022 de la Comunidad Valenciana avanza en la transparencia algorítmica, exigiendo que se publiquen descripciones

comprensibles del diseño y funcionamiento de los algoritmos utilizados en procedimientos administrativos (Boix Palop, 2022).

Aunque el marco legal español establece un planteamiento sólido en teoría, su aplicación efectiva depende de una voluntad política real y de los recursos necesarios para llevarlo a cabo (Morondo et al., 2022). En síntesis, este marco exige un compromiso sustancial para garantizar que la inteligencia artificial y los sistemas automatizados de las administraciones sanitarias se utilicen de manera justa y no discriminatoria. Este planteamiento requiere una adaptación de los procedimientos administrativos, en línea con las exigencias de la Ley 15/2022, y con las demandas del nuevo marco normativo que se analizará en los capítulos siguientes.

### 3.3. LA REPARACIÓN DE LOS DAÑOS POR DISCRIMINACIÓN ALGORÍTMICA EN EL DERECHO ESPAÑOL

En el ámbito sanitario, los daños más comunes derivados de decisiones discriminatorias basadas en sistemas de inteligencia artificial incluyen perjuicios morales, económicos y físicos. El daño moral derivado de un tratamiento discriminatorio entraña una dimensión especial por la lesión a la dignidad de la víctima. Este perjuicio se vincula a emociones negativas como zozobra, ansiedad o tristeza, reconocidas tradicionalmente en la jurisprudencia española (LO 1/1982). Así como el daño moral se presume, otros daños, como las pérdidas económicas o físicas, requieren prueba específica para ser indemnizables, a menos que se configure una presunción más amplia, como ha propuesto parte de la doctrina jurídica.

En cuanto a la cuantificación, la Ley 15/2022 introduce criterios como la gravedad de la lesión, la difusión del acto discriminatorio y la interacción de varias causas de discriminación, lo que podría justificar una compensación mayor en casos de discriminación múltiple o interseccional. Aunque algunos autores critican que la presunción de daño moral establecida en esta ley pueda interpretarse como *iuris et de iure*, alterando la función compensatoria del sistema de responsabilidad civil, otros defienden que la propia discriminación constituye un daño en sí mismo (Carrasco, 1993, 2019-a, 2019-b). Esta tesis, aunque innovadora, plantea retos conceptuales y prácticos respecto a la delimitación del alcance del «daño en sí», como veremos más adelante.

En cuanto al daño indemnizable en el ámbito de la sanidad, los perjuicios más frecuentes en estos casos suelen ser los morales en sentido estricto y los económicos puros, ya que las decisiones discriminatorias adoptadas

por los sistemas de inteligencia artificial pueden traducirse en pérdidas de oportunidades, como la imposibilidad de acceder a un determinado tratamiento o la obtención de un diagnóstico erróneo derivado de datos sesgados. Aunque el daño moral es especialmente característico en estos casos por la lesión a la dignidad de la víctima que implica toda conducta discriminatoria, se consideran indemnizables también los daños físicos que se deriven de la decisión discriminatoria que, por ejemplo, haya retrasado el diagnóstico y el tratamiento de la enfermedad. Sin embargo, como se ha dicho, salvo el daño moral, todos los demás perjuicios deberán probarse para que proceda su resarcimiento.

A lo anterior se añade que el legislador español proporciona una serie de criterios que deberán tenerse en cuenta para su cuantificación. Desempeñan un papel destacable en la reparación de los daños ocasionados por sistemas de inteligencia artificial que ya contienen criterios de valoración del daño tanto el art. 9.3 de la Ley Orgánica 1/1982, de 5 de mayo, de protección civil del derecho al honor, a la intimidad personal y familiar y a la propia imagen, como el art. 27 de la Ley 15/2022, que se refiere a factores de riesgo como las circunstancias del caso, la gravedad de la lesión efectivamente producida, la difusión o audiencia del medio a través del que se haya producido o la concurrencia o interacción de varias causas de discriminación. En este sentido, la referencia a la concurrencia o interacción de varias causas de discriminación permite afirmar que los casos de discriminación múltiple o interseccional merecen un mayor reproche que se traducirá en una indemnización mayor a la víctima por los perjuicios ocasionados.

El art. 27 de esta Ley, después de declarar que la persona que cause discriminación debe reparar el daño causado, establece que: "*[a]creditada la discriminación se presumirá la existencia de daño moral, que se valorará atendiendo a las circunstancias del caso, a la concurrencia o interacción de varias causas previstas en la ley y a la gravedad de la lesión efectivamente producida, para lo que se tendrá en cuenta, en su caso, la difusión o audiencia del medio a través del que se haya producido*".

¿Cuáles son esos daños morales presumidos por el legislador en la Ley 15/2022? Atendiendo a la interpretación habitual de su legislación inspiradora —la LO 1/1982—, dentro de la tipología de los daños morales, los que se presumen son los denominados daños morales puros, vinculados a la noción de *pretium doloris*, y que se identifican con sensaciones y emociones humanas negativas provocadas por el evento dañoso como la zozobra, la ansiedad, el dolor, la inquietud, la tristeza, el miedo, la aflicción o la an-

gustia. Así pues, lo que se estaría presumiendo en la Ley 15/2022, es que la conducta constitutiva de discriminación ha producido alguno o varios de esos sentimientos o emociones negativos en la víctima.

Como se ha adelantado más arriba, salvo los morales, el resto de los daños no gozarán de la presunción ni de la inversión de la carga de la prueba. En virtud del referido art. 27 *in fine*, una vez acreditada la discriminación se presumirá la existencia de daño moral, que se valorará atendiendo a las circunstancias del caso, a la concurrencia o interacción de varias causas de discriminación previstas en la ley y a la gravedad de la lesión efectivamente producida, para lo que se tendrá en cuenta, en su caso, la difusión o audiencia del medio a través del que se haya producido.

La calificación de la regla del art. 27.1 *in fine* como una presunción *iuris et de iure* es criticada por la doctrina diversos motivos. En primer lugar, a falta de una expresión clara en este sentido por parte del legislador, las presunciones deberían calificarse primariamente como *iuris tantum*, según el art. 385.3 LEC. Por otra parte, la presunción *iuris et de iure* altera la función compensatoria del sistema de responsabilidad civil para transformarla en una de punitiva. Se ha dicho también que la reparación de las consecuencias de una discriminación no ha de pasar necesariamente por la utilización del remedio indemnizatorio; otros remedios de conducta pueden ser suficientes para eliminar todas las consecuencias de un acto discriminatorio y restaurar la situación previa a este. la concurrencia o interacción de varias causas de discriminación previstas en la ley como factor de valoración del daño (Barba, 2023). Esta redacción puede llevar a los tribunales a entender que también esta presunción ha de calificarse como *iuris et de iure* y no permitir al reclamado aportar prueba de la falta de causación de daños a pesar de haber incurrido en una conducta discriminatoria. Estas críticas son trasladables *mutatis mutandis* a la presunción establecida en el art. 9.3 de la LO 1/1982.

En vista de los problemas que plantea la presunción prevista en la regulación española, acaso debería revisarse la consideración de esta regla y configurarla como presunción *iuris tantum* (Barba, 2023). Las reclamaciones que pueden llegar a plantearse en relación con modelos fundacionales y sistemas de inteligencia artificial generativa –dado el carácter disperso y masivo de los daños potenciales– podrán ser un buen banco de pruebas para que los tribunales revisen la imposición automática de consecuencias indemnizatorias por intromisiones ilegítimas en los derechos al honor, intimidad y propia imagen.

La posibilidad de interpretar el daño presumido en el art. 9.3 de la LO 1/1982 como un daño en sí mismo ("*injury as such*") ha sido analizada por Carrasco Perera (Carrasco, 1993, 2019-a, 2019-b). Este autor subraya el carácter *iuris et de iure* atribuido a esta presunción y la naturaleza moral de los daños implícitos en la norma. Según esta perspectiva, podría entenderse que el daño y la discriminación (o la intromisión ilegítima) se identifican como una misma entidad, lo que conferiría pleno sentido a la presunción legal. En este marco, el legislador parecería considerar que el simple hecho de discriminar constituye en sí mismo un daño para quienes lo sufren, con independencia de las consecuencias emocionales (como ansiedad, dolor o tristeza) o económicas que también puedan derivarse. Este planteamiento se alinea con los criterios de valoración establecidos en la legislación, los cuales parecen centrarse en medir la intensidad de la lesión al derecho afectado, más que las emociones o sentimientos de la víctima.

En nuestra opinión, la mención al daño moral en la Ley 15/2022 no contradice la interpretación planteada, ya que los "daños en sí" tienen un carácter claramente extrapatrimonial. Esta interpretación se alinea con el propósito principal tanto de la LO 1/1982 como de la Ley 15/2022, que es reconocer la vulneración del derecho, priorizando este reconocimiento sobre la compensación material al afectado.

No obstante, la adopción de esta interpretación plantea nuevas cuestiones, especialmente en el plano conceptual. La consistencia misma de la noción de "daño en sí" es objeto de debate. Este concepto guarda cierta afinidad con los daños fisiológicos o anatómico-funcionales reconocidos en el baremo del Anexo del Real Decreto Legislativo 8/2004, de 29 de octubre, sobre Responsabilidad Civil y Seguro en la Circulación de Vehículos a Motor, los cuales conforman el perjuicio personal básico. Sin embargo, los daños a la integridad psicofísica tienen una manifestación física o psíquica que no está presente en casos como los daños al honor, a la intimidad o al derecho a no ser discriminado. Este último, en particular, al carecer de una dimensión tangible, sugiere que se está vinculando el concepto de daño con el de antijuridicidad, o incluso con la determinación de los "intereses protegidos", como establece el PETL 2:102.

El planteamiento del "daño en sí" implica pasar de la premisa de que el daño deriva de la lesión de un derecho o interés legítimo, a sostener que ciertas lesiones de derechos o intereses legítimos, en cuanto tales, constituyen daños. Sin embargo, esta tesis plantea problemas prácticos significativos (Martín Casals, 2011). Una vez aceptada la viabilidad de este tipo de daño, surge la necesidad de delimitar su ámbito de aplica-

ción, y determinar si debe restringirse exclusivamente a los casos en los que el legislador lo contemple expresamente, como algunos defienden, o si podría extenderse a todas las lesiones de derechos de la personalidad, conforme al planteamiento del DCFR, o incluso a otros derechos subjetivos. Tal delimitación requiere, a su vez, una reflexión acerca de los criterios que fundamentan cada postura, lo que implica abordar un problema de interpretación jurídica y aplicación práctica que está lejos de estar resuelto.

### 3.4. LAS LIMITACIONES DEL DERECHO ANTIDISCRIMINATORIO PARA AFRONTAR LA DISCRIMINACIÓN ALGORÍTMICA

El marco normativo europeo y español en materia de igualdad y no discriminación ha avanzado significativamente en la protección contra formas tradicionales de discriminación, tal como se ha visto en los apartados anteriores. Sin embargo, tanto la doctrina como la jurisprudencia han señalado importantes limitaciones al enfrentarse a la complejidad de la discriminación algorítmica, especialmente en sectores como la sanidad. Según Schiek (2016-a, 2016-b), las Directivas europeas están orientadas principalmente hacia la igualdad formal, lo que dificulta abordar las dinámicas complejas y menos evidentes que emergen en entornos tecnológicos. Por su parte, Fredman (2011, 2026) subraya que la normativa europea presenta debilidades en el tratamiento de la discriminación indirecta y de la interseccionalidad, elementos clave en el contexto algorítmico. Estas carencias se agravan en el ámbito sanitario, donde los sistemas automatizados tienen un impacto directo en derechos fundamentales.

El Tribunal de Justicia de la Unión Europea (TJUE) ha proporcionado criterios relevantes que enriquecen la interpretación del marco normativo y permiten una mejor adaptación de la discriminación algorítmica. En la sentencia del caso Egenberger (C-414/16, TOL6.573.837), se reconoció el efecto directo horizontal del art. 21 de la Carta de Derechos Fundamentales de la UE, permitiendo invocar la prohibición de la discriminación en las relaciones privadas. Este precedente es particularmente relevante en la sanidad, donde los servicios pueden estar mediados por relaciones contractuales entre pacientes y proveedores de servicios tecnológicos. Asimismo, en casos como Meister (C-415/10, TOL9.916.431) y Danfoss (C-109/88) la jurisprudencia del TJUE refuerza la inversión de la carga de la prueba en situaciones de opacidad; un principio que podría extrapolarse a los sistemas algorítmicos empleados en contextos sanitarios, exigiendo a

las empresas o administraciones justificar que sus decisiones automatizadas no son discriminatorias.

A nivel nacional, hemos visto que la Ley 15/2022 incorpora disposiciones significativas para abordar los retos de la discriminación algorítmica. Su art. 23 establece la obligación de garantizar la transparencia y la rendición de cuentas en el uso de inteligencia artificial, con especial atención a sectores críticos como la sanidad. Sin embargo, persisten deficiencias en la supervisión de estos sistemas, agravadas por la falta de estándares claros sobre acceso y publicidad de los algoritmos (Cotino Hueso, 2024-a). En el caso Bosco, la Audiencia Nacional (SAN 2013/2024, TOL10.005.554) negó el acceso al código fuente de un sistema algorítmico público, argumentando riesgos de seguridad informática, dejando sin resolver aspectos esenciales sobre la transparencia y el control ciudadano en el uso de algoritmos financiados con fondos públicos.

Hemos visto que la discriminación algorítmica en el ámbito sanitario se manifiesta principalmente como discriminación indirecta, derivada de sesgos en los datos o en el diseño de los algoritmos. Estos sesgos generan resultados desproporcionados para determinados grupos protegidos, como ocurrió en el caso Deliveroo, resuelto por el Tribunal de Bolonia el 31 de diciembre 2020 (n.º 2949/2020), donde un sistema algorítmico penalizaba injustamente a trabajadores que ejercían derechos legítimos. Este fallo subraya la necesidad de que los sistemas algorítmicos integren garantías para evitar decisiones discriminatorias, una exigencia igualmente aplicable al ámbito sanitario.

Por último, el Derecho antidiscriminatorio presenta carencias notorias en la prueba del daño en contextos algorítmicos debido a la opacidad de estas tecnologías. Aunque, como sabemos, la normativa europea contempla la inversión de la carga de la prueba, esta medida tiene un alcance limitado frente a la complejidad técnica de los sistemas automatizados. La jurisprudencia europea y nacional señalan la necesidad de desarrollar herramientas específicas, como auditorías algorítmicas y acceso a datos desagregados, para garantizar una protección efectiva contra la discriminación algorítmica en sectores críticos como la sanidad.

En conclusión, aunque el marco normativo europeo y el español constituyen una base sólida, su aplicabilidad al fenómeno de la discriminación algorítmica requiere ajustes significativos. La aplicación decidida de la inversión de la carga de la prueba, la integración de estándares de transparencia y explicabilidad, junto con el reconocimiento de un derecho específico a la no discriminación algorítmica, permitiría mitigar los sesgos tecnológicos y garantizar la igualdad de trato en un entorno de creciente automatización.

*Capítulo 4*

# *El Reglamento General de Protección de Datos (RGPD)*

En los capítulos anteriores se ha destacado que la discriminación algorítmica representa una preocupación significativa en el desarrollo de productos sanitarios basados en inteligencia artificial, especialmente en decisiones críticas como diagnósticos médicos y evaluaciones de seguros. Aunque el Reglamento General de Protección de Datos (RGPD) de 2016 fue promulgado en un contexto en el que la inteligencia artificial aún no ocupaba un lugar central en el debate regulatorio, muchas de sus disposiciones han demostrado ser relevantes para abordar algunos de estos riesgos. Entre ellas, destacan las normas relacionadas con decisiones automatizadas, transparencia y explicabilidad, así como los principios generales de calidad y precisión de los datos, que proporcionan un marco inicial valioso.

Sin embargo, la capacidad del RGPD para regular de manera integral la inteligencia artificial en productos sanitarios presenta limitaciones evidentes. Entre estas se encuentran la falta de requisitos específicos para el diseño y desarrollo de algoritmos, las tensiones entre principios como la minimización de datos y la exactitud, así como las restricciones impuestas por otras normas como las de propiedad intelectual, que dificultan la transparencia de los sistemas algorítmicos. Estas carencias evidenciaron la necesidad de un marco normativo complementario que respondiera específicamente a las particularidades de la inteligencia artificial. En este sentido, el Reglamento de Inteligencia Artificial, aprobado en 2024, y que se analizará en el capítulo siguiente, tiene como objetivo llenar los vacíos normativos que el RGPD no logró abordar.

Este capítulo se centrará en examinar el modo en que el RGPD proporciona una base sólida en aspectos como la calidad de los datos y la protección frente a decisiones automatizadas, así como en destacar las limitaciones de estas disposiciones para abordar de forma completa los riesgos específicos de la discriminación algorítmica en el ámbito sanitario. En primer lugar, abordamos la calidad de los datos, considerando cómo la exactitud, la minimización y la pertinencia de los datos personales pueden afectar la equidad algorítmica, especialmente en lo que respecta a la pre-

vención de sesgos y la protección de grupos vulnerables. A continuación, analizamos las evaluaciones de impacto y la responsabilidad proactiva, conceptos clave que refuerzan la necesidad de una protección de datos desde el diseño de los sistemas de IA, así como la implementación de medidas preventivas para mitigar riesgos potenciales.

Seguidamente, el capítulo se adentra en la transparencia, explicabilidad e interpretabilidad de los algoritmos, aspectos esenciales para garantizar que tanto los profesionales de la salud como los pacientes comprendan las decisiones automatizadas que pueden afectar significativamente su bienestar. Finalmente, se examina el artículo 22 del RGPD, que regula las decisiones automatizadas y la supervisión humana, con especial atención al impacto de la jurisprudencia reciente, como el caso SCHUFA, y su aplicación en el ámbito sanitario.

A partir de esta base, se explorará la interacción entre el RGPD y el RIA, con el objetivo de evaluar la posibilidad de construir un entorno regulatorio coherente y efectivo que garantice la equidad y la transparencia en el uso de inteligencia artificial en la sanidad.

## 4.1. LA CALIDAD DE LOS DATOS

Uno de los factores determinantes en la discriminación algorítmica consiste en la calidad de los datos utilizados para entrenar los algoritmos. Hemos insistido en que, en el contexto de los productos sanitarios, los datos sesgados pueden perpetuar desigualdades históricas, afectando desproporcionadamente a grupos vulnerables. Pues bien, aunque el Reglamento General de Protección de Datos no aborda directamente los sesgos, establece principios que afectan la calidad de los datos, proporcionando un marco que puede ayudar a mitigar estos problemas.

El principio de **licitud, lealtad y transparencia** —establecido en el art. 5.1.a del RGPD— exige que los datos sean procesados de manera justa y transparente, garantizando que los interesados comprendan cómo se utilizan sus datos. La lealtad en el tratamiento de datos implica evitar cualquier práctica que pueda ser engañosa para los individuos o que los coloque en desventaja injusta debido a errores o sesgos en los sistemas algorítmicos. Este principio es esencial para asegurar que los datos utilizados en productos sanitarios con IA sean manejados de manera ética y conforme a la ley.

La **exactitud de datos** está regulada en el art. 5.1.d del RGPD y requiere que los datos sean exactos y pertinentes, lo cual implica eliminar información redundante o inexacta que podría generar resultados sesgados. Sin

embargo, la aplicación práctica del principio de exactitud se enfrenta a dificultades como la falta de definiciones técnicas claras sobre la calidad de los datos. A menudo, se carece de orientaciones específicas para garantizar que los datos utilizados sean inclusivos y representativos, lo que puede llevar a interpretaciones dispares que no aborden de forma adecuada la corrección de los sesgos. Más arriba hemos puesto el ejemplo de un sistema de IA para diagnósticos médicos que se entrena mayoritariamente con datos de una población específica —como hombres adultos de una determinada etnia—, en cuyo caso los diagnósticos para mujeres o minorías pueden ser menos precisos, perpetuando desigualdades en la atención sanitaria. Los Considerandos 75 y 85 del RGPD advierten de que el tratamiento incorrecto de datos puede generar discriminación y otros daños inmateriales, subrayando la importancia de un manejo cuidadoso y responsable de los datos.

El principio de **minimización** de datos, consagrado en el art. 5.1.c, establece que el tratamiento de datos personales debe limitarse a aquellos que sean estrictamente necesarios para cumplir con la finalidad específica para la cual se recaban. Este principio tiene como objetivo fundamental proteger la privacidad de los interesados al evitar la recopilación excesiva o innecesaria de información personal. Como destaca Bygrave (2014) en su análisis de los principios fundamentales de la protección de datos, la minimización no solo constituye una herramienta esencial para limitar el uso indebido de datos personales, sino que también refuerza la confianza de los interesados en el sistema de tratamiento de datos.

Sin embargo, en el contexto del desarrollo y entrenamiento de modelos de inteligencia artificial, la aplicación estricta de este principio plantea un conflicto significativo con las necesidades técnicas inherentes a estas tecnologías. Los modelos de IA, especialmente aquellos basados en aprendizaje profundo, requieren grandes volúmenes de datos diversos y representativos para su entrenamiento, con el fin de identificar patrones de manera efectiva y generalizar sus resultados en una amplia variedad de casos. Por contra, la limitación a la hora de la recopilación de datos respetuosa con el derecho a la privacidad de los interesados, puede generar sesgos significativos en los algoritmos por falta de representatividad de determinados colectivos.

La interacción entre el principio de minimización de datos y las necesidades del aprendizaje automático refleja uno de los dilemas más complejos de la regulación de la inteligencia artificial. Para resolver este conflicto, resulta imprescindible equilibrar el respeto por los derechos de privacidad

de los interesados con la necesidad de recopilar datos suficientes para garantizar la robustez y la justicia de los modelos algorítmicos.

La literatura especializada ha explorado diferentes estrategias para abordar el meritado equilibrio. Mantelero (2018) señala que este dilema pone de relieve la necesidad de adoptar interpretaciones flexibles y contextuales del principio de minimización, especialmente cuando el objetivo del tratamiento de datos es prevenir sesgos y promover la equidad en los sistemas de IA. Hildebrandt (2015) propone que el principio de minimización puede adaptarse mediante el uso de técnicas como la anonimización y la seudonimización, lo que permite a los desarrolladores trabajar con datos amplios sin comprometer la privacidad individual. Recordemos que, si los datos con los que se trabaja son anónimos, no es necesario aplicar los principios y reconocer los derechos contemplados en el RGPD (Nicolas, 2019). No obstante, la anonimización es considerada por el Comité Europeo de Protección de Datos un proceso dinámico que debe tener en cuenta las capacidades computacionales crecientes que hacen que datos que hoy están anonimizados, puedan ser nominativos en el futuro próximo (EDPB, 2024).

Por otro lado, Veale y Binns (2017) destacan la importancia de realizar auditorías regulares tanto de los datos utilizados como de los resultados generados por los modelos, con el fin de identificar y corregir posibles sesgos o usos indebidos de la información personal. En nuestra opinión, la tensión entre ambos principios requiere un marco regulatorio que no solo contemple el respeto a la privacidad y la protección de datos, sino que también facilite el desarrollo de sistemas de IA éticos y justos que puedan operar en beneficio de una sociedad más equitativa y eficiente. Sobre esta cuestión volveremos con ocasión del análisis del nuevo Reglamento de Inteligencia Artificial.

En definitiva, podemos concluir que, aunque los principios generales de calidad de los datos del RGPD proporcionan un marco normativo valioso para prevenir discriminaciones algorítmicas en el ámbito sanitario, su implementación efectiva requiere adaptaciones específicas a las particularidades de la inteligencia artificial. Para garantizar una aplicación justa y equitativa de las tecnologías de IA en sanidad, el equilibrio entre la minimización de datos y la necesidad de recopilar datos diversos y representativos, así como la responsabilidad proactiva en la evaluación de sesgos, son aspectos críticos que deben ser incorporados a la regulación específica que ofrece el nuevo Reglamento de Inteligencia Artificial, al que nos referiremos en los capítulos siguientes.

## 4.2. EVALUACIONES DE IMPACTO, RESPONSABILIDAD PROACTIVA Y PROTECCIÓN DESDE EL DISEÑO

La **responsabilidad proactiva y protección desde el diseño**, tal como se establece en los arts. 25 y 35 del RGPD, obliga a los responsables a implementar medidas para garantizar la calidad de los datos desde la concepción de los sistemas. Estas medidas incluyen la evaluación de posibles sesgos inherentes a los conjuntos de datos utilizados. Esta valoración de los derechos y libertades de los individuos debe hacerse desde el diseño del algoritmo, particularmente en aquellos grupos que pueden estar en riesgo de discriminación debido a características sensibles como la raza, el género o la condición socioeconómica.

El Reglamento General de Protección de Datos establece la obligatoriedad de realizar **evaluaciones de impacto** en la protección de datos (EIPD) en casos donde el tratamiento pueda entrañar un alto riesgo para los derechos y libertades de las personas (art. 35). El propósito de estas evaluaciones no es solo identificar riesgos, sino también implementar soluciones específicas que minimicen los impactos negativos como la discriminación por sesgos.

Además, las evaluaciones deben considerar los posibles efectos acumulativos de los diagnósticos automatizados. Mediante la evaluación de impacto no solo se trata de evaluar los riesgos inmediatos, sino también de entender cómo las decisiones automatizadas pueden influir en la confianza pública en los sistemas de salud basados en IA.

En nuestra opinión, las evaluaciones de impacto y las medidas preventivas siguen siendo instrumentos válidos para garantizar que los productos sanitarios basados en IA operen de manera justa y transparente. La efectividad de estas herramientas depende de su capacidad para adaptarse a los cambios operados por los nuevos sistemas algorítmicos, considerando no solo los riesgos inmediatos, sino también los impactos acumulativos a largo plazo.

## 4.3. TRANSPARENCIA, EXPLICABILIDAD E INTERPRETABILIDAD DE LOS ALGORITMOS DE SALUD

El **principio de transparencia**, consagrado en el art. 5.1.a del RGPD, requiere que los interesados tengan **acceso a información clara y comprensible sobre cómo se procesan sus datos**.

En el contexto de las decisiones automatizadas, el RGPD refuerza el principio de transparencia mediante disposiciones que obligan a proporcionar información significativa sobre la lógica detrás de los algoritmos. Aunque no establece un derecho explícito a recibir explicaciones detalladas, los arts. 13.2.f, 14.2.g y 15.1.h imponen a los responsables del tratamiento el deber de ofrecer claridad sobre cómo se toman las decisiones automatizadas y cuáles son sus posibles consecuencias para los interesados. Estas obligaciones han sido interpretadas por el TJUE en la sentencia SCHUFA (TOL9.882.990) a la que haremos referencia más tarde de forma extensa.

Los algoritmos avanzados, como los basados en aprendizaje profundo, plantean retos significativos en cuanto a la capacidad de explicar sus decisiones debido a su intrincada estructura y funcionamiento, lo que a menudo los convierte en una «caja negra» incluso para especialistas. Esta limitación pone de relieve la necesidad de que el desarrollo e implementación de sistemas de inteligencia artificial se guíe por principios como la transparencia, la explicabilidad y la interpretabilidad, esenciales para garantizar la confianza y la protección de los derechos de los usuarios.

La literatura académica, particularmente los trabajos de Wachter, Mittelstadt y Floridi (2017), ha interpretado las disposiciones del RGPD como un marco que, aunque limitado, abre la puerta a la integración del principio de explicabilidad en el ámbito de la inteligencia artificial. Sin embargo, el RGPD no aborda de manera integral las particularidades de los sistemas de IA avanzados. Como veremos en el próximo capítulo, estos principios deben ser complementados con los requisitos específicos que el Reglamento de Inteligencia Artificial establece para los sistemas de alto riesgo, consolidando un marco normativo más robusto para afrontar los retos éticos y técnicos asociados a estas tecnologías.

En la ciencia de datos la explicabilidad se entiende como la capacidad de un sistema o modelo algorítmico para comunicar de manera clara y comprensible cómo y por qué toma determinadas decisiones. Este concepto ha evolucionado significativamente con el desarrollo de la *Explainable Artificial Intelligence* (XAI), un conjunto de métodos y procesos diseñados para hacer que los resultados generados por los algoritmos sean comprensibles para los usuarios. Hsieh et al. (2024) subrayan que la XAI no solo tiene aplicaciones técnicas, sino que también es esencial para garantizar la seguridad del paciente y la confianza en los sistemas algorítmicos, particularmente en ámbitos como la salud, donde las decisiones pueden tener un impacto directo en la vida de las personas. Con el fin de explicar cómo se

alcanzaron ciertas conclusiones algorítmicas mientras se protegen los secretos comerciales, los científicos de datos están desarrollando soluciones XAI mediante el uso de informes automatizados que genera el propio sistema al ser interrogado sobre las razones de su decisión (Johnson, 2023).

En el ámbito sanitario, la explicabilidad implica que los algoritmos no solo deben ser precisos, sino también comprensibles para médicos, pacientes y reguladores. Por ejemplo, en el caso de un sistema diseñado para predecir el riesgo de enfermedades cardíacas, un modelo basado en XAI no se limitaría a emitir una predicción, sino que proporcionaría información sobre los factores que llevaron a dicha conclusión, como niveles de colesterol, presión arterial y edad del paciente. Este nivel de detalle permite a los médicos validar las decisiones algorítmicas, identificar posibles sesgos y adaptar el tratamiento a las necesidades específicas del paciente. Ritter et al. (2022) señalan que esta transparencia fortalece la confianza de los usuarios y facilita la integración de la inteligencia artificial en procesos médicos críticos.

Un aspecto importante de la explicabilidad es la modalidad conocida como «explicabilidad *ex post*», que cobra relevancia cuando las decisiones automatizadas generan controversia o disputas. Según Wachter et al. (2017), esta forma de explicabilidad permite proporcionar información detallada sobre los datos utilizados, la lógica aplicada y las consecuencias de una decisión, una vez que esta ha sido tomada. Por ejemplo, si un paciente cuestiona una decisión algorítmica que le ha negado acceso a un tratamiento, el sistema debería ser capaz de detallar cómo se llegó a esa conclusión, qué variables fueron consideradas y si se siguieron los principios normativos aplicables. Como señala Gil Membrado (2023), esta modalidad permite evaluar si las decisiones han sido justas y conformes con las normativas, especialmente en casos donde los sesgos o errores son difíciles de detectar en tiempo real.

La explicabilidad se dirige principalmente a los usuarios no técnicos (como médicos y pacientes), permitiéndoles comprender las recomendaciones del algoritmo sin necesidad de conocer los detalles técnicos internos. En cambio, la **interpretabilidad** es un estándar dirigido a desarrolladores y expertos técnicos, ya que les permite verificar y validar que el modelo funciona según lo esperado, reduciendo los riesgos asociados a errores o sesgos ocultos. En el ámbito sanitario, ambos conceptos son complementarios. La interpretabilidad asegura que los modelos sean técnicamente robustos y precisos, mientras que la explicabilidad facilita la aceptación ética y social de los sistemas al proporcionar a los profesionales médicos y a los

pacientes la confianza necesaria para utilizar estas tecnologías en decisiones críticas de salud (Escobar Mimbrera, 2021).

El desarrollo de la explicabilidad, en particular de la XAI, no solo responde a una necesidad técnica, sino también a exigencias éticas y legales en contextos críticos como la sanidad. Aunque el RGPD ofrece un marco inicial basado en principios como la calidad de los datos y la transparencia, su alcance es insuficiente para abordar integralmente las particularidades de los sistemas algorítmicos actuales.

El reconocimiento normativo de un derecho a la explicabilidad podría complementar este marco, garantizando que las decisiones automatizadas sean comprensibles, auditables y conformes con los principios de justicia e igualdad. Este reconocimiento no solo beneficiaría a los profesionales de la salud y a los pacientes, sino que también reforzaría la confianza en la inteligencia artificial como herramienta para mejorar la toma de decisiones en el ámbito sanitario.

En definitiva, la transparencia, la explicabilidad y la interpretabilidad son principios interdependientes que juegan un papel fundamental en la adopción ética de la inteligencia artificial en la sanidad. Mientras que la transparencia establece el marco general de información para los interesados, la explicabilidad proporciona herramientas prácticas para comprender decisiones específicas y, finalmente, la interpretabilidad asegura la robustez técnica de los modelos. La colaboración entre desarrolladores y reguladores será fundamental para superar los problemas inherentes a la implementación de estos principios, fomentando sistemas de IA fiables y respetuosos de los derechos de los pacientes.

## 4.4. DECISIONES AUTOMATIZADAS Y SUPERVISIÓN HUMANA: EL CASO SCHUFA

De forma pionera, el RGPD reguló las decisiones automatizadas a través de su art. 22, con el fin proteger a los individuos de decisiones basadas únicamente en el tratamiento automatizado de datos personales cuando estas producen efectos jurídicos significativos. Sin embargo, la aplicación de este artículo en el contexto de los diagnósticos médicos basados en inteligencia artificial está siendo objeto de debate doctrinal y jurisprudencial (Berenguer Albaladejo, 2024).

Como es sabido, el art. 22.1 del RGPD otorga a los interesados el derecho a no ser objeto de decisiones basadas únicamente en el tratamiento

automatizado, incluida la elaboración de perfiles, que produzcan efectos jurídicos o les afecten significativamente de modo similar.

La citada sentencia del TJUE en el caso SCHUFA aporta claridad al respecto, estableciendo que para que se aplique la prohibición del art. 22.1, deben concurrir las siguientes condiciones. En primer lugar, debe tratarse de una decisión que produzca efectos jurídicos en el interesado o le afecte significativamente de modo similar. Además, debe estar basada únicamente en el tratamiento automatizado, es decir que la decisión debe adoptarse sin intervención humana significativa; es decir, debe depender exclusivamente de procesos automatizados, incluida la elaboración de perfiles. Veamos a continuación con más detalle el razonamiento del Tribunal.

La sentencia del TJUE de 7 de diciembre de 2023, en el asunto SCHUFA Holding (C-634/21 TOL9.882.990), examina si la generación automatizada de un valor de probabilidad sobre la capacidad futura de una persona para cumplir con sus obligaciones de pago constituye una «decisión automatizada» en el sentido del art. 22 del RGPD.

En primer lugar, el TJUE falla que la mera elaboración de un «score» crediticio por parte de una agencia de información comercial no constituye, por sí misma, una decisión automatizada en los términos del art. 22. Sin embargo, si una entidad financiera utiliza dicha puntuación, aunque haya sido facilitada por un tercero, como base exclusiva para adoptar decisiones que afecten al interesado, como la concesión o denegación de un crédito, entonces sí se estaría ante una decisión basada únicamente en el tratamiento automatizado. En este contexto, el TJUE establece que el art. 22 del RGPD es aplicable cuando la **decisión final, que produce efectos jurídicos o afecta significativamente al interesado de modo similar, se toma sin intervención humana y se basa exclusivamente en el «score» proporcionado por la agencia** (FFJJ. 42 a 50).

La sentencia SCHUFA se alinea con interpretaciones previas que abogan por una **aplicación restrictiva del art. 22 del RGPD**, limitando su alcance a decisiones que cumplan estrictamente con los criterios mencionados. En este sentido, autores como Wachter, Mittelstadt y Floridi (2017) han argumentado que el RGPD no establece un derecho general a la explicación de decisiones automatizadas, sino que se centra en garantizar que los interesados no sean sujetos de decisiones automatizadas sin las debidas garantías.

En segundo lugar, el Tribunal entiende que el art. 22.1 del RGPD consagra un "**derecho**" (sic) **del interesado a no ser objeto de una decisión basada únicamente en el tratamiento automatizado, incluida la elaboración**

**de perfiles**. Dicha disposición establece una **prohibición de principio** cuya inobservancia no necesita ser invocada proactivamente por el interesado (FJ 52).

Por tanto, el Tribunal aclara que el RGPD contiene una prohibición general de que los individuos sean objeto de decisiones basadas exclusivamente en el tratamiento automatizado, incluido el perfilado, cuando estas produzcan efectos jurídicos o les afecten significativamente de manera similar. El Tribunal afirma además que «*la inobservancia [de la prohibición] no necesita ser invocada proactivamente por el interesado.*»

El art. 22.1 del RGPD no debe ser interpretado como un simple derecho individual que depende del ejercicio activo de la persona afectada para que sea exigible, sino como una prohibición de naturaleza objetiva y vinculante. En términos prácticos, el Tribunal pone énfasis en la obligación del responsable del tratamiento de garantizar el cumplimiento de esta norma de manera proactiva, sin que la reclamación de los derechos del interesado sea una condición necesaria para determinar la existencia de la infracción.

El razonamiento del TJUE se alinea, en nuestra opinión, con la estructura del art. 22 del RGPD, que otorga una protección reforzada frente a decisiones automatizadas que puedan afectar de manera significativa los derechos fundamentales. Según la interpretación sistemática del Reglamento, esta disposición no opera únicamente en beneficio de quienes son conscientes de sus derechos y los ejercen, sino que constituye una garantía de orden público que las autoridades de protección de datos y los tribunales pueden supervisar y aplicar de oficio. Como destaca la doctrina especializada, esta interpretación contribuye a garantizar una «protección uniforme y efectiva de los derechos fundamentales en el ámbito del tratamiento de datos personales» (Kuner, Bygrave y Docksey, 2020, p. 456).

Además, la naturaleza objetiva de la prohibición refuerza el deber de diligencia de los responsables del tratamiento. Estos deben adoptar medidas adecuadas para evitar decisiones basadas exclusivamente en sistemas automatizados que no cuenten con la intervención humana real y efectiva. La falta de acción o medidas proactivas puede ser considerada una violación del art. 22.1, independientemente de que exista una reclamación por parte del interesado. Esta jurisprudencia entronca con la doctrina del TJUE que entiende el RGPD como un instrumento no solo de protección individual, sino también de regulación estructural en el ámbito del mercado digital (Sentencias TJUE, Sentencia de 4 de julio de 2023, asunto C-252/21, Meta Platforms y otros, TOL9.908.883; la comentada sentencia de 7 de diciembre de 2023, asunto C-634/21, TOL9.882.990; sentencia de 26 de abril de

2022, asunto C-401/19, República de Polonia contra Parlamento Europeo y Consejo de la Unión Europea, TOL8.914.622).

En lo que respecta al uso de herramientas tecnológicas, el TJUE, apoyándose en el Considerando 71 del RGPD, recalca que los responsables del tratamiento deben emplear **procedimientos matemáticos o estadísticos que sean apropiados para el propósito del procesamiento automatizado** (FJ 59). Esta exigencia tiene como objetivo no solo la optimización técnica de los sistemas de IA, sino también el cumplimiento de estándares jurídicos y éticos que minimicen los riesgos inherentes a las decisiones automatizadas. En este sentido, se requiere que los modelos utilizados garanticen un alto nivel de precisión y reduzcan al mínimo posible el margen de error, evitando con ello resultados que puedan generar perjuicios indebidos o injustificados para los interesados. Se exige que los responsables cuenten con mecanismos robustos para identificar y corregir inexactitudes en los datos o en los resultados del procesamiento, con el fin de preservar la integridad de los derechos de las personas afectadas.

La obligación de utilizar procedimientos matemáticos o estadísticos adecuados adquiere particular relevancia en el ámbito de la inteligencia artificial, dado que estos sistemas, al basarse en patrones algorítmicos, pueden reproducir o amplificar sesgos preexistentes si no son diseñados e implementados con cuidado. El TJUE subraya la necesidad de **prevenir cualquier tipo de efecto discriminatorio** que pueda derivarse del uso de decisiones automatizadas, como también lo prevé el Considerando 71 del RGPD (FJ 66). La sentencia insiste en que es imperativo que los responsables adopten medidas que impidan la discriminación basada en características protegidas como la raza, el origen étnico, las opiniones políticas, la religión, las creencias, la orientación sexual, la salud o la afiliación sindical.

A estas obligaciones técnicas y sustantivas se añaden las **obligaciones de transparencia** previstas en los arts. 13.2.f y 14.2.g del RGPD, que requieren que los responsables proporcionen al interesado información significativa sobre la lógica detrás del procesamiento automatizado, así como sobre la importancia y las consecuencias previstas de dicho procesamiento. Estas disposiciones son complementadas por el art. 15.1.h, que reconoce a los interesados el derecho a acceder a esta información y, en última instancia, a obtener la intervención humana para expresar su punto de vista y cuestionar las decisiones automatizadas. El TJUE interpreta estos requisitos como una garantía reforzada frente a los riesgos inherentes al uso de sistemas automatizados en contextos donde se evalúan aspectos personales, como

el desempeño laboral, la situación económica o la conducta (Berenguer Albaladejo, 2024).

El fallo del TJUE en el caso SCHUFA pone de relieve la necesidad de integrar salvaguardas técnicas y organizativas que aseguren un tratamiento justo, transparente y no discriminatorio de los datos personales. Esta interpretación amplia refuerza el carácter autónomo y vinculante del art. 22.1 del RGPD, al establecer que la prevención de la discriminación, el empleo adecuado de procedimientos estadísticos y matemáticos, y la provisión de información significativa sobre los procesos automatizados no son meros requisitos técnicos, sino pilares fundamentales para garantizar los derechos fundamentales en la era digital. El TJUE exige que los responsables del tratamiento implementen medidas efectivas para garantizar resultados equitativos y evitar prácticas discriminatorias. El nuevo estándar de protección, supervisado por las autoridades competentes, es necesario para fortalecer la confianza y garantizar la seguridad jurídica en un contexto donde la automatización desempeña un papel cada vez más destacado.

La interpretación posibilista del Tribunal de Justicia de la Unión Europea destaca la robustez del Reglamento General de Protección de Datos en la protección de los derechos fundamentales. Sin embargo, el RGPD carece de disposiciones específicas que aborden la calidad de los datos desde una perspectiva de equidad, lo que genera un vacío normativo en este ámbito. Esta omisión resulta especialmente relevante en contextos donde las decisiones automatizadas pueden tener un impacto significativo en los derechos de los interesados.

Para abordar estas limitaciones, el Reglamento de Inteligencia Artificial complementa al RGPD mediante la incorporación de obligaciones específicas, como la evaluación de sesgos y la implementación de medidas correctivas, recogidas en los arts. 10.2.f y 10.2.g. Estas disposiciones están destinadas a mitigar riesgos técnicos asociados a los sistemas de IA, pero también sirven para prevenir discriminaciones estructurales derivadas de la utilización de datos históricos sesgados. Como veremos en el capítulo siguiente, el RIA establece un marco normativo que refuerza la equidad y minimiza los riesgos éticos en la aplicación de la inteligencia artificial.

Además, existe una tensión entre el derecho a la información de los usuarios y la necesidad de proteger la propiedad intelectual y los secretos comerciales de los desarrolladores que aflora en el caso de autos. La sentencia SCHUFA aborda directamente el equilibrio entre la explicabilidad en los sistemas algorítmicos y la protección de secretos comerciales o derechos de propiedad intelectual, determinando que las empresas están

obligadas a proporcionar a los interesados información significativa sobre los fundamentos de las decisiones automatizadas, incluso si estas son complejas o derivan de modelos avanzados. El Tribunal subraya que la protección de secretos comerciales no puede utilizarse como pretexto para negar a los usuarios el derecho a comprender cómo se han tomado decisiones que les afectan de manera significativa. No obstante, reconoce la necesidad de equilibrar la transparencia con el secreto comercial y los derechos de propiedad intelectual. En este sentido, la sentencia establece que se puede cumplir con el principio de explicabilidad sin revelar detalles técnicos específicos o el código fuente, mediante explicaciones comprensibles sobre los métodos y criterios generales utilizados.

En el contexto sanitario, esta jurisprudencia refuerza la importancia de proporcionar explicaciones claras sobre cómo los sistemas de IA llegan a conclusiones diagnósticas o terapéuticas, mientras se protege la innovación tecnológica. La sentencia sugiere que el equilibrio puede lograrse mediante la generación de informes que detallen de forma accesible los razonamientos del sistema, sin comprometer la confidencialidad técnica. En este mismo sentido se pronuncia el Reglamento de Inteligencia Artificial, como veremos.

## 4.5. LA APLICACIÓN DEL ART. 22 DEL RGPD A LA IA EN EL ÁMBITO SANITARIO

El art. 22.1 del Reglamento General de Protección de Datos regula las decisiones «basadas únicamente en el tratamiento automatizado, incluida la elaboración de perfiles», que produzcan efectos jurídicos sobre el interesado o lo afecten de modo similar. En el ámbito sanitario, surge la cuestión de si los sistemas de diagnóstico que incorporan inteligencia artificial toman decisiones de este tipo, o la norma se refiere a otro tipo de funcionalidades. La respuesta determinará si estos sistemas deben cumplir con las estrictas garantías establecidas en el RGPD, como la intervención humana significativa y la transparencia en el proceso decisorio.

Para responder a esta cuestión, es esencial analizar la funcionalidad específica de los sistemas de inteligencia artificial en el ámbito sanitario. Los **sistemas de apoyo al diagnóstico**, donde el resultado generado por la IA es evaluado por un profesional antes de tomar una decisión, no suelen considerarse decisiones automatizadas en el sentido del art. 22. En estos casos, el médico conserva el control final del diagnóstico, utilizando la salida del algoritmo como una herramienta para informar su juicio clínico.

Añádase que la Ley de Autonomía del Paciente aclara que las decisiones sobre el tratamiento debe tomarlas el paciente tras un proceso de información por parte del personal sanitario. La sentencia SCHUFA (C-634/21, TOL9.882.990), a la que nos acabamos de referir, aclara que una decisión no se considera automatizada si incluye una revisión significativa por parte de un humano.

Sin embargo, otros sistemas de IA en el ámbito sanitario sí generan decisiones automatizadas que podrían comprenderse en el ámbito del art. 22. Por ejemplo, los sistemas de triaje médico, que priorizan la atención de pacientes en función de parámetros como síntomas reportados y datos médicos; o las evaluaciones automatizadas para determinar la contratación de seguros de salud, pueden producir decisiones automatizadas con efectos significativos sobre los interesados. En estos casos, si el sistema toma decisiones sin revisión humana significativa, el art. 22 resultará aplicable, con las obligaciones que ello implica.

En concreto, el RGPD exige que los sistemas de IA respeten el derecho del interesado a solicitar la intervención humana, expresar su punto de vista e impugnar decisiones. Además, estos sistemas deben proporcionar información significativa sobre la lógica aplicada, los datos considerados y las consecuencias previstas de las decisiones automatizadas. Por ejemplo, un sistema de triaje automatizado que priorice pacientes en una sala de emergencias debe explicar la lógica adoptada para que sus decisiones sean justas y equilibradas. Del mismo modo, en el contexto de los seguros de salud, cualquier decisión automatizada sobre la elegibilidad del asegurado debe someterse a supervisión humana para prevenir errores derivados de sesgos en los datos de entrenamiento.

El art. 22.4 del RGPD establece restricciones adicionales para las decisiones automatizadas que utilicen categorías especiales de datos, como los relativos a la salud. Estas decisiones están prohibidas, salvo que se cumplan condiciones estrictas, como las previstas en el art. 9.2.g, que permite el tratamiento de datos sensibles cuando sea necesario por razones de interés público sustancial, siempre que esté respaldado por legislación adecuada y acompañado de medidas de protección. En el caso de una emergencia sanitaria, por ejemplo, un sistema automatizado que gestione recursos críticos —como es el caso de las camas en unidades de cuidados intensivos—, podría justificar el tratamiento automatizado de datos sensibles bajo esta excepción, siempre que se implementen medidas como auditorías, supervisión humana y transparencia en el proceso.

En conclusión, la aplicabilidad del art. 22 del RGPD a los sistemas de diagnóstico con IA dependerá de si las decisiones tomadas pueden calificarse como «basadas únicamente en el tratamiento automatizado». En este sentido, mientras que los sistemas de apoyo al diagnóstico suelen quedar fuera de este ámbito debido a la intervención humana significativa, otros sistemas como los de triaje o evaluación de seguros pueden estar sujetos a las restricciones del art. 22.

*Capítulo 5*

# *El Reglamento de Inteligencia Artificial: aspectos generales*

Este capítulo aborda el estudio del Reglamento de Inteligencia Artificial de la Unión Europea (Reglamento (UE) 2024/1689, RIA), también conocido como Ley de Inteligencia Artificial, entendido como un marco normativo pionero, diseñado para regular el uso y desarrollo de sistemas de IA, especialmente aquellos clasificados como de alto riesgo.

En primer lugar, se analiza la evolución de esta regulación, desde sus primeras propuestas hasta la incorporación de niveles de riesgo diferenciados y disposiciones específicas para garantizar la seguridad, la transparencia y la protección de derechos fundamentales. A continuación, examinamos los aspectos generales de la norma, su estructura y principios fundamentales, poniendo especial atención en cómo este marco transversal, tecnológicamente neutro y basado en la gestión de riesgos, trata de armonizar la regulación en diversos sectores y aplicaciones, incluida la sanidad. El capítulo también destaca cómo el RIA aspira a alcanzar un equilibrio entre fomentar la innovación tecnológica sin dejar de proteger la seguridad, la salud y los derechos fundamentales de las personas, estableciendo así un estándar normativo pionero en el ámbito internacional, que no está exento de críticas.

Todo ello proporcionará el contexto necesario para estudiar, en el siguiente capítulo, la regulación específica de los dispositivos sanitarios con IA clasificados como productos de alto riesgo según el Reglamento.

## 5.1. ANTECEDENTES

La propuesta de Reglamento de Inteligencia Artificial de 2021 (COM 2021) fue elaborada por la Comisión Europea tomando como base los documentos preparatorios a los que se hizo referencia en el capítulo 2º. La propuesta inicial de la Comisión se inspiró en el modelo del **Nuevo Marco Legislativo** (NML) de la Unión Europea (Decisión n.º 768/2008/CE),

un marco regulatorio técnico diseñado para garantizar la seguridad de los productos y su libre circulación en el espacio europeo.

La regulación finalmente adoptada se caracteriza por su **naturaleza transversal**, diseñada para ser aplicable a todos los sectores y usos de la inteligencia artificial, independientemente del sector en el que se implemente (EPRS, 2021). En vez de desarrollar normativas especiales para cada sector, el Reglamento propone un marco transversal que establece principios comunes aplicables a las distintas aplicaciones sectoriales de la IA. Este planteamiento de carácter horizontal y general, está alineado con el marco regulador de seguridad de productos ya establecido en la Unión Europea en el Reglamento (UE) 2023/988.

El Reglamento pretende, a su vez, **armonizar la regulación de la inteligencia artificial con otros marcos legislativos existentes**, como la Ley de Servicios Digitales (DSA), la Ley de Mercados Digitales (DMA), la Ley de Datos (DA), la Ley de Gobernanza de Datos (DGA) y el Reglamento General de Protección de Datos (RGPD) y el Reglamento del Espacio Europeo de Datos de Salud (EDDS). Se trata con todo ello de garantizar un entorno normativo coherente, en el que el ciudadano se beneficie de la vasta experiencia de la Unión Europea en la regulación de tecnologías emergentes y en la protección de los derechos fundamentales.

Por otra parte, la regulación es **tecnológicamente neutra**, lo que implica que no regula tecnologías específicas, sino que se fija en los usos y riesgos inherentes a los sistemas de IA. Esta técnica regulatoria pretende garantizar la vigencia de las normas en un entorno tecnológico en constante evolución, evitando su obsolescencia y fomentando la innovación dentro de un marco normativo amplio (Presno Linera y Meuwese, 2024).

Se trataba de adoptar un marco normativo preventivo fundamentado en la evaluación de riesgos (COM 2021), que permitiera a los responsables políticos considerar los riesgos específicos asociados a los distintos usos de la inteligencia artificial según su impacto en la seguridad, la salud y los derechos fundamentales. Este sistema tiene en cuenta tanto la gravedad de los posibles daños como la probabilidad de que ocurran, teniendo en cuenta el contexto particular en el que se utilizan los sistemas de IA. Al ajustar la regulación a las características de cada contexto y al posible daño que pueda ocasionar una aplicación, la norma pretende evitar consecuencias inesperadas y proteger el interés público. Para ello, establece mecanismos que permiten evaluar y supervisar cada sistema, valorando su impacto social y su capacidad de afectar derechos fundamentales, la salud o la seguridad. Por ejemplo, el uso de IA en diagnósticos médicos exige

una supervisión estricta debido a su alta sensibilidad, mientras que aplicaciones en áreas menos críticas, como la optimización logística, requieren un marco normativo más flexible.

Es relevante señalar que la propuesta inicial de la Comisión (COM 2021) contemplaba únicamente dos niveles de riesgo, sin incluir disposiciones relativas a riesgos mínimos ni prohibiciones al uso de la IA. Más tarde, el Parlamento Europeo introdujo importantes enmiendas, incorporando niveles adicionales de riesgo y prohibiciones específicas para proteger derechos fundamentales como la privacidad, la no discriminación y la transparencia.

La evolución de este marco regulador, que pasó de una estructura binaria a un modelo más sofisticado con cuatro niveles de riesgo, ha sido acogida con opiniones divididas. Según Orwat et al. (2022), la escala de riesgos refleja mejor las complejas implicaciones éticas y sociales de la inteligencia artificial, traduciéndose en una regulación más adaptada para abordar los problemas específicos asociados a esta tecnología. En cambio, para una gran parte de la doctrina (European Digital Rights, EDRi, 2021, European Law Institute ELI, 2023) este modelo basado en el riesgo es insuficiente para garantizar una protección efectiva de los derechos fundamentales, en particular frente a la discriminación de colectivos vulnerables.

En cuanto a sus contenidos específicos, la propuesta inicial de la Comisión se basó en los principios establecidos por el Reglamento General de Protección de Datos en materia de calidad de datos y privacidad a los que hemos hecho referencia en el capítulo precedente, pero adaptándolos a las nuevas necesidades, como la gestión de los sesgos o la explicabilidad de los modelos. En particular, se consideró que la **calidad de los datos** debía contemplarse de forma transversal en el Reglamento de Inteligencia Artificial con el fin de que fuera aplicable a cualquier sistema de IA (COM, 2020a, 2020b) en lugar de limitarse a remitir la cuestión a normas sectoriales de seguridad de los productos.

Durante la tramitación del proyecto también se produjo un debate importante en relación a si la IA autónoma debe clasificarse **como producto o como servicio**. Esta distinción conllevaba importantes implicaciones, ya que influía en los requisitos de cumplimiento y las responsabilidades de los proveedores. Finalmente, se decidió que el marco más garantista para la regulación de los sistemas autónomos de IA era el NLF referido, como hemos visto, a la de seguridad de los productos.

Mientras la iniciativa legislativa se hallaba en preparación, irrumpió la **IA generativa o de propósito general** (GPAI), y el legislador europeo se vio en la obligación de introducir provisiones específicas de última hora para

regular este fenómeno mediante una propuesta flexible capaz de adaptarse a la evolución de estos modelos generativos. Era urgente e ineludible abordar estos nuevos sistemas basados en grandes modelos de lenguaje cuya lógica escapaba a la regulación propuesta hasta la fecha (Renda, 2021).

El RIA asignó finalmente a la Oficina de IA de la Comisión Europea un papel central en el desarrollo de herramientas destinadas a evaluar riesgos sistémicos de estos grandes modelos de lenguaje, estableciendo además el encargo de reducir las asimetrías informativas y formar a los reguladores en las competencias necesarias para supervisar esta tecnología. Nótese que la aludida brecha de información entre la industria tecnológica y los organismos reguladores era una cuestión significativa, ya que los reguladores no solo deben comprender los aspectos técnicos de los (GPAI), sino también evaluar sus implicaciones éticas, sociales y económicas en distintos contextos.

## 5.2. EL NUEVO MARCO LEGISLATIVO EN MATERIA DE SEGURIDAD DE PRODUCTOS DIGITALES

En el apartado anterior se ha puesto de relieve que la Ley de Inteligencia Artificial de la UE, como parte de un marco normativo más amplio, resalta la importancia de establecer normas armonizadas para regular la comercialización, puesta en servicio y uso de sistemas de alto riesgo. Estas disposiciones se alinean con los principios del Nuevo Marco Legislativo (NML), diseñado para garantizar la protección de los derechos fundamentales y la seguridad de los consumidores en un entorno tecnológico en constante evolución.

La Directiva (UE) 2019/882, sobre los requisitos de accesibilidad de los productos y servicios, transpuesta en España por la Ley 11/2023, recoge importantes novedades en relación a los agentes económicos que guarden relación con los productos, así como los requisitos de accesibilidad universal que deben reunir tanto los productos como los servicios que se ofrecen al consumidor europeo.

Por su parte, el referido Reglamento (UE) 2023/988, relativo a la seguridad general de los productos, constituye un pilar central de este marco normativo. Este Reglamento introduce una "red de seguridad" que protege a los consumidores frente a riesgos asociados con productos digitales, incluida la ciberseguridad. Además de actualizar las definiciones sustantivas en materia de seguridad de productos, amplía su alcance a productos vendidos en línea fuera de la Unión Europea, atendiendo a los retos emer-

gentes derivados de la digitalización y la economía circular. En el ámbito de la inteligencia artificial, la interacción entre este Reglamento y otras normativas armonizadas resulta necesaria para abordar riesgos transversales que no pueden ser regulados por un único instrumento.

El Reglamento de Inteligencia Artificial incorpora los principios del NML para regular los sistemas de IA clasificados como de alto riesgo. Según el Considerando 3 del RIA, estos productos deben someterse a un proceso de evaluación de conformidad con el fin de obtener el marcado CE, que en algunos casos incluye la intervención de organismos notificados. Este procedimiento asegura que los sistemas de IA cumplan con los estándares exigidos de calidad y seguridad, reforzando la confianza en las prácticas tecnológicas alineadas con la normativa europea. Además, el Considerando 9 del RIA subraya que estas normas no solo se orientan a garantizar la seguridad de los productos, sino que también pretenden preservar un nivel elevado de protección de los derechos fundamentales, incluyendo la prohibición de discriminación consagrada en la Carta de Derechos Fundamentales.

Otro elemento destacado propio del marco NML adoptado en el RIA es la supervisión externa. Según el art. 77 del Reglamento, las autoridades responsables de proteger los derechos fundamentales deben tener acceso a la documentación de los sistemas de IA de alto riesgo. Este acceso facilita la realización de auditorías independientes que mitigan los riesgos de sesgos en las autoevaluaciones realizadas por los fabricantes. De este modo, se refuerza la protección de los ciudadanos frente a posibles impactos negativos derivados de decisiones automatizadas.

En materia de vigilancia del mercado, el RIA también establece mecanismos robustos para supervisar el cumplimiento de los requisitos de seguridad. Conforme al art. 65, las autoridades de vigilancia deben garantizar la trazabilidad de los productos a lo largo de la cadena de suministro. Este requisito permite trazar el ciclo de vida de los productos desde su fabricación hasta el consumidor final, facilitando su retirada en caso de que se detecten riesgos para la seguridad o el incumplimiento de las normativas aplicables.

En el ámbito de los productos sanitarios, la integración de estas disposiciones con el Reglamento de Productos Sanitarios (Reglamento (UE) 2017/745), al que se hará referencia en el capítulo 7º, refuerza la seguridad y la protección de los derechos fundamentales. Este esfuerzo integrador está dirigido a asegurar que los sistemas de IA utilizados en la sanidad cumplan con estándares estrictos de seguridad y calidad, al tiempo que

promueve una regulación alineada con los valores fundamentales europeos.

Finalmente, cabe destacar que el definir la estrategia de cumplimiento reglamentario de cualquier producto sanitario, hay que considerar toda la normativa y legislación aplicable. Debe considerarse pues la legislación europea de seguridad de productos y los conceptos básicos establecidos por la Comisión Europea para la comercialización de productos regulados incluidos en la denominada "Blue Guide" (COM 2022).

## 5.3. NATURALEZA MIXTA DEL REGLAMENTO DE INTELIGENCIA ARTIFICIAL: ¿UNA REGULACIÓN TÉCNICA ORIENTADA A PROTEGER DERECHOS FUNDAMENTALES?

El Reglamento de Inteligencia Artificial se caracteriza por su naturaleza dual, combinando un enfoque técnico basado en la gestión del riesgo, propio del Nuevo Marco Legislativo sobre seguridad de productos, con una dimensión ética orientada a la protección de los derechos fundamentales, tal como hemos explicado en los epígrafes anteriores. Esta combinación se evidencia especialmente en la regulación de los sistemas de alto riesgo, como los dispositivos médicos que incorporan IA, donde la seguridad y la protección de derechos, convergen.

En su preámbulo, el Reglamento subraya la necesidad de respetar los valores y derechos esenciales de la Unión Europea. Así, el Considerando 10 establece que el desarrollo y uso de los sistemas de IA deben evitar cualquier forma de discriminación y proteger a las personas frente a impactos negativos, alineándose con principios como la justicia, la equidad y la transparencia. En la misma línea, el Considerando 24 subraya la importancia de diseñar tecnologías que no perpetúen prácticas discriminatorias, lo que requiere mecanismos de evaluación y supervisión capaces de identificar y mitigar sesgos perjudiciales. El Considerando 27, al recoger las referidas directrices del Grupo de Alto Nivel sobre IA, refuerza la idea de que esta tecnología debe ser coherente, fiable y centrada en el ser humano, orientándose hacia principios éticos que promuevan la inclusión y la equidad. Los Considerandos 48 y 70, por su parte, destacan que los sistemas de alto riesgo deben someterse a pruebas continuas que garanticen un trato justo y neutral.

De este modo, el Reglamento introduce la necesidad de establecer medidas de equidad algorítmica, no solo a partir de criterios técnicos de seguridad, sino también incorporando la perspectiva ética para proteger la

privacidad, la autonomía y la dignidad de los usuarios. Asimismo, el Considerando 46 hace hincapié en la transparencia y en la supervisión humana como herramientas para prevenir decisiones automatizadas injustas o discriminatorias, mientras que el Considerando 64 exige que, antes de su despliegue, los sistemas de IA de alto riesgo demuestren su capacidad para evitar sesgos y efectos adversos sobre los derechos fundamentales.

Esta vocación protectora de los derechos fundamentales de los sujetos objeto de las inferencias algorítmicas no se limita al ámbito de los Considerandos, sino que adquiere cuerpo normativo en los arts. 9 a 15 del RIA. Aunque estos preceptos no establecen derechos subjetivos propiamente dichos, sí introducen un conjunto de obligaciones estrictas que garantizan, de manera indirecta, la protección de los derechos fundamentales. Como veremos en el capítulo siguiente, el art. 9, que regula la gestión de riesgos, junto con el art. 10, centrado en la gobernanza de datos, aseguran tanto la fiabilidad de los sistemas como la protección de la privacidad. A su vez, la transparencia recogida en el art. 13 obliga a que las decisiones automatizadas sean comprensibles para los usuarios y las autoridades competentes, evitando decisiones opacas que, además, puedan resultar discriminatorias. Por otro lado, el art. 14, al exigir la supervisión humana, protege la autonomía y la dignidad asegurando que las decisiones algorítmicas no se adopten sin un control humano. El art. 15 obliga a que los sistemas de IA estén provistos de fuertes medidas de ciberseguridad, precisión y solidez, para prevenir errores de diseño. Finalmente, el art. 86 –mediante vías de recurso— reconoce un derecho limitado de explicación ante decisiones tomadas individualmente.

Sobre la cuestión de la naturaleza de las obligaciones contempladas en los artículos citados, la doctrina se coloca en posiciones divergentes. Algunos expertos consideran que estos preceptos son principalmente obligaciones regulatorias dirigidas a los operadores de sistemas de IA. Desde esta perspectiva, el RIA establece estándares de seguridad y fiabilidad que las entidades deben cumplir, sin conferir derechos individuales directamente exigibles (García y Martínez, 2023).

Otros autores, sin embargo, sostienen que estas disposiciones, aunque no consagren derechos explícitos, crean un marco normativo protector que garantiza de forma indirecta los derechos fundamentales (López, 2022; Berenguer, 2024). Según esta interpretación, a la que nos adherimos (Alkorta Idiakez, 2024), las obligaciones sobre transparencia, supervisión humana y gestión de riesgos no solo constituyen requisitos técnicos, sino

que también actúan como mecanismos para salvaguardar la privacidad, la autonomía y la dignidad de las personas.

Es preciso reconocer que el Reglamento va más allá de una protección meramente indirecta de los derechos fundamentales. El art. 71 establece expresamente el derecho de los individuos a reclamar ante las autoridades competentes si consideran que sus derechos han sido vulnerados por el uso de un sistema de IA. Este artículo obliga a las autoridades a investigar las reclamaciones, adoptar medidas correctoras y comunicar los resultados a los reclamantes, lo que refuerza el carácter protector del RIA. Además, el Considerando 87 introduce la protección de los denunciantes, garantizando que cualquier infracción pueda ser reportada sin temor a represalias y fortaleciendo así la rendición de cuentas en el uso de la inteligencia artificial.

A la luz de todo lo anterior, es defendible que el Reglamento de Inteligencia Artificial, aunque se basa en un modelo de evaluación de riesgos característico de la regulación técnica, trasciende la visión puramente técnica centrada en la seguridad de los productos, para manifestar una preocupación por defender los valores democráticos de la Unión Europea y los derechos fundamentales.

El Considerando 6 sintetiza esta aspiración al destacar la necesidad de desarrollar una IA centrada en el ser humano, orientada al bienestar colectivo y alineada con los derechos fundamentales consagrados en el art. 2 del Tratado de la Unión Europea y en la Carta de los Derechos Fundamentales. De esta manera, el RIA comparte con el Reglamento General de Protección de Datos una inspiración en los principios de protección de los derechos fundamentales de las personas físicas (Berenguer Albaladejo, 2024). Esta vocación garantista resulta indispensable para generar confianza en una tecnología que, debido a su profundo impacto social, requiere equilibrar la innovación tecnológica con la protección efectiva de los derechos fundamentales. Así, el Reglamento de Inteligencia Artificial se configura, en nuestra opinión, como un instrumento normativo híbrido que, al tiempo que asegura la seguridad de los sistemas, tiene la vocación de proteger y promover los valores esenciales y los derechos fundamentales de las personas en la era digital.

## 5.4. DEFINICIÓN DE LA IA

El Reglamento (UE) 2024/1689 adopta una definición amplia de inteligencia artificial, concebida para abarcar una amplia gama de sistemas informáticos avanzados (RIA, art. 3.1):

> *"[un] sistema basado en una máquina diseñado para funcionar con distintos niveles de autonomía, 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 información de salida, como predicciones, contenidos, recomendaciones o decisiones, que puede influir en entornos físicos o virtuales."*

La adopción de una definición amplia en el ámbito de la inteligencia artificial refleja la intención del legislador de establecer una regulación capaz de mantenerse vigente frente a un campo en constante transformación. Este carácter dinámico, marcado por la irrupción de desarrollos impredecibles durante la redacción legislativa, requería una perspectiva que no se limitara a lo ya conocido, evitando así que futuras propuestas tecnológicas quedaran fuera del marco regulador. Además, la diversidad de aplicaciones de la inteligencia artificial en sectores tan diversos como la medicina, la manufactura o la logística exigía un concepto amplio de la IA, que abarcara contextos variados sin comprometer la coherencia normativa ni generar lagunas legales que pudieran poner en riesgo la protección de los derechos fundamentales.

Según análisis recientes, la amplitud de la definición es esencial para garantizar un control adecuado sobre tecnologías emergentes y proporcionar seguridad jurídica tanto a desarrolladores como a usuarios (Campos, 2024). Se trata de alinear el desarrollo de la inteligencia artificial con los valores éticos y jurídicos fundamentales de la Unión Europea (Smith, 2024). Sin embargo, la amplitud del ámbito de aplicación del Reglamento no está exenta de críticas: algunos desarrolladores consideran que esta amplitud puede resultar excesiva, sometiendo a tecnologías de riesgo limitado o aplicaciones específicas a supervisiones estrictas que podrían obstaculizar la innovación. Este debate pone de manifiesto la tensión entre la flexibilidad necesaria para una regulación garantista y la preocupación por evitar cargas regulatorias que inhiban la innovación en el sector (Orwat et al., 2022).

## 5.5. CLASIFICACIÓN DE LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL EN FUNCIÓN DEL RIESGO

Adoptando el enfoque basado en el riesgo propio del NML, la regulación establece, como se ha dicho más arriba, una clasificación de los sistemas de IA basada en diferentes niveles de riesgo: riesgo inaceptable, alto riesgo, riesgo limitado y riesgo mínimo. Cada nivel de riesgo está ilustrado

por casos de uso y regula los requisitos correspondientes para la introducción de productos en el mercado.

Los sistemas de IA que presentan un **riesgo inaceptable** están prohibidos debido a su potencial para causar daños significativos en la seguridad, la saludo o los derechos fundamentales de las personas. Ejemplos de estos sistemas comprenden aquellos que manipulan el comportamiento humano mediante técnicas subliminales que pueden causar daños físicos o psicológicos, sistemas que explotan vulnerabilidades de grupos específicos como niños o personas con discapacidades para distorsionar su comportamiento, y sistemas de puntuación social que evalúan a las personas en función de su comportamiento cívico. Estos sistemas están prohibidos y no pueden ser introducidos en el mercado bajo ninguna circunstancia.

El Reglamento considera de **alto riesgo** a ciertos sistemas de IA en función de su finalidad, las modalidades específicas de su uso y las posibles consecuencias negativas para la salud, la seguridad y los derechos fundamentales de las personas. Se trata, por ejemplo, de sistemas de identificación biométrica utilizados en espacios públicos, sistemas de IA utilizados en la selección de personal o sistemas de IA incorporados a dispositivos médicos. Estos supuestos, como veremos más detalladamente a continuación, deben cumplir con estrictos requisitos de conformidad antes de su comercialización, incluyendo la evaluación de riesgos, la gestión de datos, la provisión de documentación técnica, la transparencia y la supervisión postcomercialización.

Los sistemas de IA de **riesgo limitado** requieren ciertas obligaciones de transparencia para asegurar que los usuarios estén informados sobre su uso. Ejemplos de estos sistemas son aquellos que generan contenido interactivo, como chatbots, y sistemas utilizados en la publicidad personalizada. Estos sistemas deben cumplir con requisitos de transparencia, tales como notificar a los usuarios que están interactuando con un sistema de IA y proveer explicaciones claras sobre cómo funciona el sistema y cómo se toman las decisiones.

Finalmente, los sistemas de IA de **riesgo mínimo** son aquellos que presentan un riesgo bajo y no requieren requisitos adicionales específicos para su puesta en el mercado. Ejemplos de estos sistemas incluyen filtros de spam y sistemas de recomendación de música o de películas. Estos sistemas no están sujetos a requisitos adicionales específicos bajo el RIA, pero deben cumplir con las normativas generales aplicables a todos los productos en el mercado.

La mayor parte de la regulación que contiene la norma se dedica a la regulación de los SIA de alto riesgo. Veamos a continuación con mayor detenimiento los criterios para determinar qué sistemas de inteligencia artificial deben considerarse de alto riesgo que establece el art. 6 del RIA. Estos sistemas se identifican en función de dos situaciones principales.

En primer lugar, el art. 6.2 considera de alto riesgo aquellos que se emplean en ámbitos específicamente enumerados en el Reglamento, donde su uso puede tener un impacto significativo en la salud, la seguridad o los derechos fundamentales de las personas. Estos sectores están detallados en el Anexo III, que incluye áreas como la educación, el empleo o la justicia, entre otros.

En segundo lugar, el art. 6.1 dispone que un sistema también será clasificado como de alto riesgo si, de acuerdo con ciertos actos legislativos de armonización de la Unión Europea listados en el Anexo I, se requiere que el producto o servicio al que pertenece pase por un procedimiento de evaluación de conformidad con un organismo de evaluación independiente. Este criterio se aplica, por ejemplo, a productos que contienen componentes de seguridad o que deben cumplir estrictos estándares regulatorios antes de su comercialización, como los productos sanitarios y los dispositivos de diagnóstico *in vitro*.

Cabe subrayar que solo los sistemas considerados de alto riesgo por el Reglamento Europeo de Productos Sanitarios o por Reglamento Europeo de Productos Sanitarios In Vitro serán automáticamente considerados de alto riesgo para el RIA. En cambio, los productos sanitarios que según los Reglamentos de Producto Sanitario y de Productos Sanitarios In Vitro no requieran una evaluación de conformidad por terceros no serán considerados de alto riesgo para el RIA, a no ser que cumplan con las condiciones de calificación de alto riesgo por otros motivos. Esto significa que los productos sanitarios de clase I y los productos sanitarios in vitro de clase A, los cuales no requieren una evaluación por organismo notificado, no serán considerados de alto riesgo desde el punto de vista del Reglamento de Inteligencia Artificial.

Los productos sanitarios que requieren una evaluación de conformidad por parte de un organismo notificado, en cambio, están sujetos a un proceso más exhaustivo. Este procedimiento, como veremos en el capítulo 7º de esta obra, comprende verificaciones por parte de terceros independientes para garantizar que el sistema cumple con los estándares de seguridad antes de que pueda ser introducido en el mercado.

El Reglamento también reconoce la posibilidad de actualizar y ampliar la lista de sistemas de IA considerados de alto riesgo en su Anexo III, a través de actos delegados de la Comisión (art. 7), con el fin de que el marco regulatorio permanezca alineado con los avances tecnológicos y las nuevas formas de uso que puedan emerger.

Respecto a la entrada en vigor de las obligaciones asociadas a los sistemas de alto riesgo, el Reglamento de Inteligencia Artificial introduce un sistema escalonado. El art. 113 establece un calendario de implementación en función de la clase de sistema de alto riesgo, lo que permite adaptar los requisitos a las características específicas de cada tipo de sistema y al nivel de madurez tecnológica en los sectores correspondientes. Las disposiciones relativas a los sistemas de alto riesgo del Anexo I del Reglamento de Inteligencia Artificial entrarán en vigor a los doce meses de la entrada en vigor del Reglamento, es decir, el 1 de agosto de 2025.

Como veremos, este plazo debe cohonestarse con los plazos previstos en la legislación sectorial, los cuales, en algunos casos, como los productos sanitarios *legacy*, regulados por el Reglamento 2023/607 que modifica el RPS, a los que nos referiremos en el capítulo siguiente, no coincide con los establecido en el RIA.

## 5.6. DISTRIBUCIÓN DE LA RESPONSABILIDAD A LO LARGO DE LA CADENA DE VALOR

El art. 25 del Reglamento de Inteligencia Artificial aborda específicamente la distribución de la responsabilidad por daños en la cadena de valor de la IA. Este artículo establece que todos los actores involucrados en el ciclo de vida de los sistemas de IA, incluidos proveedores, distribuidores, importadores y usuarios, tienen responsabilidades claras para garantizar la seguridad y la conformidad de los sistemas de IA con los requisitos del Reglamento.

De esta manera, los proveedores deben garantizar que los sistemas de IA se desarrollen y sean supervisados de acuerdo con las regulaciones establecidas, incluyendo la realización de evaluaciones de riesgos y la implementación de medidas adecuadas de mitigación de los riesgos detectados. Los distribuidores, por su parte, tienen la obligación de verificar que los productos que comercializan cumplen con las certificaciones necesarias y de proporcionar información clara y accesible a los consumidores.

Los importadores y otros intermediarios en la cadena de valor también deben asegurarse de que los sistemas de IA que introducen en el mercado

cumplan con los requisitos del RIA. Además, los usuarios finales tienen la responsabilidad de utilizar los sistemas de IA conforme a las instrucciones proporcionadas por el propio producto y reportar cualquier incidente o mal funcionamiento.

En conjunto, estas disposiciones tienen como objetivo distribuir de manera equitativa las responsabilidades a lo largo de toda la cadena de valor, garantizando la seguridad y la protección de los derechos de los usuarios y de los afectados por las decisiones algorítmicas. Al mismo tiempo, aseguran y refuerzan que los SIA cumplan con las normativas sectoriales establecidas en materia de seguridad de productos. En particular, el Reglamento de IA sirve de base a la Directiva de Responsabilidad por Productos Defectuosos al imponer requisitos adicionales de transparencia, seguridad y documentación, que constituirán el canon de seguridad exigible a estos productos a la hora de demostrar su conformidad con la normativa europea y asignar responsabilidades en el caso de daños generados por la falta de conformidad.

*Capítulo 6*

# *Los sistemas de alto riesgo: requisitos específicos para su autorización*

Este capítulo está dedicado a estudiar las exigencias particulares que el Reglamento de Inteligencia Artificial establece para los productos considerados de alto riesgo, a los que nos hemos referido de forma somera en el capítulo anterior, ya que se trata de una categoría que abarca la mayor parte de los dispositivos médicos que incorporan IA.

En primer lugar, se abordan las exigencias que recaen sobre los proveedores de estos sistemas. Entre ellas, destaca la necesidad de realizar evaluaciones de conformidad rigurosas, que permitan verificar el cumplimiento de los requisitos técnicos y normativos. De igual modo, se analizan los mecanismos para la gestión de riesgos, destinados a identificar, prevenir y mitigar posibles impactos adversos en el funcionamiento de los dispositivos. Como veremos, la gobernanza adecuada de los datos desempeña un papel central, puesto que la recopilación y el tratamiento inadecuados podrían comprometer tanto la fiabilidad de los resultados como los derechos de los usuarios a un trato no discriminatorio.

El capítulo analiza además la implementación de sistemas de calidad que aseguren la correcta operativa en todas las etapas del ciclo de vida del producto, así como la inclusión de procesos de supervisión humana que permitan ejercer una vigilancia sobre el desempeño de estos dispositivos a lo largo de su ciclo de vida completo.

Por otro lado, se estudia la relación entre las normativas sectoriales específicas en materia de salud – a saber, el Reglamento de Productos Sanitarios— y los principios generales consagrados en el Reglamento de Inteligencia Artificial. Este análisis pondrá de manifiesto cómo, en muchos casos, ambas esferas reguladoras no logran integrarse de manera sistémica, lo que genera tensiones normativas y problemas interpretativos que deben ser resueltos para evitar inconsistencias en la aplicación de la ley.

Finalmente, el capítulo concluye que es necesario promulgar una regulación específica para el ámbito de la salud digital. La regulación propuesta debe ser capaz de abordar la complejidad inherente a estos sistemas, equilibrando de forma adecuada la promoción de la innovación tecnológica con la protección efectiva de la salud, la preservación de los derechos fundamentales y la garantía de seguridad para los usuarios. Se requiere, en definitiva, un marco normativo que responda a las particularidades del sector sanitario y ofrezca respuestas a los dilemas éticos, técnicos y sociales que plantea el desarrollo acelerado de la inteligencia artificial en este campo.

## 6.1. LOS PRODUCTOS SANITARIOS CON IA COMO SISTEMAS DE ALTO RIESGO

En el ámbito de los productos sanitarios, la consideración de estos productos como de alto riesgo se justifica por su interacción directa con personas en situaciones de vulnerabilidad y por las decisiones críticas que los sistemas pueden tomar respecto a estas personas. Un error en el diagnóstico o en la priorización de la atención sanitaria urgente —como puede ocurrir en los sistemas de triaje que incorporan inteligencia artificial—, es susceptible de acarrear consecuencias graves no solo para la salud de los pacientes, sino también para la equidad del tratamiento que reciben estas personas, lo cual es especialmente preocupante si el sistema introduce sesgos discriminatorios relacionados con factores como la edad, el género, la etnia o la discapacidad.

Hemos adelantado que el Reglamento establece una doble vía para clasificar los sistemas de IA como de alto riesgo: por su inclusión específica en el Anexo III y por su regulación previa en otros actos legislativos europeos, como los Reglamentos de Productos Sanitarios (Reglamento (UE) 2017/745, RPS) y el Reglamento de Diagnóstico In Vitro (Reglamento (UE) 2017/746, RDIV). En este último caso, el Reglamento obliga a los fabricantes y proveedores a cumplir tanto con las disposiciones específicas del Reglamento de Productos Sanitarios y del Reglamento de Productos Sanitarios para Diagnóstico In Vitro, como con los requisitos adicionales del propio Reglamento de IA. Esta doble exigencia responde a la necesidad de abordar los riesgos particulares que plantean los dispositivos sanitarios que incorporan IA, los cuales no siempre están contemplados en las normativas sectoriales tradicionales.

En línea con lo señalado en la *Guía Azul sobre la aplicación de la normativa europea relativa a los productos* (2022), el Reglamento de IA subraya que un

producto puede estar sujeto a más de un acto jurídico de la legislación de armonización de la Unión. Por lo tanto, la comercialización o puesta en servicio de un producto solo puede producirse cuando cumple con todos los actos legislativos aplicables. En el caso de productos sanitarios que integran sistemas de IA, este planteamiento reconoce que los riesgos inherentes a la IA no siempre están cubiertos por los requisitos esenciales de salud y seguridad previstos en las normativas sectoriales tradicionales, como el RPS o el RDIV. Por ello, el Reglamento de IA complementa estas normativas al abordar específicamente los peligros asociados al uso de sistemas de inteligencia artificial.

El cumplimiento integral de la normativa obliga a los proveedores a garantizar la conformidad de sus productos con todos los marcos normativos aplicables. Sin embargo, para evitar cargas administrativas y costes innecesarios, el Reglamento promueve una flexibilidad operativa que permite a los fabricantes **integrar procesos de prueba, notificación y documentación requeridos por el Reglamento de IA dentro de los procedimientos ya existentes bajo otras normativas de armonización**. Por ejemplo, un proveedor podría optimizar las evaluaciones y la documentación exigidas por el Reglamento de IA combinándolas con las requeridas por el RPS o el RDIV, siempre y cuando se cumplan plenamente los requisitos de ambas normativas.

Por otra parte, el Reglamento dispensa del cumplimiento de los requisitos de conformidad a los productos sanitarios in house, que se vayan a utilizar en el mismo hospital en el que fueron desarrollados, así como a los sistemas de IA utilizados exclusivamente con fines de **investigación científica** (art. 2, 6). Estas excepciones incluyen actividades como estudios biomédicos que integren IA de manera experimental, siempre que estos experimentos se lleven a cabo bajo el marco ético y normativo de la Unión, ya que, en todo caso, la investigación y el desarrollo de sistemas de IA deben realizarse conforme a estándares éticos y legales reconocidos, como señala el Considerando 25 del propio Reglamento.

## 6.2. REGLAS DE CONFORMIDAD DE LOS SISTEMAS DE IA

Según el Reglamento de Inteligencia Artificial, los proveedores de sistemas de IA de alto riesgo deben someter sus productos a un riguroso procedimiento de evaluación de conformidad antes de su comercialización o puesta en servicio. Las evaluaciones de conformidad, comúnmente utiliza-

das en la regulación de la seguridad de productos (NML), garantizan que los sistemas de inteligencia artificial cumplan con los requisitos establecidos antes de su comercialización. Además, asigna a los Estados miembros la responsabilidad de vigilar el mercado, asegurándose de identificar y retirar aquellos productos que no sean conformes o que representen un riesgo.

La evaluación de conformidad de los sistemas de inteligencia artificial de alto riesgo puede seguir dos métodos principales. El primero es la evaluación basada en el sistema de gestión de la calidad, al que hemos hecho referencia anteriormente. Este sistema, implementado por el proveedor, debe ser evaluado y aprobado por un organismo notificado independiente. Esta evaluación implica que el organismo notificado realice una revisión directa del sistema de IA para verificar su conformidad con los requisitos del Reglamento de Inteligencia Artificial (RIA), lo cual suele comprender pruebas y ensayos específicos para asegurar que el sistema cumple con los estándares de seguridad y calidad establecidos.

El segundo método, que será el más habitual en la práctica, es el contemplado en el art. 40: la autoevaluación de conformidad. Los proveedores pueden optar por este método siempre y cuando puedan demostrar que sus sistemas de IA cumplen con los requisitos del RIA mediante otros medios técnicos, como normas armonizadas o especificaciones técnicas reconocidas. La posibilidad de autoevaluación se sustenta en la existencia de estas normas o especificaciones técnicas que aborden los riesgos específicos del sistema de IA, permitiendo al proveedor demostrar que el sistema cumple plenamente con dichos requisitos. Aunque esta solución puede agilizar el proceso de cumplimiento, también suscita dudas sobre la imparcialidad y el rigor del mismo (Renda, 2021).

La Comisión explica (COM 2021) explica que solo un 5% de estos sistemas requerirán evaluación por parte de los organismos notificados, subrayando la necesidad de salvaguardas robustas para mantener la integridad y seguridad de los sistemas de IA. Nótese, por tanto, que la mayor parte de los sistemas de IA de alto riesgo eludirán la certificación por organismo notificado y dependerán exclusivamente de la autoevaluación.

No obstante, la opción de autoevaluación no exime a los proveedores de la obligación de proteger la salud, la seguridad y los derechos fundamentales de los usuarios. Para asegurar la fiabilidad de las autoevaluaciones, es necesario establecer salvaguardas adecuadas, como las inspecciones aleatorias por parte de los organismos de vigilancia, que equilibren la eficiencia con un cumplimiento riguroso.

En definitiva, por mucho que el Reglamento haya puesto gran énfasis en la regulación de los sistemas de alto riesgo, lo cierto es que, como se acaba de ver, la regulación actual tiende a delegar en el sector privado y en organismos no administrativos —donde los intereses privados predominan sobre el bien público—, la creación de normas y estándares de equidad para la IA, lo que podría llevar a hurtar estos sistemas de la necesaria tutela regulatoria. En el capítulo 2º, hemos criticado el sistema de autoevaluación que se limita a la perspectiva interna de los desarrolladores, sin validación externa ni adaptación a contextos específicos puesto que este sistema deja la responsabilidad de la equidad en manos de los proveedores, quienes deben evitar sesgos injustos mediante procedimientos internos no *supervisados ex ante*.

Nótese que la evaluación de conformidad basada en la verificación del producto por parte de un organismo notificado independiente puede realizarse en base a normativas sectoriales. Los organismos notificados establecidos con ocasión de la evaluación de la conformidad de productos específicos como los sanitarios, con su experiencia y marcos establecidos, constituyen una garantía importante siempre que estén bien dotados y cuenten con las capacidades requeridas para evaluar productos basados en IA. Dichos organismos serán los encargados de auditar los algoritmos de alto riesgo y de asegurarse, entre otros requisitos exigidos por la norma, que no resultan discriminatorios. La participación de estos organismos es fundamental para crear una cultura de transparencia y supervisión de los algoritmos de alto riesgo.

## 6.3. REQUISITOS DE CONFORMIDAD PARA LA IA DE ALTO RIESGO: INTERACCIÓN CON EL RPS Y EL RGPD

El Reglamento de Inteligencia Artificial establece una serie de requisitos esenciales para garantizar la seguridad, transparencia y equidad de los sistemas de IA clasificados como de alto riesgo, incluyendo los productos sanitarios con IA. Estos requisitos abarcan aspectos como la gestión de riesgos, la calidad de los datos, la documentación técnica, la trazabilidad, la supervisión humana, la ciberseguridad, la robustez y la evaluación continua.

El art. 8 del RIA dispone que estas medidas deben ser proporcionadas y eficaces, teniendo en cuenta el estado actual de la técnica en materia de IA. Además, los proveedores deben documentar y justificar sus decisiones,

con la participación de expertos y partes interesadas externas cuando sea necesario.

Recuérdese que cuando un producto contenga un sistema de IA al que se apliquen requisitos del RIA, así como otros requisitos provenientes de normas sectoriales, como es el caso de los productos sanitarios que incorporan IA, los proveedores garantizarán que el producto cumpla con los requisitos previstos en ambas normativas.

Además, conforme al art. 17 del RIA, los proveedores deben implementar un sistema de gestión de la calidad que integre políticas y procedimientos para asegurar que los sistemas cumplen con los estándares de seguridad y calidad establecidos en los arts. 9 a 15. Este sistema debe abordar la gestión de riesgos mediante la identificación, evaluación y mitigación continua de los riesgos asociados con los sistemas de IA. La gobernanza de datos es igualmente necesaria, asegurando que los datos utilizados sean pertinentes, precisos y gestionados de manera segura, protegiendo tanto la privacidad como el control de los datos personales por parte del interesado. La documentación técnica debe detallar el diseño, desarrollo y funcionamiento del sistema, incluyendo información sobre los algoritmos, los datos de entrenamiento y las pruebas realizadas. Esta documentación, que debe estar disponible para las autoridades competentes, está destinada a facilitar la evaluación y supervisión del sistema.

Por último, el RIA exige evaluaciones periódicas tanto de los sistemas de gestión de la calidad como de los sistemas de IA en sí mismos. Estas revisiones permiten identificar mejoras y asegurar el cumplimiento continuo con las normativas del Reglamento. De esta manera, se pretende garantizar que los sistemas de IA de alto riesgo operen de manera segura, ética y conforme a estándares de calidad a lo largo de todo el ciclo de vida.

Veamos con más detenimiento cada uno de los referidos requisitos.

### *a) Requisitos generales del Reglamento de Inteligencia Artificial*

El Reglamento de Inteligencia Artificial establece requisitos esenciales para garantizar la seguridad, transparencia y equidad de los sistemas de IA clasificados como de alto riesgo, incluyendo los productos sanitarios. Regulados entre los artículos 9 y 15, estos requisitos abarcan la gestión de riesgos, la calidad y gobernanza de datos, la documentación técnica, la trazabilidad, la supervisión humana, la ciberseguridad, la robustez y la evaluación continua.

El artículo 8 dispone que estas medidas deben ser proporcionadas y eficaces, documentadas y justificadas, involucrando a expertos y partes interesadas externas cuando sea necesario.

### *b) Evaluación de conformidad y certificación conjunta*

El artículo 8 del RIA dispone que los procesos de prueba y conformidad puedan integrarse dentro de la evaluación de conformidad prevista por la normativa sectorial, evitando duplicidades.

Para los nuevos sistemas de IA de riesgo alto que sean productos sanitarios, este procedimiento quedará enmarcado dentro de la evaluación regulada por el Reglamento de Productos Sanitarios. Con el fin de evitar duplicidades y reducir las cargas adicionales, la Comisión Europea publicará las directrices para la certificación conjunta del Reglamento de Productos Sanitarios y del Reglamento Europeo de IA como más tarde el 2 de febrero de 2026 (arts. 6 y 96 del RIA). Entre tanto, los proveedores podrán optar por integrar los procesos de prueba y conformidad exigidos en el RIA en el proceso de conformidad previsto en la normativa sectorial (art. 8.2).

Para sistemas de IA de riesgo alto que son productos sanitarios y que ya estén certificados (que tengan el marcado CE), sólo será necesario presentar una declaración de conformidad que acredite el cumplimiento del Reglamento Europeo de IA (art. 47 del RIA).

### *c) Sistema de Gestión de la Calidad*

El artículo 17 del RIA establece la obligación para los proveedores de sistemas de IA de alto riesgo de implementar un sistema de gestión de calidad (QMS) que asegure la conformidad con los requisitos del Reglamento. Este QMS debe integrar la documentación técnica requerida (art. 11), la gestión de riesgos (art. 9), el seguimiento posterior a la comercialización (art. 21), las acciones correctivas (art. 20), la generación automática de registros (art. 18) y la cooperación con las autoridades competentes mediante auditorías internas y externas (art. 62).

Cuando un sistema de IA también se clasifique como producto sanitario o de diagnóstico *in vitro*, el proveedor deberá garantizar que el QMS cumpla simultáneamente los requisitos del RIA y de los Reglamentos RPS e RDIV. Normalmente, los fabricantes emplean el citado sistema de calidad basado en la certificación ISO/IEC 13485, que debe adaptarse para integrar los requisitos específicos del art. 17 del RIA.

Se prevé que los proveedores de productos existentes actualicen sus QMS para adecuarlos a las nuevas exigencias, mientras que los nuevos sistemas podrán implementar un QMS único que contemple, desde el diseño, el conjunto de los requisitos. Esta estrategia de integración permite una gestión más eficiente y una reducción significativa de cargas administrativas.

### *d) Documentación técnica*

El artículo 11 del RIA exige la elaboración de una documentación técnica detallada antes de la introducción en el mercado del sistema de IA, la cual debe mantenerse actualizada durante toda su vida útil. Esta documentación debe evidenciar el cumplimiento de los requisitos del Reglamento y ser accesible para las autoridades competentes.

En el caso de sistemas de IA que sean también productos sanitarios o de diagnóstico in vitro, el Reglamento permite combinar la documentación exigida por el RIA con la requerida por los Reglamentos RPS/RDIV.

Para los sistemas que ya estén en el mercado, se deberá complementar la documentación existente con las exigencias descritas; para nuevos sistemas, se podrá presentar un expediente único que responda a ambas regulaciones.

Además, la Comisión Europea establecerá un formulario de documentación técnica simplificado para pequeñas y medianas empresas, facilitando su acceso al mercado.

### *e) Excepciones: herramientas "in house" e investigación clínica*

Existen dos supuestos de excepción parcial o total al cumplimiento de los Reglamentos RPS/RDIV: las herramientas "in house" y las herramientas en fase de investigación clínica.

En el primer caso, se trata de dispositivos fabricados y utilizados en el mismo centro sanitario, de clase I o IIa —no de clase IIb, III ni implantables—, y sin equivalente comercial. En el segundo, se incluyen las herramientas destinadas exclusivamente a verificar la eficacia o seguridad.

Aunque estas herramientas queden exentas de cumplir los MDR/RDIV, deben someterse íntegramente a los requisitos del Reglamento Europeo de IA para garantizar que, incluso en contextos menos regulados, se respeten los derechos fundamentales y la seguridad de los usuarios. Esta suje-

ción incluye la evaluación de conformidad bajo exclusiva responsabilidad del proveedor y previa aprobación de un Comité de Ética y de la AEMPS.

### *f) Gestión de Riesgos y Evaluación de Impacto*

El artículo 9 del RIA exige que los sistemas de IA de alto riesgo cuenten con un sistema de gestión de riesgos iterativo y continuo que abarque riesgos para la salud, la seguridad, los derechos fundamentales, y los derivados del uso indebido previsible.

Cuando se traten datos personales, este sistema debe integrar una Evaluación de Impacto sobre la Protección de Datos (AIPD) conforme al RGPD, identificando y mitigando riesgos para los derechos y libertades de los interesados. La interacción de la evaluación de impacto de derechos fundamentales exigida por el RIA con la AIPD será objeto de tratamiento específico en el apartado 6.8 de esta monografía.

Los proveedores de sistemas ya existentes deben revisar las AIPD realizadas, mientras que los de nuevos productos deben incorporar los riesgos detectados en el sistema de gestión de riesgos exigido por el RIA.

### *g) Gobernanza de Datos: Calidad, Trazabilidad y Equidad*

El artículo 10 del RIA establece obligaciones estrictas sobre la calidad y gobernanza de los datos utilizados en sistemas de IA de alto riesgo. Se exige que los datos sean representativos, precisos, completos y libres de errores, asegurando la equidad y reduciendo riesgos discriminatorios (Morley et al., 2023).

El RIA complementa y amplía al RGPD, al imponer requisitos adicionales de trazabilidad y documentación de sesgos (art. 13), así como medidas correctivas específicas (art. 10.2.g). Por ejemplo, en el ámbito sanitario, el uso de datos sensibles resulta crucial para desarrollar algoritmos que no reproduzcan desigualdades estructurales, siempre bajo estrictas salvaguardas técnicas.

Especialmente relevante es el artículo 10.5 del RIA, que permite el uso de datos sensibles para corregir sesgos, siempre bajo salvaguardas estrictas como la minimización, el cifrado y la anonimización. Recuérdese que el Comité Europeo de Protección de Datos (CEPD, 2024) subraya que la anonimización debe evaluarse caso por caso, si bien esta doctrina ha sido matizada por el TJUE en SCHUFA (C-634/21).

### *h) Conservación de registros y trazabilidad de eventos*

El RIA exige que los sistemas de IA de alto riesgo registren automáticamente eventos relevantes a lo largo de su ciclo de vida, incluyendo riesgos significativos y modificaciones sustanciales.

Para los sistemas enumerados en el Anexo III, esta obligación se extiende al registro de la base de datos de referencia, los resultados de búsqueda y los responsables de verificación.

En esta línea, el artículo 13 del RIA exige que los sistemas de IA se diseñen de manera que faciliten una interpretación adecuada de sus resultados. Asimismo, dispone que toda la documentación técnica debe ser accesible y comprensible para los usuarios previstos. Como explican Wachter et al. (2017), no basta con documentar las especificaciones técnicas relativas a los datos de entrada, los procesos de entrenamiento y la validación; es necesario también incluir ejemplos que ilustren los usos previstos y los excluidos, junto con advertencias sobre los posibles riesgos. Conforme al citado artículo, quienes desplieguen sistemas de IA deben proporcionar información adicional a las personas afectadas, incluyendo detalles sobre los datos utilizados, los algoritmos empleados y las medidas de seguridad implementadas. Esta información debe presentarse de forma comprensible para los usuarios, promoviendo así la confianza y el uso seguro de estos sistemas (Reglamento (UE) 2024/1689, 2024).

Para cumplir estos requisitos, los sistemas de alto riesgo deben incorporar instrucciones de uso claras, completas y adaptadas a los conocimientos y necesidades de sus usuarios finales. Dichas instrucciones deben especificar las características, capacidades y limitaciones del sistema, las condiciones de uso apropiadas y los riesgos asociados. Igualmente, deben describir las medidas de supervisión humana necesarias para interpretar correctamente los resultados y prevenir errores derivados del denominado «sesgo de automatización», que implica una confianza excesiva en las decisiones emitidas por sistemas automatizados. Este aspecto resulta crucial en ámbitos como la salud, donde las decisiones asistidas por IA —por ejemplo, en diagnósticos médicos— pueden impactar directamente en la vida de los pacientes (art. 13.3 y art. 14.4.b).

El artículo 14 del RIA refuerza el papel de la supervisión humana en los sistemas de alto riesgo, estableciendo que deben ser diseñados de manera que permitan una intervención oportuna por parte de personas capacitadas en caso de fallos o decisiones inadecuadas. Este principio busca mitigar los riesgos derivados de la autonomía de los sistemas, asegurando que

siempre exista una instancia humana capaz de interpretar y, si es necesario, corregir las decisiones automatizadas. Así, en el ámbito médico, se exige que los profesionales sanitarios puedan revisar y validar las conclusiones generadas por la IA, garantizando su alineación con los estándares éticos y clínicos aplicables. La supervisión humana se configura, por tanto, como una salvaguarda esencial para mantener el control y la responsabilidad sobre los resultados de la IA.

Por ejemplo, en un sistema diseñado para identificar enfermedades pulmonares, si el resultado influye en la decisión clínica de un médico, las instrucciones de uso deben ser lo suficientemente detalladas para explicar cómo se generan los resultados, qué datos se utilizaron en el entrenamiento y qué limitaciones tiene el sistema en distintos contextos clínicos. Además, deben advertir sobre el riesgo de confiar ciegamente en los resultados sin aplicar el juicio clínico humano, minimizando así el impacto del sesgo de automatización. La falta de explicaciones claras no solo podría derivar en decisiones médicas incorrectas, sino también comprometer la salud del paciente y generar responsabilidades legales para el proveedor.

Desde la perspectiva de los daños, la falta de transparencia en un sistema de IA puede constituir un defecto de información, particularmente si la omisión de detalles sobre su uso o riesgos desemboca en daños evitables mediante advertencias adecuadas. El defecto de información es clave no solo para evaluar la conformidad del producto, sino también para determinar la responsabilidad del fabricante en caso de daños. Como establece el artículo 7.2.b de la nueva Directiva de Responsabilidad por Productos Defectuosos —que abordaremos en profundidad en el capítulo 9º—, el «uso razonablemente previsible del producto» incluye comportamientos derivados de errores comprensibles por parte de los usuarios, lo que subraya la necesidad de instrucciones exhaustivas que minimicen el riesgo de uso indebido.

Desde un enfoque doctrinal, diversos expertos destacan la importancia de los artículos 86 y 13 del RIA para fomentar una cultura de transparencia y responsabilidad en el desarrollo y uso de la IA. Según Pérez (2024), «la combinación de los arts. 86 y 13 del Reglamento de IA representa un avance significativo en la protección de los derechos de los ciudadanos frente a las decisiones automatizadas». Los requisitos de transparencia y explicabilidad del RIA deben interpretarse de forma sistemática junto con los principios de información y transparencia previstos en el Reglamento General de Protección de Datos (RGPD), a los que hicimos referencia en el capítulo 4º. En efecto, el RGPD consagra la transparencia como principio

fundamental en su artículo 5.1.a, vinculado a la licitud y lealtad en el tratamiento de datos personales. Además, los artículos 13.2.f y 14.2.h reconocen derechos de información específicos cuando los interesados son objeto de decisiones automatizadas. Este principio implica que los responsables deben ofrecer información clara y comprensible sobre las finalidades del tratamiento, las categorías de datos utilizadas, la lógica subyacente a las decisiones automatizadas y sus consecuencias previsibles para los interesados.

Sin embargo, aplicar estos requisitos de transparencia plantea retos, especialmente cuando los sistemas de IA operan bajo marcos de protección de derechos de propiedad intelectual y secreto comercial. En el contexto sanitario, ello exige una ponderación delicada de intereses y derechos, que debe resolverse caso por caso. Los desarrolladores deben proporcionar suficiente información para que los profesionales de la salud puedan tomar decisiones informadas basadas en los resultados del sistema, sin poner en riesgo la propiedad intelectual ni los secretos comerciales de la empresa.

Un ejemplo ilustrativo de esta tensión es la sentencia del Tribunal de Justicia de la Unión Europea en el caso SCHUFA (C-634/21, TOL9.882.990), que comentaremos más adelante. En dicho fallo, el TJUE sostuvo que los responsables del tratamiento deben ofrecer explicaciones significativas sobre las decisiones automatizadas, sin necesidad de revelar detalles técnicos confidenciales ni el código fuente. Este enfoque puede extrapolarse al RIA, que permite satisfacer los requisitos de información mediante informes que describan los razonamientos y métodos generales del sistema. Sin embargo, en litigios de responsabilidad civil —como analizaremos en el capítulo 9º— podría ser necesario revelar el código fuente para probar la existencia de un defecto y el nexo causal con el daño.

### *i) Robustez, precisión y ciberseguridad*

El artículo 15 establece que los sistemas de IA de alto riesgo deben ser diseñados para mantener una precisión elevada, robustez frente a errores o ataques, y ciberseguridad a lo largo de todo su ciclo de vida. La protección contra ataques como el envenenamiento de datos o los ataques adversarios resulta esencial (Papernot et al., 2016), requiriéndose la colaboración activa con la Agencia Europea de Ciberseguridad (ENISA).

La robustez debe garantizar decisiones justas y consistentes incluso en condiciones adversas (Binns, 2020), protegiendo especialmente los derechos fundamentales y evitando impactos desproporcionados en grupos vulnerables.

## 6.4. EXPLICABILIDAD E INTERPRETABILIDAD

El derecho a la explicación recogido en el art. 86 tiene como objetivo garantizar que las personas afectadas por las decisiones algorítmicas puedan comprender las inferencias que han llevado a la toma de decisión. Sin embargo, su formulación actual ha sido criticada por su falta de especificidad y claridad operativa (Nnawuchi, George, 2024). Esta crítica se alinea con el análisis de quienes señalan que el derecho no aborda adecuadamente las preocupaciones sobre el papel de la IA en la toma de decisiones individuales (Hernández Peña, 2022; Castilla Barea, 2024).

En primer lugar, debe anotarse que el alcance del derecho a la explicación consagrado en el art. 86 se circunscribe a los sistemas de IA catalogados como de alto riesgo en el Anexo III de la RIA, excluyendo, sin embargo, sistemas cubiertos por actos legislativos de armonización de la Unión, como es el caso de los dispositivos médicos. Por tanto, la invocación del derecho está supeditada a que no existan otras normativas de la Unión o nacionales que regulen específicamente el acceso a dichas explicaciones, en cuyo caso serían de aplicación estas últimas.

Este derecho incluye, por ejemplo, sistemas utilizados por autoridades públicas para evaluar la elegibilidad de personas físicas para servicios esenciales, sistemas empleados en seguros de vida y salud, y algoritmos para clasificar llamadas de emergencia o realizar triajes en contextos de atención sanitaria urgente. El art. 86, por otra parte, restringe su aplicación a personas físicas que consideren que las decisiones afectan negativamente su salud, seguridad o derechos fundamentales.

Aunque el derecho a recibir una explicación no figuraba en la propuesta inicial de la Comisión de 2021, fue incorporado a instancia del Parlamento Europeo y reconoce el derecho de las personas afectadas por decisiones basadas en sistemas de IA de alto riesgo a recibir explicaciones claras y significativas. Haciéndose eco de lo establecido previamente en el art. 22.1 del RGPD, el art. 86 del RIA dispone que, si una decisión afecta significativamente la salud, la seguridad o los derechos fundamentales de las personas, esta tiene derecho a recibir una explicación clara sobre cómo participó el sistema de IA en el proceso de toma de la decisión que le afectó.

A diferencia del RGPD, donde la responsabilidad principal recae en los responsables de los datos, en el RIA las responsabilidades se disocian entre desarrolladores e implementadores. Esta separación resulta altamente problemática debido a los efectos negativos, ampliamente documentados,

que el procesamiento impulsado por IA puede tener sobre la capacidad de toma de decisiones.

En particular, excluir al desarrollador del ámbito de aplicación del derecho a la explicación desatiende la complejidad inherente a la toma de decisiones algorítmica. Por ejemplo, los usuarios finales, como los oficiales de inmigración, suelen carecer del conocimiento necesario para ofrecer explicaciones significativas sobre la influencia del sistema de IA en sus decisiones. Se limita el derecho a explicaciones superficiales, dejando a las personas afectadas sin herramientas efectivas para comprender los motivos de dichas decisiones ni para articular una defensa adecuada.

En el Considerando 93 del RIA se afirma que,

> *«[s]i bien los riesgos relacionados con los sistemas de IA pueden derivarse del diseño de dichos sistemas, también pueden surgir de la forma en que se utilizan»*, justificando así que los responsables del despliegue de sistemas de IA de alto riesgo desempeñen «[*u*]*n papel fundamental en garantizar que los derechos fundamentales sean protegidos, complementando las obligaciones del proveedor durante el desarrollo del sistema de IA»*. Según los legisladores de la UE, los responsables del despliegue son *«los más capacitados para entender cómo se utilizará concretamente el sistema de IA de alto riesgo y, por lo tanto, pueden identificar riesgos significativos potenciales que no se habían previsto en la fase de desarrollo, debido a un conocimiento más preciso del contexto de uso, las personas o grupos de personas probablemente afectados, incluidos los grupos vulnerables»*.

En mi opinión, estas explicaciones resultan insuficientes para dotar al derecho de la eficacia necesaria como remedio.

El ciclo de vida de la IA requiere establecer una cadena de responsabilidad clara y efectiva, que permita a los usuarios contactar a los desarrolladores para resolver posibles sesgos o errores en los sistemas. La doctrina subraya la importancia de integrar las responsabilidades tanto de los desarrolladores como de los implementadores de inteligencia artificial (Demkova, 2024). Se trata, en definitiva, de que todas las partes involucradas en el desarrollo y despliegue de sistemas de IA sean responsables de sus acciones y decisiones, promoviendo así la transparencia y la confiabilidad en el uso de estas tecnologías.

## 6.5. SUPERVISIÓN HUMANA. INTERPRETACIÓN SISTÉMICA CON EL RGPD

La supervisión humana ocupa un lugar central en la regulación de los sistemas de inteligencia artificial, especialmente en aquellos clasificados

como de alto riesgo y en sectores sensibles como el sanitario. Así, el art. 14 del RIA establece que los sistemas de IA deben estar diseñados para permitir una supervisión efectiva durante todo su ciclo de vida. Este requisito técnico tiene como objetivo garantizar la capacidad de los operadores para intervenir en el sistema y, de ser necesario, detenerlo ante riesgos o resultados imprevistos. Así, el diseño técnico de los sistemas de IA debe incluir herramientas que faciliten la monitorización y el control humano.

Por otro lado, hemos defendido que el Reglamento General de Protección de Datos aborda la supervisión humana desde una perspectiva distinta, centrada en la protección de los derechos de los interesados. Según la interpretación del Tribunal de Justicia de la Unión Europea en el caso SCHUFA (C-634/21, TOL9.882.990), el art. 22.1 del RGPD reconoce el derecho de las personas a que decisiones automatizadas con efectos jurídicos o significativamente adversos estén sujetas a una intervención humana significativa. Este derecho no se limita a una mera formalidad, sino que debe permitir que el humano pueda modificar o revertir decisiones automatizadas, asegurando así un control efectivo sobre los procesos algorítmicos. Esta interpretación refuerza la necesidad de garantizar una supervisión que combine la capacidad técnica con la protección de los derechos fundamentales.

¿Cómo se integran ambos marcos normativos a la hora de exigir que determinados productos de alto riesgo cuenten con sistemas de supervisión humana? Aunque ambos marcos normativos coinciden en la importancia de la supervisión, sus objetivos y alcance presentan diferencias. El RGPD establece un derecho a la supervisión humana como salvaguarda frente a decisiones automatizadas, centrando su atención en los responsables del tratamiento y su deber de garantizar que el personal esté debidamente capacitado. En cambio, el RIA exige a los proveedores integrar herramientas técnicas en sus sistemas que faciliten esta supervisión. Según Kiskinov (2024), esta dualidad refleja un modelo regulatorio que reparte responsabilidades entre diseñadores, operadores y usuarios, asegurando que todas las etapas del ciclo de vida del sistema respeten los principios éticos y técnicos.

Por su parte, Lazcoz Moratinos (2023) señala que la supervisión humana debe ser algo más que un mecanismo simbólico y requiere competencias tanto técnicas como éticas. Los supervisores deben tener la capacidad de interpretar resultados, detectar errores y adoptar decisiones correctivas. En el ámbito sanitario, esto se traduce en la necesidad de integrar las recomendaciones algorítmicas con el juicio clínico humano, asegurando que

las decisiones no dependan exclusivamente de los sistemas automatizados. Además, Lazcoz subraya la importancia de diseñar sistemas que incluyan limitaciones operativas, de modo que el control humano no pueda ser desactivado por el propio sistema, y aboga por fomentar una cultura de responsabilidad profesional en el uso de la IA.

Kiskinov (2024) complementa esta perspectiva al destacar que la supervisión no debe ser puramente reactiva, sino que debe prevenir riesgos desde el diseño del sistema. Los proveedores están obligados a incorporar mecanismos que aseguren una interacción efectiva entre humanos y máquinas, lo que incluye no solo herramientas técnicas, sino también medidas prácticas como guías de usuario detalladas, programas de formación y protocolos claros para la gestión de incidentes. Estos elementos son esenciales para permitir que la intervención humana sea oportuna, informada y eficaz.

En síntesis, mientras el RGPD asegura un derecho subjetivo a la supervisión humana frente a decisiones automatizadas, el RIA establece una obligación técnica que recae sobre los proveedores para dotar a los sistemas de herramientas de supervisión. Ambos marcos, aunque distintos en su alcance, se complementan al abordar diferentes dimensiones de la supervisión, garantizando la seguridad, la transparencia y el respeto por los derechos de las personas en un entorno digital cada vez más automatizado.

## 6.6. SUPERVISIÓN DE LOS SISTEMAS TRAS SU INTRODUCCIÓN EN EL MERCADO

La responsabilidad de garantizar que los sistemas de IA cumplan con las normativas no termina con su comercialización, sino que se mantiene en el tiempo. La supervisión *ex post* incluye mecanismos como auditorías, monitoreo continuo del desempeño y la revisión de posibles impactos adversos.

Esta continuidad es especialmente importante en sistemas dinámicos, cuya operación puede variar con el tiempo debido a la evolución de los datos o a los cambios en el entorno donde se utilizan. Según el RIA, los desarrolladores deben monitorear continuamente estos sistemas y actualizarlos según sea necesario, para asegurar su precisión y equidad en contextos cambiantes. Estas prácticas permiten identificar y mitigar problemas que no se detectaron durante la fase inicial de desarrollo o evaluación (De Miguel y Mursuli, 2023), al tiempo que sirve para proteger los derechos fundamentales y a prevenir impactos negativos inesperados (Binns, 2020).

La supervisión de los sistemas de IA de alto riesgo tras su puesta en servicio, articulada a través de procedimientos de evaluación de conformidad, el registro en bases de datos centralizadas y los mecanismos de vigilancia continua, representa un elemento esencial del marco normativo del RIA. Estas medidas no solo garantizan la seguridad y fiabilidad de los sistemas, sino que también están llamadas a reforzar la confianza del público en la implementación ética de estas tecnologías avanzadas.

Los proveedores tienen la obligación de registrar sus sistemas de IA de alto riesgo en una base de datos centralizada de la Unión Europea, gestionada por la Comisión Europea, según se estipula en el art. 60 del RIA. Como hemos referido más arriba, esta base de datos tiene múltiples propósitos, entre ellos aumentar la transparencia, facilitar la vigilancia por parte de las autoridades competentes y reforzar la supervisión pública. Con ello se responde a la creciente preocupación por la opacidad de los sistemas de IA, que puede dificultar la identificación de posibles riesgos o irregularidades tras su implementación (Parlamento Europeo, el Consejo y la Comisión, 2023). La trazabilidad de los sistemas permite además a los usuarios finales —investigadores y partes interesadas— conocer mejor las características de los productos comercializados, y realizar un seguimiento de su desempeño durante toda su vida útil. Además, el registro en la base de datos no solo fortalece la supervisión por parte de las autoridades, sino que también mejora la rendición de cuentas hacia el público general y las organizaciones interesadas. De hecho, estudios recientes subrayan que la transparencia activa y la rendición de cuentas aumentan la aceptación pública de la IA, mitigando preocupaciones sobre posibles riesgos éticos o de discriminación (Floridi et al., 2021).

La Agencia de la Unión Europea para la Ciberseguridad (ENISA) desempeña un papel esencial en la supervisión, ya que proporciona directrices y apoyo técnico para abordar vulnerabilidades emergentes y garantizar la integridad de los sistemas a lo largo de su ciclo de vida operativo (ENISA, 2023).

## 6.7. LA EVALUACIÓN DE IMPACTO DE DERECHOS FUNDAMENTALES

A fin de garantizar eficazmente la protección de los derechos fundamentales, los responsables del despliegue de determinados sistemas de IA de alto riesgo (quien lo utiliza en su operativa, no quien lo desarrolla), antes de ponerlos en servicio, deben llevar a cabo una evaluación del im-

pacto que la utilización de dichos sistemas puede tener en los derechos fundamentales (EIDF).

Está obligación se activa únicamente cuando el sistema de IA es de alto riesgo por remisión al Anexo III —salvo en el caso de las infraestructuras críticas que no requerirá dicha evaluación— y es utilizado por una autoridad pública o una entidad privada que ejerce funciones públicas. Los sistemas que deben someterse a dicha evaluación previa son, por tanto, los contemplados en el Anexo III (a saber, sistemas de biometría, educación y formación profesional, empleo, gestión de los trabajadores, acceso al autoempleo, acceso a servicios y prestaciones públicas esenciales como la asistencia sanitaria, solvencia de las personas físicas, evaluación de riesgos y fijación de precios para contratación de seguros de vida y salud, sistemas de priorización de emergencias, triaje de pacientes en urgencias, garantía de cumplimiento del Derecho, migración asilo y gestión de control transfronterizo, administración de justicia y procesos democráticos), salvo los sistemas de IA empleados para la gestión de infraestructuras críticas, que no requerirán dicho examen antes de su puesta en uso.

El art. 27 del RIA establece que están obligados a dicha evaluación los responsables del despliegue de sistemas de IA de alto riesgo que sean organismos de Derecho público, pero, también las entidades privadas que presten servicios públicos y las entidades bancarias que desplieguen sistemas destinados a ser utilizados para evaluar la solvencia de las personas físicas o establecer su calificación crediticia (salvo los sistemas de IA utilizados para detectar fraudes financieros), así como las entidades de seguros que desplieguen sistemas de IA que evalúan riesgos y fijan precios de seguros de vida y salud.

La meritada evaluación consistirá en las pruebas descritas en el apartado 1 del art. 27, que son las siguientes:

a) una descripción de los procesos del responsable del despliegue en los que se utilizará el sistema de IA de alto riesgo en consonancia con su finalidad prevista;

b) una descripción del período de tiempo durante el cual se prevé utilizar cada sistema de IA de alto riesgo y la frecuencia con la que está previsto utilizarlo;

c) las categorías de personas físicas y colectivos que puedan verse afectados por su utilización en el contexto específico;

d) los riesgos de perjuicio específicos que puedan afectar a las categorías de personas físicas y colectivos determinadas con arreglo a la

letra c) del presente apartado, teniendo en cuenta la información facilitada por el proveedor con arreglo al art. 13;

e) una descripción de la aplicación de medidas de supervisión humana, de acuerdo con las instrucciones de uso;

f) las medidas que deben adoptarse en caso de que dichos riesgos se materialicen, incluidos los acuerdos de gobernanza interna y los mecanismos de reclamación.

El objetivo de la evaluación de impacto relativa a los derechos fundamentales es que el responsable del despliegue determine los riesgos específicos para los derechos de las personas o colectivos de personas que probablemente se vean afectados y defina las medidas que deben adoptarse en caso de que se materialicen dichos riesgos. La evaluación de impacto debe llevarse a cabo antes del despliegue del sistema de IA de alto riesgo y debe actualizarse cuando el responsable del despliegue considere que alguno de los factores pertinentes ha cambiado. La evaluación de impacto debe determinar los procesos pertinentes del responsable del despliegue en los que se utilizará el sistema de IA de alto riesgo en consonancia con su finalidad prevista y debe incluir una descripción del plazo de tiempo y la frecuencia en que se pretende utilizar el sistema, así como de las categorías concretas de personas físicas y de colectivos de personas que probablemente se vean afectados por el uso del sistema de IA de alto riesgo en ese contexto de uso específico.

Según consta en el Considerando 96, al llevar a cabo esta evaluación, el responsable del despliegue debe tener en cuenta la información pertinente para una evaluación adecuada del impacto, lo que incluye, por ejemplo, la información facilitada por el proveedor del sistema de IA de alto riesgo en las instrucciones de uso. Cuando proceda, para recopilar la información pertinente necesaria para llevar a cabo la evaluación de impacto, los responsables del despliegue de un sistema de IA de alto riesgo, en particular cuando el sistema de IA se utilice en el sector público, pueden contar con la participación de las partes interesadas pertinentes, como, por ejemplo, los representantes de colectivos de personas que probablemente se vean afectados por el sistema de IA, expertos independientes u organizaciones de la sociedad civil, en la realización de dichas evaluaciones de impacto y en el diseño de las medidas que deben adoptarse en caso de materialización de los riesgos. La Oficina Europea de Inteligencia Artificial (en lo sucesivo, "Oficina de IA") debe elaborar un modelo de cuestionario con el fin de facilitar el cumplimiento y reducir la carga administrativa para los responsables del despliegue.

A la luz de los riesgos detectados, los responsables del despliegue deben determinar las medidas que han de adoptarse en caso de que se materialicen dichos riesgos, entre las que se incluyen, por ejemplo, sistemas de gobernanza para ese contexto de uso específico, como los mecanismos de supervisión humana con arreglo a las instrucciones de uso, los procedimientos de tramitación de reclamaciones y de recurso, ya que podrían ser fundamentales para mitigar los riesgos para los derechos fundamentales en casos de uso concretos. Tras llevar a cabo dicha evaluación de impacto, el responsable del despliegue debe notificarlo a la autoridad de vigilancia del mercado competente, que en nuestro caso será la propia AESIA.

Nótese que esta evaluación de impacto es distinta de la prevista en el RGPD. El propio Reglamento de Inteligencia Artificial (art. 27.9) reconoce que cuando proceda, los responsables del despliegue de sistemas de IA de alto riesgo utilizarán la información facilitada conforme al art. 13 para cumplir la obligación de llevar a cabo una evaluación de impacto relativa a la protección de datos que les imponen el art. 35 del RGPD. La evaluación de impacto de protección de datos (EIPD) estará por tanto las más de las veces comprendida dentro de la EIDF, aunque esta última es más amplia, incluyendo no solo la protección de los datos, sino también otros derechos fundamentales como la transparencia y la no discriminación en los contextos y las poblaciones concretas en las que se prevé su aplicación.

En lo que se refiere a los productos sanitarios, la obligación de realizar una evaluación de impacto sobre los derechos fundamentales, prevista en el art. 27 del Reglamento de IA, **no se aplica automáticamente a todos los productos sanitarios que incorporen IA**. El artículo 27.1 del RIA establece una exclusión expresa para los sistemas de IA que se integran como parte de productos ya regulados por otra legislación sectorial de la UE contemplada en el Anexo I del Reglamento.

Esta cláusula de exclusión parcial significa que, en tanto el RPS ya imponga requisitos de seguridad, rendimiento, evaluación clínica y gestión del riesgo, los sistemas de IA que funcionen exclusivamente como parte de estos productos no estarán sujetos a las exigencias adicionales del RIA, incluida la obligación de realizar una EIDF, siempre que los riesgos asociados a los derechos fundamentales estén razonablemente cubiertos por el marco existente. En este sentido, debe observarse que el RPS no contempla de manera explícita una evaluación de impacto sobre derechos fundamentales, pero sí contiene, como veremos más adelante, mecanismos de evaluación de riesgos para la salud y la seguridad, así como requisitos éticos indirectos relacionados con el bienestar del paciente. El grado de

equivalencia funcional entre estos requisitos y los del RIA es, por tanto, el fundamento de la exención.

Con todo, la exención no es absoluta. En escenarios de reutilización, recontextualización o ampliación funcional del sistema de IA que incorpora el producto sanitario, dicha exención podría dejar de ser aplicable. Entonces el responsable del despliegue podría verse obligado a realizar una evaluación de impacto sobre los derechos fundamentales, especialmente si se trata de un ente público o una autoridad competente en salud. La coordinación entre ambos marcos regulatorios es, por tanto, esencial para garantizar tanto la seguridad del paciente como la protección efectiva de los derechos fundamentales.

## 6.8. LA AGENCIA ESPAÑOLA DE SUPERVISIÓN DE LA INTELIGENCIA ARTIFICIAL

La creación de la Agencia Española de Supervisión de la Inteligencia Artificial (AESIA), que tuvo lugar mediante Real Decreto 729/2023, de 22 de agosto —antes de la promulgación del Reglamento de Inteligencia Artificial de la Unión Europea—, constituye un paso decisivo en el compromiso del Estado español con la regulación de los sistemas de inteligencia artificial.

En consonancia con las disposiciones del Reglamento de Inteligencia Artificial a los que acabamos de hacer referencia la AESIA desempeña una serie de funciones orientadas a garantizar la correcta aplicación de dicha normativa.

En concreto, el art. 4 del Real Decreto 729/2023, de 22 de agosto, que aprueba el Estatuto de la AESIA, le atribuye la misión de supervisar los sistemas tras su comercialización y, en su caso, sancionar aquellos sistemas de inteligencia artificial que puedan vulnerar derechos como la integridad, la intimidad o la igualdad de oportunidades, con una especial atención a los sesgos de género y étnico-raciales.

La AESIA actúa además como un organismo notificante que acredita a otros organismos notificados para que realicen las evaluaciones de conformidad de los productos de alto riesgo a las que nos hemos referido anteriormente.

Otro de los encargos fundamentales que recoge el Estatuto de la AESIA es la elaboración de guías operativas para la implementación de los requisitos del Reglamento de IA. Estas guías tienen como objetivo facilitar a las or-

ganizaciones la adaptación a la normativa, garantizando que las políticas y procedimientos incorporen principios de equidad y transparencia. En esta misma línea, el art. 10.1.c) del Estatuto de la Agencia menciona de manera expresa la necesidad de apoyar el desarrollo de sistemas de IA con perspectiva de género, integrando el principio de igualdad de oportunidades entre mujeres y hombres en su diseño y ejecución, e impulsando evaluaciones de impacto que identifiquen y eliminen posibles sesgos discriminatorios.

Además, la Agencia asume un papel relevante en la formación y asesoramiento a entidades públicas y privadas, promoviendo prácticas responsables y concienciando sobre los beneficios y los riesgos que plantea la inteligencia artificial.

Para que la AESIA pueda cumplir con éxito sus objetivos, es imprescindible dotarla de recursos suficientes y de la capacidad necesaria para ejercer labores de control, evaluación y sanción frente a casos de discriminación algorítmica. Además, resulta fundamental implementar auditorías y procesos de certificación de algoritmos, los cuales podrían inspirarse en modelos ya existentes, como las auditorías salariales (Morondo et al., 2023). Estas certificaciones podrían establecerse como un requisito previo para acceder a fondos públicos o participar en contratos con la Administración, siguiendo prácticas similares a las que actualmente se aplican en materia de igualdad.

Por otra parte, la lucha contra la discriminación algorítmica exige una coordinación estrecha entre distintos agentes públicos. La AESIA debe servir como eje central en este proceso, colaborando con organismos ya encargados de velar por la igualdad y la no discriminación a los que nos hemos referido en el capítulo 2º. Asimismo, se debe fomentar un diálogo constante con asociaciones, fundaciones y otros grupos de interés que aporten una perspectiva humanista y crítica sobre los posibles impactos negativos de los sistemas de IA en los derechos fundamentales.

*Capítulo 7*

# *Regulación de los productos y dispositivos sanitarios que incorporan IA*

En este capítulo, abordaremos la normativa sectorial relativa a los productos sanitarios de software, centrada en el Reglamento de Productos Sanitarios (RPS) y el Reglamento de Productos Sanitarios para Diagnóstico In Vitro (RDIV), en su caso.

Analizaremos también, los requisitos añadidos para que los productos sanitarios con software puedan introducirse en el mercado, con especial atención a los requisitos que garantizan la equidad de los aplicativos sanitarios que se prevén en el Reglamento del Espacio Europeo de Datos Sanitarios (REEDS) y en el Reglamento de evaluación de tecnologías sanitarias (REETS).

Finamente, nos referiremos a la responsabilidad de los organismos encargados de evaluar la conformidad.

## 7.1. EL REGLAMENTO DE PRODUCTOS SANITARIOS

A lo largo del capítulo anterior, hemos examinado el marco regulador de los sistemas de inteligencia artificial, incluyendo los requisitos esenciales establecidos en el Reglamento de Inteligencia Artificial. Hemos observado que, en el caso de los productos sanitarios que incorporan IA, el Reglamento se remite a la normativa específica aplicable a estos productos, pero también deja claro que los requisitos esenciales del RIA son siempre aplicables.

La normativa específica que regula los productos sanitarios es el Reglamento (UE) 2017/745 de Productos Sanitarios (RPS) y el Reglamento (UE) 2017/746 sobre productos sanitarios para diagnóstico in vitro (RPS-DIV). El primero tiene como objetivo garantizar que los productos sanitarios comercializados en la Unión Europea cumplan con estrictas normas de seguridad y eficacia para proteger la salud de pacientes y usuarios. A su vez, el Reglamento (UE) 2017/746 sobre productos sanitarios para diagnóstico in vitro (RDIV) regula específicamente los dispositivos destinados

al análisis de muestras biológicas humanas, como sangre, tejidos u orina. Ambas normas se encuadran en el Nuevo Marco Legislativo (NML) de la UE, al que nos hemos referido en el capítulo 5º. Debe tenerse en cuenta así mismo el Reglamento (UE) 2024/1860 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se modifican los Reglamentos (UE) 2017/745 y (UE) 2017/746 en lo que respecta a la aplicación gradual de Eudamed, la obligación de informar en caso de interrupción o cese en el suministro y las disposiciones transitorias para determinados productos sanitarios para diagnóstico in vitro. Y, en el caso español, el Real Decreto 192/2023, de 21 de marzo, por el que se regulan los productos sanitarios.[4]

Al igual que el RIA, el RPS adopta un enfoque basado en el riesgo, lo que significa que los requisitos aplicables son proporcionales al nivel de peligrosidad de cada producto. De esta manera, los productos de mayor riesgo, como los clasificados en las clases IIa, IIb y III, están sujetos a evaluaciones clínicas y técnicas más estrictas, con el fin de garantizar un escrutinio adecuado a los posibles riesgos que presentan.

Una de las novedades más importantes que aporta el RPS es la introducción de mecanismos para mejorar la trazabilidad y el control de los productos a lo largo de su ciclo de vida. El sistema de identificación única de dispositivos (UDI) facilita la vigilancia y el seguimiento de los productos sanitarios, mientras que las exigencias de gestión de calidad, conforme a la norma ISO 13485 + ISO 42001 aseguran que los fabricantes implementen procesos que garanticen el cumplimiento de los requisitos regulatorios. Además, el Reglamento establece obligaciones claras para fabricantes, importadores y distribuidores, quienes deben asumir responsabilidades tanto en la fase de diseño y producción como en la comercialización y uso del producto.

En cuanto a los plazos de entrada en vigor de los mecanismos de evaluación de la conformidad de productos de riesgo alto, se establecen plazos transitorios para permitir a los fabricantes adaptarse a los requisitos del nuevo Reglamento asegurando al mismo tiempo la disponibilidad continua de estos productos en el mercado. El Reglamento 2023/607 alargó dichos plazos, que varían según la clase del producto *legacy* de que se trate. Los certificados de clase III y productos implantables de clase IIb, serán válidos hasta el 31 de diciembre de 2027. Los de clase IIb (no implantables),

---

4 Actualmente, se está trabajando en el proyecto de un nuevo Real Decreto para regular específicamente los productos sanitarios para diagnóstico in vitro, con lo que se completarán los requisitos nacionales para esta categoría de productos.

clase IIa y productos de clase I estériles o con función de medición serán válidos hasta el 31 de diciembre de 2028. Estos plazos no coinciden con los previstos en el art. 113 del RIA, por lo que habrá que optar por una u otra norma, que en este caso deberá ser la *lex specialis* de los Reglamentos sectoriales 2017/745 y 2017/746.

Veamos a continuación con más detalle los requisitos que deben cumplir los productos sanitarios para poder ser introducidos en el mercado, según la normativa prevista en el Reglamento de Productos Sanitarios y el Real Decreto que lo desarrolla, el Real Decreto 192/2023, de 21 de marzo, por el que se regulan los productos sanitarios.

### *a) Calificación de los programas informáticos como productos sanitarios*

El RPS abarca dentro de su ámbito de aplicación cualquier dispositivo destinado a fines médicos en humanos, exceptuando los medicamentos. Incluye productos para diagnóstico, tratamiento, apoyo a la concepción y aquellos utilizados en limpieza, desinfección o esterilización. Además, regula productos sin finalidad médica listados en el Anexo XVI, como implantes mamarios, rellenos faciales o equipos de depilación láser. Por su parte, el RDIV se centra en dispositivos diseñados para analizar muestras biológicas, como reactivos, analizadores y contenedores para muestras, así como en sus accesorios necesarios o complementarios (por ejemplo, electrodos de ECG o colgadores de bolsas de suero). Aunque existen productos regulados por el RDIV que integran IA para analizar datos complejos de muestras biológicas, en este trabajo nos centraremos en las disposiciones del RPS, dado que la mayoría de los productos sanitarios que incorporan IA se enmarcan dentro de este Reglamento, haciendo referencia cuando proceda a la regulación de los productos *in vitro*.

En general, siempre que se plantee la posibilidad de que una aplicación informática específica puede tener una finalidad médica, debe seguirse una metodología de calificación y clasificación del producto. En primer lugar, debe precisarse qué uso tendrá la información de salida del software en el contexto médico o clínico. A continuación, debe establecerse si la finalidad prevista entra dentro del ámbito de aplicación del RPS. Si es el caso, habrá que aplicar las reglas de clasificación del producto contenidas en el Anexo VIII del RPS. El proceso de calificación y clasificación se documentará en un informe de clasificación que servirá para justificar las decisiones tomadas.

Cabe destacar que la calificación del producto como un producto sanitario y su subsiguiente clasificación será por su aplicación clínica según la finalidad prevista definida, independientemente de que se base en un sistema de IA.

Para comenzar con la calificación del producto, hay que subrayar que un programa informático será calificado como un producto sanitario, si el mismo debe ser utilizado con un fin médico conforme al art. 2.1 del RPS y al (MDCG 2019-11, p. 5):

> «[*p*]*roducto sanitario es todo instrumento, aparato, dispositivo, programa informático, implante, reactivo, material u otro art. destinado por el fabricante a ser utilizado, solo o en combinación, en seres humanos para uno o varios de los siguientes fines médicos específicos:*
>
> *— diagnóstico, prevención, seguimiento, predicción, pronóstico, tratamiento o alivio de enfermedades,*
>
> *— diagnóstico, seguimiento, tratamiento, alivio o compensación de una lesión o discapacidad,*
>
> *— investigación, sustitución o modificación de la anatomía o de un proceso o estado fisiológico o patológico,*
>
> *— que proporcione información mediante el examen in vitro de muestras derivadas del cuerpo humano, incluidas las donaciones de órganos, sangre y tejidos, y que no ejerza su acción principal prevista por medios farmacológicos, inmunológicos o metabólicos, en o sobre el cuerpo humano, pero que pueda ser asistida en su función por dichos medios*».

Así pues, los programas informáticos pueden considerarse productos sanitarios con el nuevo Reglamento. Esta cuestión resultaba dudosa a la luz de las Directivas anteriores (Directiva 93/42/CEE, art. 1.2.a), que no contemplaban el software como tal. La amplitud de los fines médicos definidos en el punto 1 del art. 2 es garantía de que prácticamente todo sistema de IA o solución informática que pueda ser utilizada para el diagnóstico, prevención, seguimiento, prognosis o tratamiento sea calificado como producto sanitario.

Quedan fuera de dicha clasificación los aplicativos de bienestar o entrenamiento. Estos aplicativos, que incluyen herramientas de seguimiento de la salud y el bienestar, están contemplados en el proyecto de Reglamento sobre el Espacio Europeo de Datos de Salud (EEDS), que establece normas específicas para los aplicativos de bienestar y detalla las obligaciones de los proveedores en cuanto a la protección de datos, interoperabilidad y la seguridad.

Para determinar si un sistema de software de inteligencia artificial debe clasificarse como producto sanitario o no, resulta útil la guía publicada por

el Grupo de Coordinación de Dispositivos Médicos de la UE con respecto al software sanitario (MDCG, 2019-11, 2021-24).[5]

En los casos en que se utilicen productos de software cuya calificación resulte dudosa, es recomendable solicitar a la autoridad competente una aclaración formal para confirmar que no se trata de un producto sanitario, en nuestro caso, a la Agencia Española de Medicamentos y Productos Sanitarios (AEMPS). Si persisten dudas sobre la calificación o clasificación de un producto, la AEMPS dispone de un servicio de consultas que permite a fabricantes y patrocinadores resolver estas incertidumbres (AEMPS, 2022).

Teniendo en cuenta que todo sistema de IA es un programa informático, esta guía es aplicable a la calificación y clasificación de cualquier sistema de IA. Nótese que la diferencia entre un programa informático que pueda ser calificado como IA y uno que no, será relevante de cara a la aplicación del Reglamento de IA, pero no de los Reglamentos de Productos Sanitarios.

---

[5] La Guía MDGC, 2019 contribuye a la calificación y clasificación del software sanitario como producto sanitario según la Regla 11 del Anexo VIII del RPS, aunque no sea esta la única Regla aplicable a dichos productos. Por su parte, el MDCG, 2021 ofrece ejemplos actualizados de programas informáticos calificados como productos sanitarios. Por ejemplo, "Software de ayuda a la toma de decisiones", en general, se trata de herramientas informáticas que combinan bases de datos y algoritmos de información médica general con datos específicos de cada paciente. Su objetivo es proporcionar a los profesionales sanitarios y/o usuarios recomendaciones para el diagnóstico, pronóstico, seguimiento y tratamiento de pacientes concretos. (Por ejemplo, los sistemas de planificación de fármacos (como la quimioterapia) están destinados a calcular la dosis de fármaco que debe administrarse a un paciente concreto y, por tanto, se califican como productos sanitarios). "Sistemas de información": Los sistemas de información destinados únicamente a transferir, almacenar, convertir, formatear o archivar datos no se consideran productos sanitarios en sí mismos. Sin embargo, pueden utilizarse con módulos adicionales que pueden calificarse por derecho propio como productos sanitarios (MDSW). (Por ejemplo, Sistema de electrocardiógrafo (ECG) prehospitalario: Un sistema de gestión de ECG prehospitalario es un sistema basado en software destinado a los servicios de ambulancia para almacenar y transferir información de los pacientes a un médico en una ubicación remota. Normalmente, el sistema contiene información sobre la identificación del paciente, sus parámetros vitales y otras observaciones clínicas documentadas. Estos sistemas de electrocardiógrafo (ECG) prehospitalario no están calificados como productos sanitarios). "Comunicación": Se califica como producto sanitario (MDSW) un módulo de software que genera alarmas basadas en la monitorización y análisis de parámetros fisiológicos específicos del paciente. Sistemas web para la supervisión de datos: Los módulos destinados a supervisar el rendimiento médico de los productos sanitarios entran en el ámbito de aplicación de la normativa sobre productos sanitarios.

En relación al software, a menudo se plantea la cuestión de si el software que es un mero componente que forma parte de un producto más amplio hace del conjunto un producto sanitario o no. El TJUE (sentencia de 7 de diciembre de 2017, Asunto C-329/16, TOL9.913.342) resolvió al respecto que los programas informáticos pueden constar de varios módulos diferenciados, algunos de los cuales tendrán una finalidad médica y otros no. En ese caso, solo los módulos con finalidad médica entrarán en el ámbito de aplicación del RPS. Es responsabilidad del fabricante identificar los límites e interfaces de los diferentes módulos, que deben estar claramente identificados por el fabricante y basados en el uso que se hará del producto.

### *b) Clasificación de productos sanitarios de software*

La clasificación del riesgo del Anexo VIII depende de la finalidad del producto y del impacto potencial que sus decisiones puedan tener sobre la salud humana, lo que genera un marco exigente tanto para fabricantes como para organismos notificados.

Conforme al Reglamento de Productos Sanitarios, estos se clasifican en cuatro categorías o clases según su riesgo intrínseco: I, Is, Im, Ir, IIa, IIb y III (mientras que, para los productos regulados por el Reglamento (UE) 2017/746 de diagnóstico in vitro, las clases son A, B, C y D).

El fabricante, al solicitar una investigación clínica (IC) con un producto sanitario, debe presentar a las Autoridades Sanitarias (AASS) una Declaración de Conformidad con los requisitos generales de seguridad y funcionamiento del RPS o del RDIV, exceptuando aquellos requisitos que sean objeto del resultado de la propia investigación.

En cuanto a los programas informáticos, la Regla 11 del Anexo VIII del RPS establece que los softwares diseñados para proporcionar información utilizada en la toma de decisiones con fines diagnósticos o terapéuticos se clasifican, como norma general, en clase IIa. Sin embargo, si las decisiones derivadas del uso del software pueden tener un impacto más grave, la clasificación aumenta si existe la posibilidad de causar la muerte o un deterioro irreversible del estado de salud, el software se clasifica en clase III; si el impacto puede derivar en un deterioro grave de la salud o una intervención quirúrgica, la clasificación será clase IIb.

Por otro lado, los softwares destinados a supervisar procesos fisiológicos también se clasifican en clase IIa, salvo que supervisen parámetros vitales cuya variación pueda generar un peligro inmediato para el paciente, en cuyo caso la clasificación será IIb.

Finalmente, todos los demás programas informáticos se clasifican en clase I. Cabe remarcar que la clasificación de un SIA médico como clase I va asociada a que éste no proporciona información que se utiliza para tomar decisiones con fines terapéuticos o de diagnóstico y por lo tanto es necesario establecer claramente este punto a la justificación de la clasificación.

La práctica demuestra que la gran mayoría de los softwares destinados al diagnóstico, tratamiento, seguimiento o predicción de enfermedades proporcionan información relevante para la toma de decisiones terapéuticas o asistenciales. En la mayoría de los casos, estos productos se clasifican como clase IIa o superior, lo que conlleva procedimientos complejos de certificación con organismos notificados. Este proceso no solo aumenta los costes de producción y certificación, sino que también impone un escrutinio más riguroso sobre la seguridad y el rendimiento del software.

La falta de claridad en la Regla 11, especialmente al considerar cualquier posible consecuencia para la salud humana sin atender a su probabilidad, ha suscitado críticas (Rühlicke, 2024). Al incluir todas las posibles causas de daño, incluso las más remotas o improbables, la regla puede interpretarse como la imposición de un sistema de responsabilidad objetiva que limita la innovación en software sanitario. Estas críticas han sido recogidas en la guía MDCG 2019-11, la cual defiende una clasificación basada en la probabilidad del daño, no solo en su mera posibilidad (MDCG, 2019, p. 8). Por su parte el *Forum Internacional de Reguladores de Dispositivos Médicos* (IMDRF por sus siglas en inglés) propuso un marco para la categorización de riesgos del software como dispositivo médico que ha contribuido a clarificar la terminología utilizada en la clasificación de riesgos que puede contribuir a la normalización del procedimiento, aunque no se trata de una regulación vinculante (IMDRF, 2014).

### *c) Evaluación de la conformidad*

Todos los productos sanitarios deben someterse a un procedimiento de evaluación de la conformidad antes de su uso rutinario con pacientes y deben llevar el marcado CE. Exceptuando los productos de más bajo riesgo (Clase I del MDR y Clase A del RDIV) y los productos de desarrollo "in-house".

La evaluación de la conformidad incluye la intervención de una entidad evaluadora externa denominada Organismo Notificado (ON), que se designa y se supervisa por las Autoridades Competentes (AACC) de cada país.

Una vez demostrada su competencia técnica específica, se incluyen en el listado NANDO que mantiene la Comisión Europea.

Asimismo, la organización responsable que se identifica como el fabricante del producto debe tener una licencia de establecimiento de fabricación de productos sanitarios emitida por la AEMPS y debe registrar el producto en la base de datos EUDAMED gestionada por las AACC tanto a nivel nacional como europeo.

### *d) Evaluación clínica e investigación clínica en el contexto de la RPS*

El Reglamento (UE) 2017/745 regula la evaluación clínica y la investigación clínica como partes fundamentales del proceso de evaluación de la conformidad. La evaluación clínica, que corresponde al análisis de datos para verificar la seguridad y el rendimiento del producto, puede incluir, cuando sea necesario, la realización de una investigación clínica. Esta última consiste en un estudio sistemático con participación de sujetos humanos, orientado a demostrar el cumplimiento de los requisitos generales de seguridad y funcionamiento definidos en el Anexo I del RPS.

Según el art. 62 del RPS, las investigaciones clínicas deben diseñarse y llevarse a cabo conforme a lo establecido en los arts. 63 a 80 y en el Anexo XV del RPS. Su finalidad es múltiple: verificar que el producto, en condiciones normales de uso, cumple con la finalidad prevista por el fabricante y alcanza las prestaciones especificadas; establecer las ventajas clínicas del producto; y evaluar su seguridad clínica y determinar los posibles efectos secundarios, asegurando que los riesgos son aceptables en relación con los beneficios previstos.

El ámbito de aplicación de las investigaciones clínicas depende de la clasificación del producto. En principio, la realización de una investigación clínica es obligatoria para los productos de clase III y los productos invasivos, como establece el art. 61.4. Para productos de clase IIa y IIb, incluida la mayoría del software sanitario, la conformidad puede basarse en datos clínicos suficientes que demuestren la seguridad y las prestaciones del producto, sin necesidad de realizar una investigación clínica completa, siempre que el fabricante justifique el nivel de evidencia requerido en función de la finalidad prevista del dispositivo (art. 61.1). La profundidad y el nivel de evidencia que debe obtenerse se refiere al cumplimiento de los requisitos aplicables, en concreto, su seguridad y su funcionamiento clínico, varía en función del nivel de riesgo del producto. Por ejemplo, los productos de Clase III, al representar el mayor riesgo, están sujetos a controles mucho más estrictos y exhaustivos.

En este contexto, el fabricante dispone de dos opciones para cumplir con los requisitos de evaluación clínica: realizar una investigación clínica, que debe ajustarse a los procedimientos establecidos en los arts. 62 a 80 del RPS, o bien aportar datos clínicos relativos a un producto sanitario equivalente, conforme a lo indicado en el art. 61.3 y en el Anexo XIV, A, 3. Habitualmente, esta justificación se basa en demostrar la equivalencia o similitud con otro producto preexistente con una finalidad parecida, sobre todo en el contexto clínico de que se trata.

Si los datos utilizados para la investigación no se utilizarán en una evaluación de conformidad posterior (art. 82 del RPS), únicamente se aplicarán los apartados 2 y 3 del art. 62; las letras b, c, d, f, h y l del apartado 4; el art. 62.6; así como la legislación nacional aplicable.

Por lo tanto, el promotor o el fabricante de un producto debe planificar y justificar el proceso de evaluación clínica y, si es necesario, llevar a cabo una investigación clínica para demostrar el cumplimiento de los requisitos regulatorios.

Si bien para los productos más complejos, como los de clase III, la investigación clínica resulta obligatoria, para otros productos de menor riesgo, incluidos los de clase IIa, el fabricante puede no necesitar una investigación clínica que valide el sistema de software o, en su caso, optar por utilizar datos equivalentes que sustenten la evidencia clínica. En todos los casos, la transparencia y la trazabilidad de los datos clínicos son exigencias esenciales, supervisadas a lo largo del ciclo de vida del producto.

### *e) Requisitos de la investigación clínica para evitar el sesgo en productos con IA*

En el apartado anterior hemos visto que demostrar la conformidad de productos sanitarios basados en software con inteligencia artificial clasificados como III o, en determinados casos, también los productos IIb, requiere llevar a cabo una investigación clínica en los términos de los arts. 62 a 80 del RPS, además de aquellos recogidos en el Anexo XV. Estos requisitos garantizan la validez científica, la protección de los derechos de los participantes y la mitigación de sesgos, contribuyendo así al derecho fundamental a no ser discriminado.

El primer aspecto reseñable reside en el diseño y la ejecución ética y científica de la investigación. El art. 62.3 establece que la seguridad, la dignidad y el bienestar de los sujetos de investigación deben prevalecer sobre cualquier otro interés, asegurando, al mismo tiempo, que los datos clínicos generados sean científicamente válidos, fiables y sólidos. Este principio

obliga a realizar una revisión ética y científica rigurosa antes de iniciar cualquier estudio, especialmente cuando se trata de software con IA, donde las implicaciones pueden ser complejas e imprevisibles.

La protección de poblaciones vulnerables, conforme al art. 62.4.d, requiere implementar medidas específicas para resguardar los derechos de sujetos como menores, mujeres embarazadas, personas inmunodeficientes y ancianos. Estas poblaciones, al ser más susceptibles a riesgos y discriminaciones, deben ser consideradas cuidadosamente en el diseño de la investigación para evitar sesgos sistemáticos en los resultados.

El consentimiento informado, descrito en el art. 62.4.f, constituye otro pilar fundamental de estas evaluaciones. Los sujetos o sus representantes legales deben recibir información clara, comprensible y transparente sobre el propósito, riesgos y beneficios de la investigación, garantizando así una participación consciente y voluntaria.

La salvaguarda de derechos, según el art. 62.4.h, obliga a proteger la integridad física y psíquica, así como la intimidad y los datos personales de los participantes en investigaciones con software sanitario, donde la recopilación y el tratamiento de grandes volúmenes de datos pueden presentar riesgos de privacidad si no se gestionan adecuadamente.

La mitigación de sesgos en la investigación clínica requiere medidas adicionales, como las establecidas en el Anexo XV, apartado 3.6.4, que obliga a implementar una distribución aleatoria en los ensayos y gestionar de forma proactiva los factores de discriminación. El objetivo es minimizar los sesgos en la selección, asignación y análisis de los datos, garantizando así la robustez y fiabilidad de los resultados obtenidos.

Un aspecto fundamental consiste en asegurar la representatividad de la población (RPS, Anexo XV, apartado 3.6.3). Para evitar sesgos discriminatorios, es imprescindible que la población sometida a investigación sea representativa del grupo objetivo, asegurando una adecuada diversidad en términos de edad, género, etnia y condiciones de salud. Esta representatividad es crítica en el desarrollo de software con IA, ya que datos no representativos pueden dar lugar a algoritmos sesgados que perpetúan desigualdades.

La transparencia y trazabilidad, conforme al Anexo XV, apartado 3.2, requieren una identificación y descripción claras del producto, su finalidad prevista, incluida la población a la que va dirigida y su fabricante. Este requisito facilita el análisis independiente de los datos y asegura la integridad del proceso de evaluación.

Además, el art. 62.4.e exige una evaluación continua de los riesgos y beneficios, que debe llevarse a cabo durante toda la investigación. Es necesario monitorizar de forma constante que los beneficios esperados justifiquen los riesgos y los inconvenientes previsibles, garantizando así que el estudio se mantiene dentro de los límites éticos y científicos adecuados.

Por último, la calificación del personal desempeña un papel esencial. Según el art. 62.6, el investigador principal debe ser un profesional acreditado, con conocimientos científicos y experiencia suficiente en la atención al paciente y en la metodología de la investigación clínica. Todo el equipo involucrado debe contar con formación adecuada y estar cualificado para llevar a cabo sus tareas de manera rigurosa y conforme a los estándares establecidos.

No hay duda de que el marco establecido para las investigaciones clínicas con participación física de sujetos humanos por el Reglamento sobre productos sanitarios es un marco riguroso. Sin embargo, en nuestra opinión, estos requisitos resultan insuficientes para estudios basados únicamente en datos, como es el caso de los sistemas de inteligencia artificial utilizados en productos sanitarios. Como sabemos, la investigación clínica con IA plantea problemas específicos relacionados con la transparencia, la validación y la mitigación de sesgos, que exigen ser complementados con las disposiciones previstas en el Reglamento de Inteligencia Artificial.

Por el juego de las remisiones normativas al que hemos hecho referencia en el capítulo anterior, cualquier sistema de IA que se utiliza con fines de diagnóstico y se califica como producto sanitario (clase IIa o superior en virtud del RPS), se considerará un sistema de alto riesgo según el RIA. En estos casos, el fabricante va a tener que demostrar que el sistema cumple con los requisitos obligatorios establecidos en los arts. 9 a 15 del RIA, relacionados con la fiabilidad, la transparencia y la ausencia de sesgos. El cumplimiento simultáneo de estos requisitos y de los previstos en el RPS es indispensable no solo para la aprobación del producto, sino también para determinar si el sistema podría ser considerado defectuoso.

Además, las normas de seguridad y ciberseguridad tienen un peso específico en la apreciación de un producto de IA como defectuoso. Anotemos que de acuerdo con el art. 7.2.f de la Directiva sobre Responsabilidad por los Daños Causados por Productos Defectuosos, como se verá en el capítulo 9º, los requisitos de seguridad pertinentes, incluidos los de ciberseguridad se consideran circunstancias relevantes para valorar el carácter defectuoso de un sistema. En este mismo sentido, se ha pronunciado recientemente la

En este sentido, creo que la evaluación de sistemas de inteligencia artificial utilizados con fines médicos debe considerar tanto las exigencias del RPS como los requisitos específicos del RIA, especialmente en relación a la apreciación de la responsabilidad derivada de productos defectuosos. La integración de los tres marcos regulatorios no solo permite un control más robusto de la conformidad de los productos sanitarios, sino que también refuerza la protección del paciente y del usuario. Al aplicar las disposiciones del RIA en conjunción con el RPS, se consigue un nivel de seguridad equivalente, adaptado a las particularidades que presentan los sistemas de IA en el ámbito médico.

El *Grupo de Coordinación de Productos Sanitarios* (MDCG, según sus siglas en inglés) creado por el Reglamento 2017/745 prepara actualmente una Guía para la evaluación de productos sanitarios de software con IA en la que se aclarará las dudas que plantea la aplicación conjunta de ambos marcos regulatorios en cuanto a los requisitos requeridos de certificación. En el capítulo siguiente nos referiremos en concreto a la coordinación de los requisitos referentes a la calidad de los datos y la necesidad de demostrar la neutralidad o no discriminación de los productos.

### *f) Requisitos para el marcado CE*

El marcado CE de un producto sanitario se obtiene mediante un proceso de evaluación por parte de un organismo notificado o bien para los de bajo riesgo (clases I o A) por el propio fabricante.

EUDAMED será la base de datos donde se gestionarán las IC realizadas en Europa e incluye un módulo especifico de Investigación Clínica y de Evaluación del Funcionamiento.

De particular relevancia a estos efectos es la aplicación en España del Real Decreto 192/2023, de 21 de marzo, por el que se regulan los productos sanitarios, que en su art. 30 establece que se aplicarán a las investigaciones clínicas realizadas para demostrar la conformidad de los productos, además de lo establecido en el capítulo VI y Anexos XIV y XV del RPS, los principios éticos, metodológicos y de protección de los sujetos del ensayo, contemplados en el Real Decreto 1090/2015, de 4 de diciembre, por el que se regulan los ensayos clínicos con medicamentos.

### *g) Organismos notificados de evaluación de la conformidad*

Los organismos notificados autorizados conforme al Reglamento (UE) 2017/745 sobre productos sanitarios tienen la facultad de controlar la con-

formidad de un producto sanitario con IA siempre que cumplan con los requisitos establecidos en la sección 2 del Capítulo III del RIA. Como se ha dicho más arriba, aunque a un sistema de IA se le vaya a aplicar también en el futuro los requisitos del Reglamento de IA, en realidad esta vía normativa será igualmente aplicable en relación a la certificación. Por tanto, los requisitos de evaluación establecidos en la sección 2 del RIA y los apartados 4.3, 4.4, 4.5 y 4.6 del Anexo VII forman parte del proceso de control.

Como dispone el art. 43.3 del RIA, corresponde a los organismos notificados verificar la conformidad del producto en estos aspectos, siempre que se cumplan las exigencias del art. 31. Para ejercer esta función, dichos organismos deben cumplir las disposiciones del art. 31 del RIA, que exigen contar con personal cualificado en áreas administrativas, técnicas, jurídicas y científicas. En particular, se requiere experiencia específica en sistemas de inteligencia artificial, procesamiento de datos y en los requisitos técnicos aplicables.

Por tanto, los fabricantes españoles de productos sanitarios de clase III, IIb, IIa, Is, Ir y Im, así como de productos sanitarios para diagnóstico in vitro clases D, C, B, y A estéril, antes de colocar el marcado CE, tienen que elegir un organismo notificado con el que quieren certificar sus productos y seguir un procedimiento de evaluación de la conformidad que resulte de aplicación.

El organismo español designado para evaluar la conformidad de los productos sanitarios y productos sanitarios de diagnóstico in vitro, para que estos puedan comercializarse dentro de la Unión Europea es el Centro Nacional de Certificación de Productos Sanitarios adscrito a la AEMPS (CNCps). El CNCps, Organismo Notificado 0318, es el único organismo notificado en España designado por el Ministerio de Sanidad para la evaluación de la conformidad de productos sanitarios según el Reglamento 2017/745, de 5 de abril y el Reglamento 2017/746.

Además, el Organismo Notificado 0318 mantiene la designación para las directivas de productos sanitarios y de productos sanitarios para diagnóstico in vitro (Directiva 93/42/CE y Directiva 98/79/CE, respectivamente), que, sobrepasada la fecha de aplicación de nuevos reglamentos de productos sanitarios, las actuaciones se limitan al seguimiento de los certificados CE ya emitidos, que incluyen únicamente auditorías de seguimiento y modificaciones no significativas de los productos certificados.

El alcance y estado de la designación del Organismo Notificado 0318 puede ser consultado en la página web del registro NANDO la Comisión Europea.

*h) Supervisión postcomercialización*

Tras someter sus productos al procedimiento de evaluación por parte de un organismo notificado y después de su comercialización, los productos sanitarios de alto riesgo están sometidos a inspecciones del producto tras su introducción en el mercado.

Para ello el fabricante debe implementar un plan de vigilancia conforme al RPS en el que podrá integrar los elementos del art. 72 del RIA. Esta integración permitirá al producto sanitario con IA alcanzar un nivel de protección equivalente al previsto en el RIA, facilitando así la alineación entre ambos Reglamentos. De esta manera, se asegura que los procesos de vigilancia incorporen los requisitos relativos a la seguridad y fiabilidad de los sistemas de IA, garantizando una supervisión continua tras su introducción en el mercado o puesta en servicio.

Una novedad a destacar de los Reglamentos RPS y RDIV es el establecimiento en el anexo III de la información mínima que debe incluirse en el seguimiento postcomercialización. Esta información debe formar parte de la documentación técnica, la cual debe actualizarse de manera continuada para asegurar el cumplimiento con la normativa vigente.

## 7.2. EL REAL DECRETO 192/2023, DE 21 DE MARZO POR EL QUE SE REGULAN LOS PRODUCTOS SANITARIOS

Los requisitos de los reglamentos se aplican por igual en todo el ámbito de la EEE, pero cada Estado miembro puede regular ciertos aspectos reglamentarios adicionales (por ejemplo, el idioma del etiquetado de los productos, las licencias de establecimientos o los requisitos relacionados con la fabricación intrahospitalaria), bien porque no se desarrollan en el reglamento o bien porque la legislación previa nacional así lo contempla. En el caso español, se promulgó el Real Decreto 192/2023, de 21 de marzo por el que se regulan los productos sanitarios regula las evaluaciones e investigaciones clínicas con productos sanitarios.

En su art. 2.2. se dice que "investigación clínica con producto sanitario" es cualquier investigación sistemática en uno o más sujetos humanos con objeto de evaluar la seguridad o las prestaciones de un producto.

El art. 30.1 del Reglamento de Productos Sanitarios (RPS) prevé que, en la realización de investigaciones clínicas con productos sanitarios incluidos en el art. 1 del RPS, se aplicarán los principios éticos, metodológicos y de protección de los sujetos del ensayo establecidos en el Real Decreto 1090/2015, que regu-

la los ensayos clínicos con medicamentos. Estos requisitos deben adaptarse a los productos sanitarios basados en software, los cuales se desarrollan a partir de datos y no requieren intervención física en el paciente (ANCEI, 2021).

En concreto, los productos destinados a la investigación clínica solo podrán ser puestos a disposición de los facultativos médicos si la investigación cuenta con el dictamen favorable del Comité de Ética de la Investigación con medicamentos (CEIm) y la conformidad del centro donde se realizará la investigación clínica. Los CEIm son responsables de velar por la representatividad de los datos clínicos conforme al mandato del Real Decreto 1090/2015, que exige que los datos estén libres de sesgo y sean suficientemente representativos.

Una vez obtenido el visto bueno del CEIm el promotor solicitará la autorización correspondiente a la Agencia Española de Medicamentos y Productos Sanitarios, junto con la documentación exigida en el RPS. Por tanto, a diferencia de la investigación que se lleva a cabo bajo la Ley de investigación biomédica, la investigación clínica con producto sanitario requerirá autorización del AEMPS.

La solicitud, el manual del investigador, el plan de investigación clínica, el consentimiento informado y las instrucciones y etiquetado del producto en investigación deberán acompañar a la solicitud.

Por otra parte, será la propia AEMPS, independientemente de las competencias de otras autoridades sanitarias, la que decidirá sobre la aplicación de las definiciones y los criterios de clasificación a los que nos hemos referido anteriormente.

El Ministerio de Sanidad es la autoridad responsable de los organismos notificados encargados de llevar a cabo la evaluación de conformidad de los productos sanitarios con vistas al marcado CE. El Ministerio designará los organismos que efectuarán los procedimientos citados en el RPS, tras evaluar su aptitud en orden a su designación y para verificar el mantenimiento de estas aptitudes en el tiempo (art. 16). Nótese que en este caso la competencia de la AESIA cede ante esta *lex specialis*.

Una vez obtenida la certificación de la evaluación de conformidad por parte del organismo notificado autorizado, los agentes económicos, antes de la introducción del producto en el mercado deberán registrarse en el Registro de Comercialización de la AEMPS.

Tras la comercialización el producto sanitario los agentes deberán mantener un registro documentado de los productos que pongan a disposición en territorio español con el fin de que estos sean trazables.

El Decreto puntualiza que, para garantizar la correcta utilización de los productos, los profesionales que los utilicen deberán estar debidamente cualificados y formados.

El Real Decreto también regula la fabricación de productos sanitarios por parte de hospitales para su propio y exclusivo uso. Los hospitales deben comunicar el inicio de esta actividad a la AEMPS y cumplir con los requisitos establecidos en el artículo 5.5 del Reglamento (UE) 2017/745. Esta fabricación no está dirigida a la comercialización y debe garantizar la seguridad y el adecuado funcionamiento de los productos.

## 7.3. EL CASO ESPECÍFICO DE LOS SISTEMAS DE APRENDIZAJE AUTOMÁTICO, LAS RECOMENDACIONES TRIPOD+AI

El Reglamento de Productos Sanitarios no se refiere explícitamente la comercialización de software sanitario basado en grandes modelos de lenguaje que incorporen aprendizaje profundo, sin embargo, constatamos que hoy en día no se están autorizando los productos sanitarios que incorporan una IA basada en aprendizaje. La Agencia Europea para el Medicamento (EMA) y la AEMPS han establecido una moratoria sobre estos productos hasta que se desarrollen las guías adecuadas para poder evaluarlos.

En concreto, con el fin de facilitar la evaluación y uso de algoritmos de aprendizaje en medicina, el **HMA-EMA Big Data Steering Group** (2023) ha elaborado un plan hasta 2028 con el fin de maximizar los beneficios del aprendizaje automático en la regulación de medicamentos, minimizando al mismo tiempo los riesgos asociados. El plan incluye la orientación sobre el uso de IA a lo largo del ciclo de vida de los medicamentos, así como el apoyo continuo a productos en desarrollo. Además, el documento referido declara querer proporcionar marcos para el uso de herramientas de IA que aumenten la eficiencia, mejoren la comprensión y el análisis de datos, y apoyen la toma de decisiones. También, en el marco de este plan, se diseñarán iniciativas para desarrollar la capacidad y la capacidad de respuesta de la red, los socios y las partes interesadas para mantenerse al día con los avances en el campo de la IA.

Por su parte, el **TRIPOD+AI Statement** (Collins et al., 2024) al que nos hemos referido en el capítulo segundo, ofrece una guía detallada para la elaboración de informes sobre estudios que desarrollan o evalúan modelos de predicción clínica basados en métodos de regresión o de aprendizaje automático.

La propia guía declara que la adopción de estas directrices mejora la adopción de tecnologías sanitarias basadas en algoritmos de aprendizaje automático, promoviendo una cultura de transparencia en la investigación, mejorando la calidad y la reproducibilidad de los estudios, y facilitando la evaluación crítica de los modelos predictivos en diversos contextos clínicos.

El propósito fundamental de TRIPOD+AI es garantizar que los proyectos y memorias de prototipos basados en LLM y algoritmos de aprendizaje automático sean completos, precisos y transparentes, lo que no solo facilita la replicabilidad de los estudios, sino que también fortalece la confianza en la validez de los modelos. Su estructura está diseñada para orientar tanto a quienes desarrollan modelos predictivos como a quienes evalúan su desempeño, destacando la importancia de documentar con rigor los métodos empleados, los datos utilizados y los resultados obtenidos.

En cuanto a los ítems requeridos por la guía, se contempla una lista de 27 elementos esenciales que abarcan desde el título del estudio hasta la discusión de los resultados. Entre los más relevantes se encuentra la descripción del contexto clínico y la justificación del estudio, donde se insta a explicar claramente el objetivo del modelo y su relevancia en la toma de decisiones clínicas. Asimismo, se hace hincapié en la transparencia respecto a la procedencia de los datos, los criterios de inclusión de los participantes y los métodos de análisis, incluyendo detalles sobre el manejo de datos y la validación interna del modelo.

En relación con la evolución respecto a la guía anterior (TRIPOD 2015), destaca la ampliación del alcance de la misma para incluir modelos basados en IA y aprendizaje automático, así como la incorporación de nuevos ítems relacionados con la participación de pacientes y la adopción de prácticas de ciencia abierta, como la compartición de datos y códigos de análisis. Además, se ha reforzado la evaluación del rendimiento del modelo en subgrupos clave, promoviendo una mayor sensibilidad hacia la equidad en salud.

Un aspecto destacado de esta guía consiste en la atención a la equidad o *fairness*, que se integra de forma transversal en varios ítems de la misma. Se recomienda informar explícitamente sobre cómo se ha abordado la representatividad de los datos, la existencia de posibles sesgos y las estrategias empleadas para mitigar desigualdades. Esto implica detallar si el rendimiento del modelo varía entre subgrupos poblacionales y qué medidas se han tomado para garantizar su equidad.

No es casual que la equidad constituya uno de los pilares fundamentales en la actualización del TRIPOD+AI Statement, reflejando la creciente preocupación por la equidad en el uso de modelos de predicción clínica, especialmente aquellos basados en inteligencia artificial y aprendizaje automático. La inclusión explícita de este concepto en la guía responde a la necesidad de garantizar que los modelos predictivos no perpetúen ni amplifiquen desigualdades existentes en el ámbito de la salud.

En el contexto de TRIPOD+AI, la *fairness* se define como la propiedad de los modelos de predicción que asegura la ausencia de discriminación injusta contra individuos o grupos basados en características demográficas, como la edad, el género, la raza o el estatus socioeconómico. La equidad no se limita a la representación de diversos subgrupos en los datos de entrenamiento, sino que abarca también cómo se diseñan, desarrollan, validan e implementan estos modelos en la práctica clínica. La falta de equidad puede traducirse en decisiones clínicas sesgadas que afecten negativamente a poblaciones vulnerables, perpetuando desigualdades en el acceso a diagnósticos, tratamientos y resultados en salud.

El TRIPOD+AI se refiere a la cuestión de la equidad de manera transversal a lo largo de toda la guía, destacándose en varios ítems del checklist que propone:

1. Contexto y justificación del estudio (Ítem 3c):

   Se recomienda que los autores describan cualquier desigualdad en salud conocida entre diferentes grupos sociodemográficos. Esto implica identificar, desde el planteamiento del estudio, si existen disparidades que puedan influir en la validez o la aplicabilidad del modelo en distintas poblaciones.

2. Fuentes de datos y representatividad (Ítems 5a y 7):

   En la descripción de los datos utilizados, se insta a detallar la representatividad de los mismos, especificando si abarcan adecuadamente la diversidad de la población objetivo. Además, se debe informar sobre cualquier preprocesamiento de datos y si este proceso se realizó de manera uniforme entre los diferentes grupos demográficos.

3. Definición del resultado y de los predictores (Ítems 8a, 8b y 9c):

   Aquí se enfatiza la necesidad de garantizar que la evaluación de los resultados y la medición de los predictores sean coherentes y equitativos entre los subgrupos poblacionales. Si el proceso de evaluación requiere interpretación subjetiva, es fundamental describir las cua-

lificaciones y características demográficas de los evaluadores, lo que puede influir en la aparición de sesgos.

4. Evaluación del rendimiento del modelo (Ítem 12f):

   Se recomienda reportar cómo se ha evaluado el rendimiento del modelo en diferentes subgrupos, considerando posibles variaciones en la precisión predictiva entre ellos. Esto contribuye a identificar si el modelo funciona de manera equitativa para todas las poblaciones relevantes.

5. Enfoques para abordar la fairness (Ítem 14):

   Este ítem se centra específicamente en la equidad, exigiendo que los autores describan cualquier enfoque adoptado para abordar la fairness en el desarrollo o la evaluación del modelo, así como la justificación de dichas estrategias. Esto puede incluir técnicas de mitigación de sesgos, como el reequilibrio de datos, la calibración diferencial o el uso de métricas de equidad para evaluar el desempeño del modelo en distintos grupos.

6. Resultados y análisis por subgrupos (Ítems 20b y 23a):

   En la sección de resultados, se debe presentar información desglosada por subgrupos relevantes, permitiendo evaluar si existen diferencias significativas en el rendimiento del modelo entre ellos. Se recomienda incluir intervalos de confianza para estas comparaciones y, si es pertinente, gráficos que faciliten la visualización de posibles disparidades.

7. Interpretación y limitaciones (Ítems 25 y 26):

   Finalmente, en la discusión, se espera que los autores reflexionen sobre los hallazgos relacionados con la *fairness*, incluyendo cómo las limitaciones del estudio podrían haber afectado la equidad del modelo. Esto abarca desde la falta de representatividad de los datos hasta posibles sesgos introducidos en el análisis.

La inclusión de estos ítems refleja un cambio de paradigma en la investigación de modelos predictivos. La equidad deja de ser un aspecto secundario para convertirse en un criterio central en la evaluación de la calidad de los estudios, puesto que la falta de *fairness* no solo tiene implicaciones éticas, sino también prácticas, ya que un modelo que funciona bien en una población, pero mal en otra puede generar errores clínicos costosos y perjudicar la salud de los pacientes.

Además, TRIPOD+AI subraya la importancia de la participación de pacientes y del público en la investigación (Ítem 19), fomentando un enfoque inclusivo que considera la diversidad de perspectivas en el diseño, la implementación y la interpretación de los modelos. Esta participación contribuye a identificar posibles fuentes de sesgo que podrían pasar desapercibidas desde una perspectiva exclusivamente técnica.

Finalmente, el apartado propositivo de la guía profundiza en la importancia de la transparencia en la investigación de modelos predictivos y en cómo TRIPOD+AI contribuye a mejorar la calidad de los estudios en este campo. Se subraya que, si bien la guía proporciona recomendaciones mínimas de reporte, su adopción puede tener un impacto significativo en la reproducibilidad de la investigación, la implementación de modelos en la práctica clínica y la confianza pública en la IA aplicada a la salud. Se destaca, además, la necesidad de actualizaciones periódicas para adaptarse a los avances tecnológicos, como los modelos de IA generativa, y se enfatiza la relevancia de la participación activa de diversos actores, incluidos pacientes y representantes del público, en la investigación y la validación de estos modelos.

Por su parte, el documento TRIPOD LLM (Gallifant et al., 2025) presenta una guía de reporte diseñada específicamente para prototipos que utilizan modelos de lenguaje en aplicaciones biomédicas. Es una extensión del documento anterior, TRIPOD+ AI, y aborda los retos que plantean los LLM en salud.

La guía comprende una lista de 19 elementos principales y 50 ítems de desarrollo que cubren aspectos diversos de la evaluación de estos sistemas. Además, la guía ofrece un formato modular que permite a los desarrolladores cuestionar la misma según el modelo de desarrollo que estén diseñando. El objetivo declarado de la declaración TRIPOD +LLM consiste en plantear respuestas a los desafíos de transparencia, reproducibilidad, interpretabilidad y neutralidad o falta de sesgo de estos grandes modelos de lenguaje en el ámbito de la salud.

## 7.4. RESPONSABILIDAD DE LOS ORGANISMOS ENCARGADOS DE EVALUAR LA CONFORMIDAD

La responsabilidad de los organismos notificados y de las autoridades encargadas de la vigilancia del mercado en el desempeño de sus funciones es un aspecto de gran importancia, particularmente en la etapa de vigilancia postcomercialización. Como se mencionó anteriormente, el modelo

europeo de certificación para productos de alto riesgo establece la necesidad de una evaluación previa llevada a cabo por los organismos notificados; sin embargo, una gran parte de las aplicaciones de salud que no califican como productos sanitarios se basa en sistemas de autoevaluación. Lo anterior contrasta con las exigencias normativas para aquellas aplicaciones que deben certificarse como interoperables con el historial médico electrónico europeo (HME), tema que se abordará más adelante en este capítulo.

La autoevaluación, por su propia naturaleza, demanda la existencia de un sistema sólido de vigilancia, capaz de identificar posibles fraudes y riesgos asociados a estos dispositivos. Para que dicho sistema sea efectivo, es imprescindible que se apoye en normas claras que regulen la responsabilidad de los organismos encargados de la vigilancia del mercado. Aunque en términos generales se considera que la responsabilidad de estos organismos tiene un carácter residual, hay situaciones en las que dicha responsabilidad se ha activado, especialmente cuando los fabricantes han resultado insolventes.

En este sentido, el caso Schmitt resulta particularmente ilustrativo. En esta sentencia, el Tribunal de Justicia de la Unión Europea (sentencia de 16 de febrero de 2017, Asunto C-219/15, Schmitt contra TÜV Rheinland LGA Products GmbH, TOL5.958.920) abordó la cuestión de la responsabilidad de los organismos notificados en el marco de su función de vigilancia. El tribunal reconoció que estos organismos gozan de un amplio margen de discrecionalidad al ejercer sus funciones, pero subrayó que dicho margen no es ilimitado. Además, el TJUE menciona que una supervisión insuficiente puede poner en peligro la seguridad de los productos sanitarios, generando riesgos para los usuarios y comprometiendo la eficacia del sistema de certificación. Este caso, resuelto antes de la entrada en vigor del Reglamento (UE) 2017/745, evidencia la necesidad de contar con un marco claro y efectivo para delimitar las responsabilidades en el ámbito de la vigilancia del mercado.

Por otro lado, respecto a la responsabilidad de las administraciones públicas, destaca el papel de la Agencia Española de Medicamentos y Productos Sanitarios (AEMPS) como autoridad competente en la certificación y vigilancia de productos sanitarios. Con la entrada en vigor del Reglamento 2017/745, se han introducido mecanismos más estrictos, como la obligación de realizar auditorías quinquenales sin previo aviso, lo que podría generar responsabilidad si un producto defectuoso no detectado causara daños a un paciente. En estos casos, el incumplimiento de los procedimientos de evaluación podría imputarse al organismo notificado.

La responsabilidad de la AEMPS se considera subsidiaria y se activa únicamente cuando el organismo notificado no ha cumplido con sus funciones y el fabricante no ha implementado las medidas correctivas necesarias. No obstante, Grimalt Servera (2023) defiende que la AEMPS puede incurrir en responsabilidad si no actúa de forma oportuna y proporcional ante un riesgo grave e inminente para la salud pública. En contra, García Micó (2024) argumenta que la imputación de responsabilidad patrimonial a la AEMPS resulta difícil de justificar debido a la presencia de múltiples concausas en los daños sufridos por los pacientes. Según este autor, el papel de la AEMPS en la vigilancia de los productos sanitarios es secundario respecto al de los organismos notificados, limitándose a actuar en casos donde la intervención estatal sea imprescindible.

En conclusión, si bien los organismos notificados y las autoridades notificantes, como la AEMPS, cumplen roles distintos en el control y vigilancia de los productos sanitarios, ambos pueden incurrir en responsabilidad bajo circunstancias específicas. La adopción de mecanismos de evaluación más rigurosos, como los establecidos en el Reglamento (UE) 2017/745, tiene como objetivo reducir los riesgos, aunque también eleva significativamente el nivel de exigencia para todas las partes involucradas.

## 7.5. EL REGLAMENTO DEL ESPACIO EUROPEO DE DATOS SANITARIOS

Como adelantemos en el capítulo anterior, el Reglamento (UE) 2025/327 sobre el Espacio Europeo de Datos Sanitarios (REEDS), aprobado pero pendiente de publicación en el momento de redactar esta obra, establece un marco legislativo diseñado para garantizar el acceso a datos sanitarios electrónicos de alta calidad, promoviendo su uso seguro y protegido (Alkorta Idiakez, 2022-a). Entre otros objetivos, el Reglamento está dirigido a establecer un marco común de compartición de datos sanitarios provenientes de los historiales médicos electrónicos (HME) y de los dispositivos y aplicaciones de bienestar que recopilan y procesan datos de salud, que no sean considerados productos sanitarios según el Reglamento (UE) 2017/745.

Este Reglamento es aplicable a los productos sanitarios con IA regulados por el Reglamento (UE) 2017/745, en la medida en que dichos dispositivos hacen uso de datos de validación y entrenamiento del Espacio Europeo de Datos Sanitarios. Dichos dispositivos deben cumplir tanto con sus disposiciones específicas del Reglamento de productos sanitarios a los

que ya se ha hecho referencia, como con las estipulaciones del EEDS en lo relativo a la generación, almacenamiento y compartición de datos sanitarios electrónicos.

Aunque ambos Reglamentos tienen objetivos distintos, se complementan al establecer estándares rigurosos para la seguridad, fiabilidad e interoperabilidad de los dispositivos y los datos de salud que manejan. El Reglamento (UE) 2017/745 prioriza la seguridad y eficacia de los dispositivos médicos, exigiendo estrictos controles en la evaluación de conformidad, trazabilidad y vigilancia postcomercialización. Por su parte, el EEDS se centra en la interoperabilidad, privacidad y seguridad de los datos, garantizando su protección, accesibilidad y reutilización para fines autorizados. La interacción de ambos Reglamentos crea un marco regulador que equilibra innovación tecnológica con protección de datos, fomentando dispositivos seguros y respetuosos con los principios de privacidad exigidos en la Unión Europea.

El Reglamento del Espacio Europeo de Datos Sanitarios establece normas específicas aplicables a los sistemas HME interoperables, en tanto que productos sanitarios, superponiéndose al Reglamento de productos sanitarios en los arts. 14 a 30 bis en materia de requisitos para la evaluación de conformidad, certificación y vigilancia tras la introducción en el mercado.

El Historial Médico Electrónico europeo (HME) consiste en un sistema digital diseñado para recopilar, almacenar y gestionar datos sanitarios electrónicos personales. Incluye información sobre el historial médico de los pacientes, diagnósticos, tratamientos, medicamentos, alergias, inmunizaciones, imágenes médicas y resultados de laboratorio. Estos datos están distribuidos entre distintos actores del sistema sanitario, como médicos, hospitales y farmacias. El HME está diseñado para garantizar el acceso seguro, interoperable y eficiente a los datos sanitarios para los pacientes y los profesionales de la salud, promoviendo la continuidad asistencial y mejorando la calidad del tratamiento, especialmente en contextos transfronterizos. A continuación, señalamos las normas y especificaciones técnicas comunes recogidas en el REEDS para asegurar la interoperabilidad entre los sistemas HME de distintos Estados miembros.

El art. 14 introduce las categorías prioritarias de datos sanitarios electrónicos que deben integrarse en los sistemas HME. Estas incluyen datos como historiales médicos, recetas electrónicas, imágenes médicas y resultados de laboratorio. Se exige que los sistemas sean capaces de transmitir e interoperar utilizando formatos armonizados a nivel europeo.

El art. 15 establece las obligaciones para los fabricantes de sistemas HME, quienes deben demostrar el cumplimiento de los requisitos esenciales de interoperabilidad y seguridad. Estos requisitos incluyen la utilización de especificaciones comunes definidas por la Comisión para la transmisión de datos y la gestión de la identificación de usuarios.

En el art. 16, se regula el acceso de los profesionales sanitarios a los datos, con el objetivo de garantizar una continuidad asistencial efectiva. Los prestadores deben contar con las herramientas necesarias para facilitar este acceso dentro de un marco que priorice la minimización de datos, limitando el acceso a la información estrictamente necesaria y relevante para el caso clínico.

Los arts. 23 y 24 regulan la infraestructura transfronteriza MiSalud@UE (MyHealth@EU), una plataforma obligatoria para los Estados miembros, que garantiza la interoperabilidad y el intercambio seguro de datos sanitarios en toda la Unión. Los Estados miembros son responsables de organizar puntos de contacto nacionales que respeten las especificaciones técnicas y estándares de seguridad definidos a nivel europeo.

El art. 27 aborda los límites y excepciones en la regulación, como la exclusión de programas de software genéricos no diseñados específicamente para usos sanitarios. Además, aclara que los módulos que se consideren productos sanitarios deben cumplir también con los Reglamentos (UE) 2017/745 y 2017/746.

El art. 30 destaca que los sistemas HME deben someterse a un esquema obligatorio de autoevaluación de conformidad para garantizar la interoperabilidad y la capacidad de registrar y transmitir datos en el formato europeo armonizado, lo cual comprende la creación de entornos de ensayo para probar el cumplimiento técnico de los sistemas antes de su comercialización. Así mismo, prevé que los Estados miembros puedan adoptar normas adicionales para la certificación, contratación o reembolso de sistemas HME en sus territorios, siempre que estas no interfieran con la libre circulación de estos sistemas en el mercado interior.

En cuanto a la **regulación sobre sesgos**, el Reglamento, en su art. 19.2.h, establece que los sistemas HME deben garantizar una gestión de datos equitativa, evitando discriminaciones en el tratamiento de información basada en factores sensibles, como origen étnico, género o condiciones médicas. Además, los sistemas de IA integrados deben cumplir con el Reglamento de IA, que prevé mecanismos específicos para mitigar posibles sesgos algorítmicos.

La preocupación por la mitigación de los sesgos en dispositivos sanitarios con IA se refleja en el número de alusiones que recibe la cuestión en el Preámbulo del Reglamento. En los Considerandos se abordan cuestiones relativas a la inteligencia artificial y los sesgos, enfatizando la importancia de adoptar medidas para mitigar posibles efectos discriminatorios y garantizar la equidad en el uso de estas tecnologías. El Considerando 29 nos recuerda que los sistemas sanitarios que incorporan inteligencia artificial clasificados como productos sanitarios de alto riesgo deben cumplir con los requisitos establecidos en el Reglamento de IA. Este cumplimiento incluye la evaluación de conformidad para asegurar la transparencia, la trazabilidad y la reducción de riesgos asociados, como los sesgos algorítmicos que puedan influir negativamente en la equidad de los resultados clínicos. Sobre el cumplimiento de los referidos requisitos remitimos al capítulo 6º de esta obra.

El Considerando 42 subraya la necesidad de que los sistemas HME diseñados con componentes de inteligencia artificial respeten especificaciones comunes que incluyan garantías relacionadas con la seguridad, la confidencialidad, la protección de los datos personales y la integridad de los procesos. Estas especificaciones se diseñan, entre otros objetivos, para prevenir la discriminación derivada de los datos utilizados, garantizando que los algoritmos no se vean afectados por variables como género, origen étnico, edad o estado de salud.

Por otro lado, el Considerando 24 resalta la relevancia de los principios éticos europeos para la salud digital, que incluyen directrices sobre el tratamiento igualitario y la protección frente a sesgos. Estos principios, adoptados por la red de sanidad electrónica en 2022 (Red de Sanidad Electrónica, 2022), a los que hemos hecho referencia en el capítulo 2º de esta obra, son aplicables tanto a los desarrolladores como a los operadores de sistemas de inteligencia artificial en el ámbito sanitario. Asimismo, el Considerando 27 menciona que, para garantizar una gestión equitativa de los datos, es necesario evaluar los riesgos inherentes a los modelos algorítmicos y demostrar que se han implementado medidas adecuadas para mitigarlos antes de introducir los sistemas en contextos clínicos.

Además, el Considerando citado aborda la importancia de contar con datos de calidad y estandarizados para evitar sesgos derivados de fuentes fragmentadas o incompletas. La interoperabilidad entre los sistemas HME, a través de especificaciones técnicas comunes, se considera un mecanismo esencial para garantizar que los datos sean representativos y fiables, reduciendo el impacto de posibles desigualdades en los resultados de salud.

Este marco normativo persigue generar confianza en la inteligencia artificial en el ámbito sanitario, al tiempo que establece salvaguardas técnicas y éticas destinadas a prevenir discriminaciones y asegurar un tratamiento justo para todos los pacientes.

El texto general refuerza la obligación de los Estados miembros de establecer autoridades de sanidad digital (art. 19) para supervisar el cumplimiento de estos requisitos, promoviendo la interoperabilidad y la equidad en la gestión de datos sanitarios en el marco de la Unión. Estas autoridades cooperarán con organismos europeos como el Comité Europeo de Protección de Datos y las agencias de regulación de productos sanitarios, a las que nos hemos referido más arriba, garantizando una aplicación coordinada de las normas.

Por otra parte, el EEDS regula específicamente el etiquetado **de las aplicaciones de bienestar que recopilan y procesan datos de salud**, que no sean considerados productos sanitarios según el Reglamento (UE) 2017/745. Por ejemplo, wearables como Apple Watch, Fitbit y Withings, que monitorizan signos vitales y realizan análisis cardíacos, deben cumplir con los requisitos de seguridad, interoperabilidad y privacidad; otros dispositivos, como pulseras de actividad o aplicaciones de salud y bienestar, aunque no realicen diagnósticos médicos, deben ajustarse también a las disposiciones del EEDS en cuanto a protección de datos y medidas de seguridad.

El art. 47 del Reglamento establece los requisitos de etiquetado para las aplicaciones sobre bienestar que declaren su interoperabilidad con los sistemas HME. En el apartado 1, se dispone que los fabricantes, al declarar el cumplimiento de los requisitos esenciales establecidos en el Anexo II, deben acompañar sus aplicaciones con una etiqueta que certifique dicha conformidad. Esta etiqueta, que es emitida por el propio fabricante, deberá incluir, según el apartado 2, la lista de categorías de datos sanitarios electrónicos que cumplen con los requisitos, una referencia a las especificaciones comunes utilizadas y el período de validez, que no podrá superar los tres años, conforme al apartado 5.

El formato y el contenido de la etiqueta serán definidos por la Comisión mediante actos de ejecución, tal como se señala en el apartado 3. Además, en el apartado 4, se establece que la etiqueta deberá estar redactada en una o varias lenguas oficiales de la Unión o en aquellas que sean comprensibles en el Estado miembro en el que se comercialice la aplicación. En cuanto a su integración, el apartado 6 prevé que, si la aplicación es parte de un dispositivo o se integra en él, la etiqueta deberá figurar en la aplicación misma o en el dispositivo correspondiente. En el caso de programas informáticos,

la etiqueta podrá ser digital, permitiéndose también el uso de códigos de barras bidimensionales.

Por otra parte, los apartados 7, 8 y 9 introducen obligaciones para diferentes agentes económicos: las autoridades de vigilancia del mercado serán responsables de verificar la conformidad de las aplicaciones con los requisitos esenciales; los proveedores deberán garantizar que cada unidad comercializada incluya gratuitamente la etiqueta correspondiente, y los distribuidores deberán ponerla a disposición de los clientes en formato electrónico en el punto de venta.

El art. 48 aborda la interoperabilidad de las aplicaciones sobre bienestar con los sistemas HME. En el apartado 1, se establece que los fabricantes pueden declarar dicha interoperabilidad si cumplen las condiciones pertinentes, debiendo informar adecuadamente a los usuarios sobre esta capacidad y sus implicaciones. Sin embargo, el apartado 2 aclara que esta interoperabilidad no supone que los datos sanitarios sean automáticamente intercambiados o transmitidos al sistema HME. Para ello, es imprescindible el consentimiento expreso de la persona física, en conformidad con el art. 8 ter del Reglamento. Adicionalmente, el mismo apartado garantiza que el usuario pueda elegir qué categorías de datos desea compartir y en qué circunstancias se llevará a cabo dicho intercambio o transmisión.

El art. 49 regula el registro de sistemas HME y aplicaciones sobre bienestar en una base de datos pública de la Unión Europea. Según el apartado 1, la Comisión creará y mantendrá dicha base de datos, que incluirá información sobre los sistemas HME que hayan emitido una declaración de conformidad y las aplicaciones sobre bienestar que cuenten con una etiqueta conforme al art. 31. De acuerdo con el apartado 2, antes de comercializar o poner en servicio un sistema o una aplicación, los fabricantes o sus representantes autorizados deberán registrar los datos requeridos, incluyendo los resultados de pruebas del entorno conforme al art. 26 bis.

Asimismo, el apartado 3 prevé que los productos sanitarios, los productos para diagnóstico in vitro y los sistemas de inteligencia artificial de alto riesgo, cuando correspondan, también deberán registrarse en esta base de datos, en coherencia con otros Reglamentos de la Unión, como los Reglamentos (UE) 2017/745, (UE) 2017/746 y la Ley de Inteligencia Artificial.

Finalmente, el apartado 4 faculta a la Comisión para adoptar actos delegados que determinen qué datos específicos deberán registrarse en esta base de datos, asegurando la inclusión de la información pertinente para garantizar su adecuada funcionalidad y alcance.

## 7.6. EL REGLAMENTO DE EVALUACIÓN DE TECNOLOGÍAS SANITARIAS

La aprobación del Reglamento Europeo de Evaluación de Tecnologías Sanitarias (Reglamento (UE) 2021/2282, REETS) introduce un procedimiento armonizado, aunque no vinculante, para la evaluación de tecnologías sanitarias, incluyendo productos basados en inteligencia artificial. Su objetivo principal es proporcionar a los Estados miembros un marco que permita tomar decisiones fundamentadas en evidencia científica sobre la adopción de tecnologías innovadoras en sus sistemas sanitarios.

Una de sus características más relevantes consiste en la promoción de la participación activa de pacientes, profesionales sanitarios y otros agentes relevantes durante el proceso de evaluación, asegurando que las tecnologías sean seguras, eficaces y adecuadas a las necesidades reales de los sistemas de salud. Se refuerza su aplicabilidad práctica y fomenta una atención médica más efectiva.

El REETS complementa el marcado CE regulado por el Reglamento (UE) 2017/745, que certifica el cumplimiento de requisitos esenciales de seguridad y rendimiento de los productos sanitarios. Además, establece exigencias adicionales enfocadas en la efectividad clínica real y el impacto de las tecnologías en los sistemas sanitarios. Los fabricantes deben proporcionar evidencia clínica sólida que demuestre los beneficios y riesgos de sus productos en condiciones de uso continuo, superando las obligaciones establecidas en el art. 61 del Reglamento (UE) 2017/745.

El art. 8 del REETS introduce la cooperación entre Estados miembros para realizar evaluaciones clínicas conjuntas, promoviendo una metodología estandarizada que evita duplicidades y asegura criterios uniformes en toda la Unión Europea. Esta colaboración garantiza que las evaluaciones se basen en evidencia científica compartida y un análisis riguroso de la eficacia clínica, económica y ética de las tecnologías.

La transparencia es un componente esencial del REETS. Los fabricantes deben documentar con detalle los algoritmos de inteligencia artificial, los datos utilizados en su entrenamiento y validación, y los procesos de decisión automatizados. En cumplimiento con el Reglamento General de Protección de Datos (RGPD) y el Reglamento del Espacio Europeo de Datos Sanitarios (EEDS), se requiere garantizar la seguridad y privacidad de los datos personales.

El Reglamento también refuerza la vigilancia postcomercialización al exigir la monitorización continua y la reevaluación de los dispositivos in-

cluso después de obtener el marcado CE. Esta exigencia es especialmente relevante, como se ha apuntado en varias ocasiones, para tecnologías basadas en inteligencia artificial, donde el rendimiento puede variar con el tiempo o presentar desviaciones no previstas. Además, se establecen requisitos más estrictos para la calidad y cantidad de datos clínicos aportados, promoviendo una supervisión constante de su efectividad en escenarios reales de uso.

Por último, el REETS facilita la integración de datos generados por dispositivos en los historiales clínicos electrónicos, permitiendo una visión más completa del estado de salud del paciente. Esta integración está supeditada al cumplimiento de los requisitos de interoperabilidad y seguridad definidos en el EEDS2 y la Directiva NIS2, que establecen estándares para la protección de redes y sistemas información en la Unión Europea.

En conjunto, el REETS fortalece el marco regulador mediante la cooperación entre Estados miembros, evaluaciones basadas en evidencia rigurosa y la armonización de criterios. Su énfasis en la transparencia, la evaluación continua y la participación de las partes interesadas asegura que las tecnologías innovadoras sean no solo seguras y eficaces, sino también sostenibles y efectivas en los sistemas de salud europeos.

## 7.7. DESARROLLO REGLAMENTARIO PROYECTADO DE LA EVALUACIÓN DE IMPACTO DE LAS TECNOLOGÍAS SANITARIAS EN ESPAÑA

El Proyecto de Real Decreto sobre la evaluación de tecnologías sanitarias que ha elaborado el Ministerio de Sanidad, implementa el art. 8 del Reglamento (UE) 2021/2282 de manera integral y detallada.

La evaluación de tecnologías sanitarias debe comprobar que los sistemas de IA usen datos representativos de poblaciones diversas, lo que ayuda a evitar sesgos. Un subgrupo específico se encargará de delimitar el ámbito de evaluación, determinando los parámetros pertinentes como la población de pacientes, las intervenciones, los comparadores y los resultados en salud. Este proceso inclusivo debe reflejar las necesidades específicas de los Estados miembros y tomar en cuenta las observaciones de pacientes, expertos clínicos y otros especialistas.

Además, el Proyecto de Real Decreto exige que la documentación técnica sea exhaustiva. Los expedientes relativos a productos sanitarios deben incluir informes de evaluación clínica, documentación del fabricante, dic-

támenes científicos de paneles de expertos y toda la información disponible sobre estudios del producto sanitario. Esta transparencia es esencial para asegurar que las evaluaciones sean completas y representativas.

La participación de diversas partes interesadas es otro aspecto fundamental. El Proyecto fomenta la inclusión de pacientes, proveedores de salud y otros expertos en el proceso de evaluación, asegurando que se consideren una amplia gama de perspectivas y necesidades, lo que contribuye a la equidad y justicia en la aplicación de las tecnologías sanitarias.

Por último, el Proyecto de Real Decreto incorpora consideraciones éticas y sociales en el proceso de evaluación, alineándose con el Considerando 28 del REETS al exigir que las evaluaciones de tecnologías sanitarias sean inclusivas, representativas y transparentes, fomentando la participación de diversas partes interesadas y considerando aspectos éticos y sociales.

*Capítulo 8*

# *El test de equidad de los productos y dispositivos sanitarios inteligentes*

En los capítulos anteriores hemos analizado los requisitos generales que debe reunir el sistema de IA de alto riesgo, prestando especial atención a las IAs sanitarias en las que el riesgo fundamental es el proveniente de los sesgos incorporados al modelo o a los diversos sets de datos.

Las normas aplicables resultan redundantes y a menudo poco claras a la hora de valorar el nivel de equidad de los sistemas. Por ello, el test de equidad que planteamos en este capítulo pretende aportar una guía al juzgador para evaluar el grado de adecuación de los sistemas de IA a las exigencias normativas vigentes a la hora de valorar la diligencia del operador en el desarrollo e implementación de la IA.

Por su especial relevancia en materia de equidad algorítmica, se desarrolla el aspecto relativo a la depuración de los datos.

## 8.1. GARANTÍAS DE EQUIDAD REQUERIDAS A LO LARGO DEL CICLO DE VIDA DEL SISTEMA

El marco normativo europeo regula los productos y dispositivos sanitarios que incorporan inteligencia artificial mediante disposiciones diseñadas para proteger los derechos fundamentales y promover la equidad en el desarrollo y uso de estos sistemas. Estas normativas incluyen el Reglamento de Inteligencia Artificial, el Reglamento General de Protección de Datos, el Reglamento de Productos Sanitarios y el Reglamento del Espacio Europeo de Datos Sanitarios; así como las directivas antidiscriminatorias 2000/43/CE y 2000/78/CE, que garantizan la igualdad de trato en función de origen racial, religión, discapacidad, edad y orientación sexual. Además, en España, la Ley 15/2022 amplía estas protecciones al abordar aspectos como la identidad de género, la condición socioeconómica y la salud. Este marco, que se complementa con directrices como el Real De-

creto 1090/2015 y el Reglamento de Evaluación de Tecnologías Sanitarias, establece un conjunto de requisitos para mitigar los sesgos algorítmicos y prevenir la discriminación. A todas estas normas se ha hecho referencia en los apartados anteriores, analizando específicamente las previsiones encaminadas a prevenir, mitigar y notificar los posibles sesgos incorporados en los productos y dispositivos sanitarios.

A continuación, ofrecemos un resumen de los requisitos legales previstos en el conjunto normativo citado que son aplicables a todo producto o dispositivo sanitario que incorpore IA. Aquellos que califiquen como productos sanitarios de tipo IIa, IIb o III según el Reglamento de Productos Sanitarios, deberán además cumplir con los requisitos específicos de seguridad de esta *lex specialis*. Este resumen aspira a proporcionar a los desarrolladores de inteligencia artificial en el ámbito sanitario una guía práctica basada en la normativa vigente, orientada a garantizar la equidad, transparencia y respeto a los derechos fundamentales en todas las fases del ciclo de vida de sus sistemas.

Los requisitos listados se aplican en todas las etapas del ciclo de vida de los sistemas de IA y son esenciales para garantizar que los dispositivos sanitarios que incorporan estas tecnologías funcionen de manera justa, transparente y ética.

Desde la **fase de concepción**, los sistemas de IA deben alinearse con principios éticos y legales, asegurando que los objetivos sean viables y respeten los derechos fundamentales. El art. 9.9 del Reglamento IA exige que los desarrolladores implementen sistemas de gestión de riesgos que permitan identificar problemas en etapas tempranas. Además, es indispensable garantizar la representatividad y calidad de los datos empleados, según lo dispuesto en el art. 10.2 del Reglamento IA y los arts. 5.1.c y 5.1.d del RGPD, para evitar sesgos que puedan perjudicar a ciertos grupos de la población.

La inclusión de expertos técnicos, responsables éticos y usuarios finales en el diseño inicial permite detectar y abordar riesgos técnicos, éticos y regulatorios. Este proceso no solo mejora la equidad del sistema, sino que también refuerza su fiabilidad. El cumplimiento de estándares de calidad en el procesamiento de datos, en el que destaca la depuración de los sets de datos, también conocido como *data cleansing*, es especialmente relevante en el sector sanitario, donde los datos son limitados y pueden no reflejar adecuadamente la diversidad poblacional. Dada su importancia para la equidad algorítmica, dedicaremos el epígrafe siguiente a esta cuestión.

En las **fases de desarrollo y validación**, los sistemas deben someterse a pruebas iterativas para garantizar su precisión y sensibilidad, tal como exigen los arts. 10.3 y 10.4 del Reglamento IA. Estas pruebas también deben

incluir evaluaciones de impacto en derechos fundamentales y protección de datos, recogidas en el art. 35 del RGPD y el art. 27 del Reglamento IA. Las evaluaciones éticas de ensayos clínicos, reguladas por el Real Decreto 1090/2015, son igualmente necesarias para justificar que los beneficios del sistema superan sus riesgos, especialmente en poblaciones vulnerables.

La garantía de equidad de los sistemas en las fases de desarrollo y validación depende de la diversidad y representatividad de las muestras y set de datos utilizadas durante esta segunda fase de pruebas y evaluaciones, como señala el art. 8 del Reglamento de Evaluación de Tecnologías Sanitarias. La ausencia de sesgos en estas etapas permite evitar que se perpetúen desigualdades en salud, una preocupación recogida en numerosas directrices regulatorias y estudios doctrinales (Binns, 2020; Goodfellow et al., 2016).

Una vez completado el desarrollo, el despliegue y la implementación del sistema en la práctica clínica requieren de mecanismos de supervisión humana, tal como establecen el art. 14 del Reglamento IA y el art. 22 del RGPD. La supervisión no solo permite revisar y, si es necesario, modificar las decisiones automatizadas, sino que también refuerza la transparencia y explicabilidad del sistema. Según el art. 10.5 del Reglamento IA y el art. 13.2.f del RGPD, los desarrolladores deben garantizar que los usuarios comprendan la lógica y limitaciones del sistema.

El **monitoreo continuo** del sistema es indispensable para asegurar su equidad a lo largo del tiempo, incluida la recopilación de retroalimentación de los usuarios y el ajuste constante del modelo para adaptarlo a nuevos datos o situaciones emergentes. La obligación de mantener registros detallados, como exige el art. 10.3 del Reglamento IA, facilita la supervisión de las operaciones del sistema y asegura su conformidad con las normativas vigentes.

Como hemos visto en el capítulo 6º de esta obra, el Reglamento de IA también exige a los desarrolladores y operadores de sistemas de alto riesgo que realicen **evaluaciones de impacto** sobre derechos fundamentales (art. 35), incluyendo la supervisión humana y las medidas correctivas en caso de materialización de riesgos. Estos resultados deben ser comunicados a las autoridades de vigilancia del mercado, como establece el art. 38 del Reglamento IA.

## 8.2. IMPLICACIONES JURÍDICAS DE LA DEPURACIÓN DE LOS DATOS CLÍNICOS

Los especialistas advierten que, aunque el cumplimiento de las exigencias normativas a las que nos hemos referido en el apartado anterior es

una condición necesaria para garantizar la equidad de los algoritmos de salud, no asegura por sí mismo la plena equidad de los sistemas de inteligencia artificial en dispositivos sanitarios. De hecho, las normativas suelen establecer un marco general que no aborda en detalle las particularidades técnicas y contextuales de cada sistema, dejando margen para la persistencia de sesgos no detectados durante las evaluaciones iniciales (Mehrabi, et al. 2021).

Estas deficiencias pueden comprometer la capacidad del sistema para generar resultados precisos y equitativos, especialmente en contextos donde las decisiones automatizadas tienen un impacto directo en la salud de los pacientes. Además, la calidad y representatividad de los datos utilizados dependen en gran medida de procesos específicos de diseño y desarrollo que pueden no estar totalmente controlados por las regulaciones (Stöger, 2020).

El proceso de depuración de datos clínicos constituye una etapa crítica en el desarrollo de sistemas de inteligencia artificial, ya que implica la identificación y corrección de errores, inconsistencias y sesgos en los datos utilizados para entrenar y validar los modelos (Stöger et al., 2020). En el ámbito sanitario, esta tarea adquiere una complejidad particular debido a las características de los datos clínicos, que suelen ser limitados, heterogéneos y, en muchos casos, poco representativos de la población objetivo (Ritter, 2022).

Para mitigar estos riesgos, la literatura especializada resalta el uso de técnicas como la aleatorización, el cegamiento, los registros públicos y los grupos de control. Estas herramientas son esenciales para garantizar que los datos sean fiables, representativos y menos susceptibles de contener sesgos.

La aleatorización asigna a los participantes de un estudio a diferentes grupos de manera aleatoria, asegurando que las diferencias observadas se deban exclusivamente al tratamiento y no a variables externas. Este procedimiento es ampliamente utilizado en ensayos clínicos y constituye una medida eficaz para prevenir sesgos en el diseño experimental (Piantadosi, 2005). Por su parte, el cegamiento evita que los participantes o investigadores conozcan la asignación de los grupos, minimizando sesgos de observación y de confirmación. En estudios como los realizados por Friedman et al. (2015) sobre tratamientos para la diabetes, el uso de diseños doble ciego ha demostrado ser eficaz para garantizar resultados más objetivos y fiables.

Además, el registro de investigaciones clínicas en bases de datos públicas como *ClinicalTrials.gov* o el Registro Español de Estudios Clínicos desempeña un papel importante en la promoción de la transparencia. Estas

plataformas aseguran que los resultados, sean favorables o no, estén disponibles para la comunidad científica, reduciendo el sesgo de publicación y proporcionando una visión más equilibrada de la efectividad de los tratamientos (Sim et al., 2006).

El uso de grupos de control en los estudios permite una comparación directa entre los efectos de un tratamiento y la ausencia de intervención. Por ejemplo, en investigaciones sobre terapias físicas, los grupos de control que no reciben el tratamiento específico ofrecen una medida precisa del impacto de la intervención (Kahan et al., 2014). Este tipo de comparaciones es esencial para evaluar objetivamente la eficacia de nuevas tecnologías sanitarias.

A pesar de su importancia, el proceso de depuración afronta importantes limitaciones. Stöger et al. (2020) señalan que la complejidad inherente a los datos médicos, junto con su heterogeneidad y sensibilidad, dificulta la eliminación total de errores y sesgos. Por consiguiente, resulta imprescindible reforzar el proceso de depuración de los datos con otras medidas como el uso de datos sintéticos al que nos referiremos en el apartado siguiente.

Estas limitaciones subrayan la necesidad de adoptar un enfoque multidisciplinar que combine técnicas avanzadas de depuración con una supervisión constante y un diseño ético de los algoritmos. Además, la capacitación continua del personal y la estandarización de los procesos de recopilación y análisis de datos son elementos esenciales para reducir errores y sesgos. Estudios como el de Ramírez et al. (2017) en Colombia subrayan que la falta de formación adecuada y la ausencia de protocolos claros pueden generar inconsistencias en los datos, afectando la calidad de la investigación clínica. Por ello, invertir en formación y en procedimientos claros no solo mejora la calidad de los datos, sino también la confianza en los sistemas desarrollados.

La existencia de sesgos no detectados pese a los procesos de depuración plantea importantes cuestiones jurídicas relacionadas con la responsabilidad. Ante la posibilidad de que los sistemas de IA ocasionen daños debido a errores derivados de datos deficientes, autores como Stöger et al. (2020) abogan por la adopción de un modelo de responsabilidad objetiva. Este sistema permitiría compensar a las víctimas sin necesidad de demostrar negligencia por parte de los desarrolladores, distribuyendo los riesgos de forma equitativa. Otros autores como Atienza Navarro (2024) ponen en cuestión la virtud de este modelo para regular los daños derivados de la IA. En un contexto en el que la tecnología sigue evolucionando y sus errores pueden tener implicaciones graves para la salud, un modelo de responsabilidad subjetiva

podría proporcionar un marco más adecuado para proteger los derechos de los usuarios y garantizar una gestión justa de los riesgos asociados a los dispositivos con IA. En cualquier caso, la regulación de la responsabilidad por los daños de la IA resulta una pieza fundamental en la política europea de garantizar una IA más equitativa en el sector de la salud.

A continuación, nos ocuparemos del marco regulatorio previsto y en desarrollo para regular los datos sintéticos en la clínica y en la atención sanitaria, en general, como alternativa o complemento al uso de datos no suficientemente representativos o sesgados.

## 8.3. EL USO DE DATOS SINTÉTICOS EN DISPOSITIVOS MÉDICOS BASADOS EN IA

### *a) Regulación del Reglamento de Dispositivos Médicos de la UE (RPS 2017/745) sobre los datos sintéticos*

El Reglamento de Dispositivos Médicos de la UE (MDR 2017/745) no menciona explícitamente los datos sintéticos ni los datos específicos de entrenamiento de la IA. El software basado en IA está regulado por el RPS como cualquier otro software de dispositivos médicos, lo que significa que debe cumplir los Requisitos Generales de Seguridad y Rendimiento (GSPR) del Anexo I. En particular, el Anexo I exige que el software (incluidos los algoritmos de IA) esté «diseñado y fabricado de acuerdo con el estado de la técnica». En la práctica, esto implica que cualquier dato utilizado para desarrollar o validar una IA médica (real o sintética) debe reflejar las mejores prácticas actuales para garantizar la seguridad y el rendimiento.

Según el RPS, los fabricantes deben realizar una evaluación exhaustiva de riesgos y una evaluación clínica de los dispositivos de IA. Aunque no está prohibido utilizar datos sintéticos de entrenamiento, el fabricante tiene la responsabilidad de demostrar que el dispositivo es seguro y funciona eficazmente para el fin previsto. Los reguladores advierten que, por ahora, no se asume que los conjuntos de datos sintéticos superen a los datos clínicos reales en la demostración de la seguridad o la eficacia. Si se utilizan datos sintéticos, se espera que un riguroso análisis de riesgos y beneficios justifique su uso, garantizando que la seguridad del paciente no se vea comprometida. En otras palabras, los datos sintéticos pueden complementar el desarrollo, pero no pueden sustituir la prueba de que el dispositivo funciona con datos reales de pacientes sin una justificación sólida.

Por ejemplo, el artículo 61 del RPS exige pruebas clínicas para los dispositivos, normalmente derivadas de investigaciones clínicas o estudios publicados en los que participan sujetos humanos. Por definición, los conjuntos de datos puramente sintéticos no se derivan del uso real en humanos, por lo que generalmente no cuentan como «datos clínicos» según el RPS (la definición de datos clínicos del RPS hace referencia a la información recopilada a partir del uso de un dispositivo en sujetos humanos o en entornos clínicos). Esto significa que, si un modelo de IA se entrena o prueba con datos sintéticos, los reguladores seguirán esperando ver la validación de datos o pacientes reales para satisfacer los requisitos de evidencia clínica del RPS (LSTS).

### *b) Disposiciones de la Ley de IA de la UE sobre datos sintéticos*

La Ley de Inteligencia Artificial de la UE (Ley de IA), promulgada como Reglamento (UE) 2024/1689, impone requisitos adicionales a los sistemas de IA, incluidos los de los productos sanitarios. Según el esquema basado en el riesgo de la Ley de IA, una IA utilizada en un producto sanitario se clasifica como un sistema de IA de «alto riesgo» (ya que se trata de una aplicación crítica para la seguridad según el RPS). Los sistemas de IA de alto riesgo deben cumplir con obligaciones estrictas, especialmente en materia de gobernanza de datos y documentación.

El artículo 10 de la Ley de IA aborda específicamente la calidad de los datos para el entrenamiento, la validación y las pruebas de la IA de alto riesgo. Exige que estos conjuntos de datos sean de alta calidad, relevantes, representativos y libres de errores o sesgos en la medida de lo posible. Es importante destacar que la Ley no prohíbe ni desaprueba los datos sintéticos, sino que, por el contrario, contempla su uso. De hecho, el artículo 10.5 reconoce los datos sintéticos (o anonimizados) como una alternativa preferible al uso de datos personales sensibles a la hora de comprobar el sesgo algorítmico. La ley establece que, si la detección y corrección de sesgos puede realizarse eficazmente con «datos anónimos o sintéticos», los desarrolladores deben utilizarlos en lugar de procesar datos personales sensibles sobre la salud. Solo si la mitigación del sesgo no puede lograrse con datos sintéticos/anonimizados, los desarrolladores pueden recurrir al uso de datos personales de categoría especial reales (como registros sanitarios de pacientes reales), e incluso entonces bajo estrictas salvaguardas. Esta disposición fomenta esencialmente el uso de datos sintéticos como medida de preservación de la privacidad durante el desarrollo de la IA, en consonancia con el principio de minimización de datos. Aparte de este

contexto de pruebas de sesgo, la Ley de IA no regula explícitamente los datos sintéticos de forma diferente a los datos reales; más bien, mantiene todos los datos de entrenamiento con los mismos estándares de calidad. En otras palabras, los datos sintéticos están permitidos por la Ley de IA, pero deben cumplir los criterios de calidad pertinentes (por ejemplo, ser representativos de la población de pacientes prevista, tener propiedades estadísticas adecuadas, etc.) al igual que cualquier otro conjunto de datos. Los proveedores deberán documentar la procedencia y las características de su conjunto de datos en la documentación técnica del sistema de IA. Si se utilizan datos sintéticos para entrenar o validar el modelo, ese hecho y su justificación deben formar parte de la documentación y del archivo de gestión de riesgos.

Por último, la Ley de IA introduce obligaciones de transparencia y supervisión que se relacionan indirectamente con el uso de datos sintéticos. Por ejemplo, los proveedores de IA de alto riesgo deben implementar un sistema de gestión de calidad y un sistema de seguimiento posterior a la comercialización (Reglamento (UE) 2024/1689). Todos los modelos de IA deben ser monitoreados en el uso en el mundo real para su rendimiento y riesgos, y cualquier deficiencia en los datos de entrenamiento (por ejemplo, si un conjunto de datos sintéticos no logró capturar algún escenario del mundo real) debe ser detectada a través de este monitoreo continuo. La Ley no impone ninguna obligación adicional posterior a la comercialización solo porque se utilizaron datos sintéticos, pero en general requiere vigilancia para toda la IA de alto riesgo.

Más adelante analizamos la vigilancia posterior a la comercialización, pero el punto clave aquí es que la Ley de IA reconoce explícitamente los datos sintéticos como una herramienta útil (especialmente para proteger la privacidad y reducir los sesgos) dentro de un marco sólido de gobernanza de datos. Los fabricantes que utilicen datos sintéticos deberán asegurarse de que siguen cumpliendo con todas las obligaciones de la Ley de IA en materia de calidad de datos, gestión de riesgos, transparencia y supervisión.

### *c) Orientación y mejores prácticas para el uso de datos sintéticos*

Los reguladores y expertos han publicado directrices y principios de mejores prácticas en materia de uso de datos sintéticos. Parece que se está llegando a un consenso de que los datos sintéticos pueden ser muy útiles en el desarrollo de la IA médica, pero su uso debe supervisarse teniendo en cuenta los factores que se listan a continuación.

En primer lugar, debe darse preferencia por los datos reales cuando sea posible. El consenso científico actual (equivalente al «estado de la técnica») es que los datos clínicos reales siguen siendo el estándar para entrenar y validar modelos de IA. Los datos sintéticos no deben sustituir por completo a los datos del mundo real, sino aumentarlos o complementarlos. Por ejemplo, un modelo de IA para diagnósticos radiológicos podría entrenarse con un gran conjunto de imágenes médicas reales, y se pueden añadir imágenes sintéticas para aumentar la diversidad (como simular tumores raros o anatomías atípicas).

Por lo general, se permite el uso de datos sintéticos en varias etapas del desarrollo de la IA médica y, de hecho, se recomienda en algunos escenarios. Los casos de uso comunes incluyen una serie de requisitos como son los siguientes:

Por otra parte, se debe ampliar los conjuntos de datos pequeños o sesgados. Es decir, si una enfermedad es poco frecuente o ciertos grupos de pacientes están infrarrepresentados, los datos sintéticos pueden crear ejemplos adicionales para equilibrar el conjunto de entrenamiento. Esto puede mejorar la generalización y la equidad de un modelo de IA. Por ejemplo, en una IA para una enfermedad rara, algunos registros de pacientes reales podrían ampliarse con perfiles de pacientes generados algorítmicamente para crear un corpus de entrenamiento suficientemente grande y diverso.

Conviene, además, simular casos y escenarios extremos. Los desarrolladores suelen generar datos sintéticos para probar el comportamiento de la IA en situaciones extremas o inusuales que pueden no estar bien capturadas en datos reales (AEPD). Por ejemplo, se podrían sintetizar datos de signos vitales que simulen un dispositivo médico en condiciones fisiológicas extremas o complicaciones poco frecuentes, para garantizar que la IA pueda manejarlos de forma segura. Esto es útil para la verificación y validación, ya que permite realizar «pruebas de estrés» del algoritmo sin esperar a que ocurran de forma natural estos eventos poco frecuentes.

El intercambio de datos debe hacerse preservando la privacidad. Si los conjuntos de datos sintéticos permiten compartir y colaborar sin exponer información real de los pacientes, esto se considera una técnica de mejora de la privacidad en virtud del RGPD, ya que los datos sintéticos generados correctamente no contienen información personal identificable (AEPD). Un hospital podría, por ejemplo, generar una versión sintética de su conjunto de datos de pacientes y compartirla con un fabricante de dispositivos para ayudarle a desarrollar un modelo de IA; dado que los datos compartidos son artificiales, pero estadísticamente similares a los de pacientes rea-

les, pueden avanzar en el entrenamiento del modelo respetando la privacidad del paciente. Tanto los organismos europeos como los nacionales de protección de datos (como la AEPD de España) respaldan la generación de datos sintéticos como medio para aplicar la «protección de datos desde el diseño», siempre que los datos sintéticos estén realmente anonimizados (ningún individuo debe ser reidentificable a partir de ellos) (AEPD 2023).

Como se ha señalado, la Ley de IA prevé explícitamente el uso de datos sintéticos para reducir los sesgos. Una de las mejores prácticas es generar datos sintéticos para grupos infrarrepresentados a fin de equilibrar la exposición de un modelo. Por ejemplo, si una IA para la detección de lesiones cutáneas tiene principalmente ejemplos de piel clara, los desarrolladores podrían generar imágenes adicionales que reflejen tonos de piel más oscuros para mitigar el sesgo en la precisión de la detección. Este uso se ajusta tanto a los principios éticos de la IA como a los próximos requisitos legales de no discriminación.

El rendimiento clínico inicial debe confirmarse a continuación en exploraciones de pacientes reales. Los reguladores enfatizan que confiar completamente en conjuntos de datos sintéticos sin validación en el mundo real es arriesgado: cualquier uso de datos sintéticos debe justificarse con un beneficio claro (por ejemplo, abordar una brecha de datos) y evaluarse para detectar posibles sesgos o errores (LSTS).

Al utilizar datos sintéticos, los desarrolladores deben seguir prácticas de validación sólidas. Los datos sintéticos deben ser realistas y relevantes: deben imitar las propiedades estadísticas de los datos del mundo real para el uso previsto (AEPD). Los datos sintéticos de mala calidad pueden inducir a error a un modelo de IA (principio de «garbage in, garbage out»). Por lo tanto, las directrices recomiendan evaluar los datos sintéticos en cuanto a fidelidad (qué tan bien representan las distribuciones reales de los pacientes) y utilidad (que el modelo de IA entrenado con ellos funcione bien con datos reales). Muchas organizaciones aconsejan tratar la generación de datos sintéticos como parte del ciclo de vida del desarrollo que requiere documentación y verificación (LSTS). Por ejemplo, si se utiliza una GAN (red generativa antagónica) para crear imágenes médicas sintéticas, el equipo debe documentar el entrenamiento de esa GAN, los datos de origen utilizados y proporcionar pruebas (como métricas de similitud estadística o revisión de expertos) de que las imágenes resultantes son de calidad adecuada para el entrenamiento del modelo.

Tanto el RPS como la Ley de IA obligan a los fabricantes a gestionar los riesgos a lo largo del ciclo de vida del dispositivo de IA. La mejor práctica es

incluir cualquier uso de datos sintéticos en el archivo de gestión de riesgos del dispositivo. Los riesgos potenciales a considerar son: la posibilidad de que los datos sintéticos introduzcan sesgos no vistos o enmascaren modos de fallo poco frecuentes, y la posibilidad de que los datos sintéticos no reflejen perfectamente la realidad. Los fabricantes deben documentar cómo se mitigan estos riesgos, por ejemplo, realizando pruebas adicionales con datos reales o seleccionando la generación de datos sintéticos para evitar la replicación de sesgos. Un plan de vigilancia posterior a la comercialización (véase la siguiente sección) es otra medida de mitigación: si una IA se ha entrenado en parte con datos sintéticos, se debe vigilar de cerca su rendimiento en el mundo real para detectar cualquier discrepancia entre las condiciones simuladas y las reales.

En resumen, las directrices actuales respaldan el uso de datos sintéticos como una herramienta poderosa, especialmente para superar la escasez de datos y los obstáculos de privacidad, pero siempre junto con una validación exhaustiva. La clave es que el uso de datos sintéticos debe mejorar el proceso de desarrollo (proporcionando más datos o datos variados) sin convertirse en una muleta que debilite la base probatoria del dispositivo. En última instancia, los reguladores esperan que un dispositivo médico de IA se pruebe en pacientes reales, incluso si los datos sintéticos ayudaron a conseguirlo. La comunidad de desarrolladores está compartiendo activamente las mejores prácticas sobre datos sintéticos (por ejemplo, a través de iniciativas como el proyecto SYNTHIA, financiado por la UE, para desarrollar herramientas de datos sintéticos validados para la asistencia sanitaria), lo que refleja la tendencia hacia el uso responsable de esta tecnología de acuerdo con las normas reglamentarias (SYNTHIA, 2023).

### *d) Requisitos de vigilancia y seguimiento posteriores a la comercialización*

Independientemente de que se utilicen o no datos sintéticos en el desarrollo, una vez que un dispositivo médico de IA está en el mercado, el fabricante debe llevar a cabo una vigilancia posterior a la comercialización de acuerdo con el RPS y el RIA. No existen obligaciones adicionales posteriores a la comercialización solo porque un dispositivo haya utilizado datos sintéticos de entrenamiento, pero el uso de datos sintéticos aumenta la importancia de monitorear el dispositivo con datos del mundo real después de su implementación. Esto se debe a que cualquier limitación de los datos sintéticos (por ejemplo, si ciertas variaciones del mundo real no se captaron completamente durante el desarrollo) debe identificarse y abordarse a través de la retroalimentación del rendimiento en el mundo real.

Según el RPS, los fabricantes deben mantener un sistema de postcomercialización y, para dispositivos de mayor riesgo o tecnologías novedosas, a menudo un plan de seguimiento clínico posterior a la comercialización (PMCF). Esto implica recopilar datos clínicos reales sobre el rendimiento del dispositivo en la práctica y comprobar si hay problemas de seguridad o rendimiento. Si el algoritmo de una IA se entrenó o validó en parte con entradas sintéticas, el PMCF podría diseñarse para prestar especial atención al comportamiento del dispositivo en escenarios simulados. Por ejemplo, si se utilizaron datos sintéticos para representar arritmias cardíacas poco frecuentes en el entrenamiento de una IA de diagnóstico, el PMCF podría recopilar activamente datos o estudios de casos de pacientes reales con esas arritmias para confirmar la precisión de la IA. Los reguladores pueden esperar que dicha supervisión proactiva garantice que las suposiciones hechas durante el desarrollo se mantengan en la realidad. En los casos en que el rendimiento de un modelo de IA se considere deficiente en un subconjunto del mundo real, es posible que el fabricante tenga que volver a entrenar o actualizar el modelo (incorporando potencialmente nuevos datos reales) y posiblemente presentarlo como una actualización ante su organismo notificado, dependiendo de la importancia del cambio.

La Ley de IA de la UE, por su parte, también exige explícitamente la supervisión posterior a la comercialización de los sistemas de IA de alto riesgo. Los proveedores (fabricantes) de IA de alto riesgo deben implementar un plan y un sistema de supervisión posterior a la comercialización para «recopilar, documentar y analizar de forma activa y sistemática los datos pertinentes sobre el rendimiento de los sistemas de IA» en uso (Reglamento (UE) 2024/1689). Lo anterior es muy similar a los requisitos del postcomercialización del RPS a los que se hizo referencia en el capítulo 7º. La Ley de IA exige que se incluya un plan de seguimiento posterior a la comercialización en la documentación técnica, y la Comisión Europea tiene la tarea de desarrollar una plantilla estandarizada para este plan para 2026 (Reglamento (UE) 2024/1689). Para un dispositivo médico basado en IA, tiene sentido integrarlo con el plan postcomercialización del RPS. En la práctica, los fabricantes dispondrán de un sistema de vigilancia integral que cumpla con ambas normativas: seguimiento del rendimiento del dispositivo, de cualquier incidente o error y de los comentarios de los usuarios. Si una IA se desarrolló con datos sintéticos, el plan de seguimiento podría señalar cualquier hipótesis específica que deba verificarse. Por ejemplo, «el modelo no se entrenó con datos reales de pacientes pediátricos, sino con datos sintéticos a escala; por lo tanto, supervisaremos especialmente el rendimiento en el uso pediátrico y recopilaremos datos adicionales para esa

cohorte». Si surge algún incidente o tendencia grave, el fabricante debe informarlo (el RPS tiene requisitos de notificación de vigilancia, y el RIA también introduce la obligación de informar de incidentes o fallos graves de IA). El Reglamento de IA también obliga a tomar medidas correctivas si se descubre que el sistema de IA no cumple o es arriesgado, lo que es paralelo al requisito del RPS de tomar medidas correctivas de seguridad de campo para los dispositivos.

Si existieran lagunas desde el desarrollo (como datos reales limitados en una determinada población), estas deberían abordarse mediante la recopilación de datos específicos en la fase posterior a la comercialización. Es probable que los reguladores europeos examinen de cerca los dispositivos de IA de alto riesgo en los primeros años del RIA, por lo que es esencial contar con un sólido plan de supervisión y actualización. Siempre que el dispositivo funcione según lo previsto en la práctica, el hecho de que los datos sintéticos formaran parte de su desarrollo no debería suponer un problema; en todo caso, demostrar que se ha supervisado y confirmado el rendimiento de la IA en el mundo real validará la idoneidad de utilizar datos sintéticos en su desarrollo.

Por tanto, con la Ley de IA ya adoptada, los reguladores y fabricantes de dispositivos médicos se están preparando para cumplir con dos regímenes. Podemos esperar que surjan nuevas normas armonizadas y documentos de orientación para ayudar a poner en práctica los requisitos de alto nivel de la Ley de IA para la gestión de datos. Por ejemplo, la Comisión Europea desarrollará normas sobre calidad de los datos y, como se ha mencionado, una plantilla para los planes de seguimiento postcomercialización. Es probable que estas aclaren cómo documentar el uso de datos sintéticos en el ciclo de vida de un sistema de IA. También se está debatiendo la actualización de la guía pertinente del RPS (a través del MDCG, el Grupo de Coordinación de Productos Sanitarios) para abordar explícitamente la IA y los datos. Los organismos notificados han comenzado a redactar posiciones comunes sobre la IA (como un reciente cuestionario del Team-NB para la IA en productos sanitarios), y reconocen que se seguirán directrices adicionales para incorporar los detalles de la Ley de IA en las evaluaciones de conformidad del RPS. En resumen, las expectativas normativas en torno a los datos (formación y pruebas) se harán más concretas a medida que se aplique la Ley de IA, lo que establecerá indirectamente puntos de referencia más claros para el uso de datos sintéticos (por ejemplo, definiendo qué significa en la práctica datos «suficientemente representativos», posiblemente a través de normas).

Las actitudes hacia los datos de salud sintéticos tienden a ser más positivas a medida que mejora la tecnología. Los reguladores están viendo ejemplos exitosos de datos sintéticos que ayudan al desarrollo sin comprometer la seguridad. Es probable que veamos un reconocimiento formal de los datos sintéticos en las directrices como una técnica válida para cosas como aumentar los conjuntos de datos para enfermedades raras, siempre que se cumplan ciertas condiciones (como demostrar la fidelidad de los datos sintéticos). La EMA (Agencia Europea de Medicamentos) en el sector de los medicamentos ya ha abordado este tema: en un documento de reflexión de 2021 sobre la IA en la medicina, la EMA señaló que los datos sintéticos pueden ser útiles para ampliar los conjuntos de datos de entrenamiento, aunque advierte que dichos datos deben evaluarse cuidadosamente (...). Es posible que veamos una orientación análoga en el ámbito de los dispositivos médicos, posiblemente a través de las directrices del MDCG o del IRPSF (Foro Internacional de Reguladores de Dispositivos Médicos), que tratan explícitamente de cómo generar y utilizar datos clínicos sintéticos en una presentación reglamentaria. La tendencia es armonizar la forma en que los reguladores de todo el mundo ven los datos sintéticos; por ejemplo, la FDA de EE. UU. se ha mostrado abierta al control sintéticos en los ensayos, y Europa está observando estos avances. En los próximos años, no se sorprenda si los datos sintéticos pasan de ser un concepto exótico a una herramienta estandarizada en el manual de normas (con las correspondientes normas de buenas prácticas sobre su validación).

Por otra parte, la UE ha aprobado el Reglamento del EEDS, Reglamento (UE) 2025/327 del Parlamento Europeo y del Consejo, de 11 de febrero de 2025, cuyo objetivo es facilitar el intercambio de datos sanitarios para la asistencia sanitaria y la investigación. El EHDS menciona mecanismos para el altruismo de datos, la anonimización, etc., y existe interés en cómo los datos sintéticos podrían permitir un intercambio seguro de datos sin las estrictas limitaciones del RGPD. Esto podría normalizar aún más los datos sintéticos en el desarrollo de modelos de IA.

Finalmente, cabe señalar que se están llevando a cabo esfuerzos regulatorios para recopilar pruebas sobre cuándo son más eficaces los datos sintéticos. Las partes interesadas, como el Supervisor Europeo de Protección de Datos (SEPD), también subrayan la importancia de la técnica: los datos sintéticos deben ser realmente anónimos para ser una solución legal al uso de datos personales de salud (AEPD). Podemos esperar mejoras en las técnicas y en las normas legales para garantizar que los conjuntos de datos sintéticos no solo sean técnicamente sólidos, sino también legalmente no identificables.

### *e) Enfoque regulatorio e iniciativas en España*

Actualmente no existen leyes específicas de España ni reglamentos vinculantes que aborden explícitamente los datos sintéticos en los dispositivos médicos de IA. En su lugar, España aplica el RPS de la UE para los dispositivos médicos y el RIA de la UE como principal normativa en materia de IA.

No obstante, se han puesto en marcha iniciativas de apoyo, como un proyecto piloto de entorno de pruebas de IA, un entorno controlado para que las empresas prueben sistemas innovadores de IA bajo supervisión regulatoria. La IA en el ámbito de la salud se ha identificado como un área clave en la que los entornos de pruebas pueden ayudar a perfeccionar las directrices. En estos proyectos piloto, se podría explorar el uso de datos sintéticos como una forma de compartir datos sanitarios con los reguladores o entre los participantes en el entorno de pruebas sin riesgos para la privacidad. Además, la AEPD (2023) ha publicado una guía sobre IA y protección de datos (que incluye directrices de transparencia y notas sobre datos sintéticos como herramienta de privacidad). Estas no son específicas para dispositivos médicos, pero refuerzan los principios generales: por ejemplo, si un hospital o empresa española utiliza datos sintéticos de pacientes, debe asegurarse de que los datos se generaron de acuerdo con las normas de privacidad (sin identificabilidad residual) y de que se informó adecuadamente a los pacientes si se utilizaron datos reales para generar el modelo.

En el ámbito de la investigación y la innovación, España participa en proyectos internacionales para promover el uso de datos sintéticos en el ámbito de la salud. Un buen ejemplo es el proyecto SYNTHIA (Marco de generación de datos sintéticos para la asistencia sanitaria), coordinado por un instituto de investigación de Valencia (SYNTHIA, 2023). Este consorcio financiado por la UE está desarrollando métodos validados para los datos sintéticos sanitarios en diversos ámbitos (imágenes, genómica, datos de HCE, etc.), con el objetivo de abordar la calidad, la confianza y la aceptación normativa de los datos sintéticos. Aunque no es un organismo regulador, este tipo de proyecto informa indirectamente sobre la política futura al proporcionar pruebas y normas para la generación de datos sintéticos. El hecho de que las partes interesadas españolas (académicas, empresas y reguladores) formen parte de estos esfuerzos significa que España está ayudando a configurar las mejores prácticas que podrían reflejarse más adelante en directrices o normas.

En conclusión, podría decirse que la mejor práctica es tratar los datos sintéticos con precaución, es decir, utilizarlos para aumentar o mejorar

los conjuntos de datos reales, pero verificando el rendimiento de la IA con datos reales de pacientes antes de comercializarlos. Todos los procesos habituales del RPS (gestión de riesgos, gestión de calidad, etc.) se aplican por igual, por lo que el uso de datos sintéticos debe documentarse y justificarse en el expediente técnico y en el informe de evaluación clínica. No existe ninguna excepción especial ni ningún obstáculo adicional del RPS únicamente para el uso de datos sintéticos; se rige por las mismas normas probatorias que cualquier otra fuente de datos que respalde la seguridad y eficacia de un dispositivo.

*Capítulo 9*

# *Responsabilidad por daños causados por los productos sanitarios con inteligencia artificial*

La interacción entre los avances tecnológicos y los marcos normativos tradicionales de la responsabilidad civil ha suscitado un debate profundo sobre el modelo adecuado para la reparación de los daños ocasionados por sistemas de inteligencia artificial, particularmente en relación con decisiones automatizadas y sesgos algorítmicos que pueden derivar en prácticas discriminatorias. No en vano, los sistemas inteligentes aplicados a la sanidad, tales como herramientas de diagnóstico basadas en IA o sistemas autónomos de triaje, presentan una doble faz. Por un lado, ofrecen avances significativos en eficiencia y precisión, pero también introducen riesgos de daños físicos, morales y económicos cuando fallan o reproducen sesgos. En este contexto, la regulación de la responsabilidad derivada de la IA en el ámbito médico adquiere una trascendencia singular.

La Unión Europea ha asumido un papel preeminente en este debate, con la aprobación de un paquete normativo que aborda la armonización legislativa en materia de responsabilidad frente a daños causados por sistemas de IA.

En primer lugar, se promulgó el Reglamento de Inteligencia Artificial que, como hemos visto, reconoce en su art. 85 a las personas afectadas su derecho a la reparación sin necesidad de una relación contractual con el proveedor. Esta norma se complementa con la Directiva de 2024 sobre Productos Defectuosos (DPD), que extiende la responsabilidad a los daños causados por software, y con la futura Directiva sobre responsabilidad de la IA actualmente en elaboración. Además, se está preparando la Directiva sobre responsabilidad en materia de IA.

Este marco normativo especializado o segmentado, pretende abordar aspectos que van más allá de los defectos de producto, como la protección de la privacidad o los problemas éticos asociados a los sistemas inteligentes. El marco está diseñado para afrontar las complejidades propias de los sistemas de IA, cuya opacidad y autonomía complican tanto la atribución de responsabilidad como la demostración de causalidad en casos de daño. Mientras

la Directiva de 2024 se basa en un régimen de responsabilidad objetiva, permitiendo reclamar daños sin necesidad de probar la culpa del productor, la Propuesta de Directiva sobre responsabilidad en materia de IA adopta un régimen de responsabilidad subjetiva, que exige demostrar incumplimientos del deber de diligencia por parte de usuarios o proveedores.

La coexistencia de los modelos de responsabilidad objetiva y subjetiva plantea importantes interrogantes sobre la armonización del Derecho europeo y su interacción con los regímenes nacionales de responsabilidad. Al mismo tiempo, se quiere alcanzar un equilibrio que permita proteger de manera efectiva a las víctimas y fomentar la innovación tecnológica, garantizando una reparación adecuada de los daños (COM, 23).

La responsabilidad objetiva, tradicionalmente aplicada a productos defectuosos, encuentra en la Directiva de 2024 un modelo claro y relativamente sencillo de implementar. Sin embargo, la responsabilidad subjetiva, que se apoya en la normativa de los Estados miembro, demanda una adaptación más compleja para integrar las medidas procesales y sustantivas introducidas por la normativa europea sobre IA en aquellos. Esta confluencia normativa tendrá un impacto transformador en la jurisprudencia española, especialmente en el ámbito sanitario, donde la *lex artis* establece un estándar de diligencia que deberá coexistir con las presunciones y mecanismos probatorios contemplados por las nuevas normativas europeas (Atienza Navarro, 2024).

Este capítulo examina, en primera instancia, las características generales de la Directiva de Productos Defectuosos, explorando las dificultades de armonización en el contexto europeo. A continuación, se analiza el ámbito de aplicación de estas normativas, con especial énfasis en su adaptación a los productos y sistemas de IA autónomos empleados en el sector sanitario. Posteriormente, se estudian las medidas procesales diseñadas para simplificar la prueba, como la exhibición de evidencias y las presunciones de causalidad, evaluando su relación con las normas existentes en el Derecho europeo y español. Finalmente, se aborda el régimen de responsabilidad en el Derecho español para los daños ocasionados por sistemas de IA médica, considerando las interacciones entre las normativas nacionales y europeas.

## 9.1. DIRECTIVA DE RESPONSABILIDAD POR PRODUCTOS DEFECTUOSOS DE LA UE

La nueva Directiva (UE) 2024/2853 de responsabilidad por productos defectuosos declara estar diseñada para equilibrar los intereses del consumidor con la necesidad de proporcionar claridad y seguridad jurídica a los

operadores económicos. Desde el punto de vista del consumidor, uno de los principales beneficios es la mayor protección que se ofrece. La Directiva amplía su ámbito de aplicación para incluir no solo productos físicos, sino también software y servicios digitales. Se asegura que los consumidores estén mejor protegidos frente a productos defectuosos en una economía que cada vez depende más de la tecnología digital. Además, la responsabilidad objetiva continúa siendo un principio fundamental, lo que significa que los productores son responsables de los daños causados por productos defectuosos sin necesidad de demostrar culpa o negligencia. La inclusión de consideraciones sobre sostenibilidad y conformidad con las expectativas del consumidor puede incentivar a los productores a mejorar la calidad y seguridad de sus productos.

No obstante, la implementación práctica y la interpretación de algunos de sus aspectos pueden presentar deficiencias tanto para los consumidores como para los productores. La carga probatoria, aunque aliviada en algunos puntos, sigue siendo un tema de debate. Algunos expertos consideran que los criterios establecidos son todavía demasiado restrictivos, especialmente en litigios relacionados con productos de alta tecnología. Peña López (2024) señala que, a Directiva prioriza excesivamente los intereses de los fabricantes, particularmente en productos basados en IA. Aunque la Directiva introduce medidas para aliviar la carga probatoria de los consumidores, señala el autor, estas no son suficientes para contrarrestar las barreras estructurales que deben afrontar las víctimas en litigios relacionados con la IA. El acceso limitado a pruebas técnicas y la complejidad inherente de estos productos perpetúan la desventaja del consumidor frente a fabricantes que gozan de amplias posibilidades de exoneración. Bajo la apariencia de un equilibrio entre intereses, la Directiva favorece notablemente a los fabricantes, lo que podría desalentar las reclamaciones de los consumidores y debilitar la protección efectiva de sus derechos. Según Peña López (2024), esta inclinación se evidencia en la permisividad de la doctrina de riesgos del desarrollo, que beneficia principalmente a los fabricantes al ofrecerles una protección desproporcionada frente a demandas legítimas de consumidores. En particular, los sistemas de alto riesgo utilizados en el sector sanitario donde la seguridad, la salud y la vida demandan un estándar más estricto, se deberían haber previsto limitaciones adicionales a los riesgos de desarrollo.

### *a) Definición de «producto»*

La consideración del software sanitario como producto en la definición dada a dicho término por la Directiva 85/374/CEE, de 25 de julio

de 1985, era dudosa. La Directiva de 1985, ahora derogada, se centraba en bienes muebles corporales, lo que generaba incertidumbre sobre la inclusión del software, especialmente cuando se distribuía de forma independiente y no como parte de un dispositivo físico. Esta ambigüedad se debía, en parte, a la dificultad de atribuir al software defectos que causaran daños físicos, tal como se concebía en el régimen de responsabilidad de la época.

La creciente integración de la inteligencia artificial en productos sanitarios ha acarreado la necesidad de redefinir el concepto de «producto» en el ámbito jurídico de la Unión Europea. La Directiva (UE) 2024/2853 refleja esta transformación al incluir explícitamente el software, tanto independiente como integrado, dentro de su ámbito de aplicación. En este sentido, la nueva Directiva representa un avance significativo al adaptar el marco jurídico de la responsabilidad por productos defectuosos a la realidad tecnológica actual. Al reconocer al software sanitario como producto, se refuerza la protección de los consumidores y se clarifica la responsabilidad de los fabricantes en un entorno cada vez más digitalizado. En este sentido, la Directiva (UE) 2024/2853 amplía la definición de «producto» para abarcar no solo bienes tangibles, sino también otros componentes digitales que sirven a la funcionalidad y seguridad de los productos, zanjando de esta manera un debate que se prolongó en el tiempo en relación a la consideración que merecían los programas informáticos.

El art. 4.1 define producto como sigue:

> *"[A] efectos de la presente Directiva se entenderá como "producto": cualquier bien mueble, aun cuando esté incorporado a otro bien mueble o aun bien inmueble o interconectado con estos; incluye la electricidad, los archivos de fabricación digital, las materias primas y los programas informáticos"* (art. 4.1)

La definición incluye, por tanto, programas informáticos y archivos de fabricación digital, independientemente de su modo de suministro o uso. La Directiva también comprende en su ámbito de aplicación los servicios digitales conexos, siempre que la falta de conexión impidiera al producto desempeñar alguna de sus funciones esenciales. Estos servicios digitales conexos se consideran componentes del producto cuando están bajo el control del fabricante, ya sea porque son suministrados directamente por él o porque este influye en su suministro por parte de terceros. Con dicha inclusión la Directiva reconoce que la seguridad y funcionalidad de un producto pueden depender tanto de sus componentes físicos como de los digitales y de los servicios asociados.

Es importante destacar que la Directiva excluye del concepto de producto al código fuente, considerándolo como pura información, así como al software libre y de código abierto suministrados fuera del ámbito de una actividad comercial. Esta exclusión encuentra respaldo en la jurisprudencia del Tribunal de Justicia de la Unión Europea (TJUE), que en la sentencia de 10 de junio de 2021 (Asunto C-65/20, VI contra Krone, TOL9.907.951), abordó la cuestión de si un consejo de salud incorrecto incluido en un periódico impreso podría atribuir un carácter defectuoso al periódico. El TJUE falló que la información no constituía un producto defectuoso, ya que no formaba parte de los elementos intrínsecos del soporte físico (el periódico) necesarios para evaluar su defectuosidad. De acuerdo con el TJUE, la información *per se* no puede ser objeto de responsabilidad bajo el régimen de productos defectuosos. Esta línea argumental refuerza la exclusión de «información pura» como producto bajo la Directiva (UE) 2024/2853.

La Directiva es aplicable a los medicamentos, incluidos los medicamentos que incorporan inteligencia artificial para su administración. Si un medicamento produce efectos adversos graves por defectos de fabricación o falta de información adecuada, el fabricante puede ser responsable. Nótese que, en España, los medicamentos están sujetos a un régimen especial de responsabilidad que establece requisitos adicionales y específicos para la responsabilidad relacionada con medicamentos y productos sanitarios. Estos requisitos se concretan en la Ley 22/2006, de 4 de julio, de Garantías y Uso Racional de los Medicamentos y Productos Sanitaros, la cual establece las bases para la regulación de los medicamentos, incluyendo el control de calidad, fármaco-vigilancia y la gestión de riesgos relacionados con su uso. También define las obligaciones de las empresas farmacéuticas respecto a los efectos adversos y su comunicación. Adicionalmente, el Real Decreto 1345/2007, de 11 de octubre, regula el procedimiento de autorización, registro y condiciones de dispensación de medicamentos de uso humano. Incluye disposiciones sobre la fármaco-vigilancia y la gestión de posibles efectos adversos.

### *b) El ciclo de vida del producto por el que se responde*

La Directiva (UE) 2024/2853 introduce un nuevo marco normativo que no solo delimita con precisión las fases del ciclo de vida del producto, sino que también extiende la responsabilidad del fabricante más allá de la introducción del SIA en el mercado. Se pretende con ello establecer un marco claro de responsabilidad en cada una de las fases por las que transcurre

un producto, particularmente en aquellos casos en los que la interacción entre el fabricante y el producto persiste más allá de su introducción inicial en el mercado.

En este nuevo paradigma, se identifican tres momentos clave en el ciclo de vida de los productos: la introducción en el mercado, la puesta en servicio y la comercialización, conceptos que ayudan a delimitar las responsabilidades de los diversos operadores económicos involucrados.

La **introducción en el mercado**, definida en el art. 4.8 como "primera comercialización de un producto en el mercado de la Unión", representa el punto en el cual el producto, ya sea un dispositivo físico o un software, sale del control del fabricante y se pone a disposición en el mercado de la Unión Europea para su distribución o suministro. Esta etapa constituye el inicio formal del ciclo de vida del producto desde la perspectiva de la responsabilidad. Sin embargo, en el caso de los productos basados en IA, la mera transferencia inicial no siempre garantiza que el fabricante pierda completamente el control sobre el producto, debido a la posibilidad de actualizaciones o modificaciones posteriores, como veremos más adelante.

La **comercialización**, contemplada en el art. 4.9, se entiende como el momento en que el producto se transfiere del distribuidor al consumidor o usuario final en el transcurso de una actividad comercial. A diferencia de los momentos previos, que se centran en el control del fabricante, esta etapa se refiera a la entrega física o venta del producto al destinatario final. La comercialización marca el término del ciclo de distribución del producto y, desde una perspectiva jurídica, permite delimitar las responsabilidades entre los distintos actores involucrados.

Por su parte, la **puesta en servicio**, regulada en el art. 4.10, se refiere al momento en que un producto comienza a ser utilizado para su finalidad prevista dentro del territorio de la Unión. Este momento es particularmente relevante en el caso de los productos sanitarios basados en IA, los cuales suelen requerir una instalación, configuración o pruebas específicas antes de poder ser empleados por los usuarios finales. Este proceso de preparación técnica, muchas veces bajo la supervisión o el control del fabricante, introduce un punto adicional en el ciclo de vida que puede generar responsabilidades específicas en caso de defectos relacionados con dicha instalación o configuración.

La Directiva establece, además, que los operadores económicos pueden eximirse de responsabilidad si logran demostrar que no introdujeron el producto en el mercado, no lo pusieron en servicio o no participaron en su comercialización. En el caso de los distribuidores, por ejemplo, no se les

podrá imputar responsabilidad alguna si prueban que no participaron en el proceso de comercialización del producto defectuoso, tal como señala el art. 11.1.b.

Uno de los aspectos más innovadores de la Directiva consiste en la **extensión de la responsabilidad del fabricante** más allá de los momentos tradicionales de introducción en el mercado y puesta en servicio, en reconocimiento de las características específicas de los sistemas de IA. Como señala Jorqui Azofra (2024), esta ampliación de la responsabilidad del fabricante más allá de los momentos tradicionales de control inicial, responde a las particularidades de los sistemas basados en IA, cuyo rendimiento y seguridad pueden evolucionar a lo largo del tiempo, garantizando así una protección más adecuada para los consumidores y un reparto más justo del riesgo entre los diferentes operadores económicos.

En este sentido, el art. 7.2.e, interpretado por el Considerando 50, establece que los fabricantes no solo son responsables por los defectos presentes en el producto en el momento de su introducción, sino también por aquellos que puedan surgir posteriormente debido a factores que se encuentran bajo su control. Dado que las tecnologías digitales permiten a los fabricantes ejercer control más allá del momento de la introducción del producto en el mercado o de la puesta en servicio, los fabricantes deben seguir siendo responsables de las deficiencias que se originen después de ese momento como resultado de programas informáticos o servicios conexos que estén bajo su control, ya sea en forma de actualizaciones o mejoras o de algoritmos de aprendizaje automático. Debe considerarse que estos programas informáticos o servicios conexos están bajo control del fabricante cuando sean suministrados por él o cuando este autorice o de cualquier otro modo consienta en su suministro por un tercero. Por ejemplo, un sistema de diagnóstico médico basado en IA que, tras recibir una actualización periódica del fabricante, presenta un error en su algoritmo que afecta la precisión de los diagnósticos: aunque este defecto no existiera en el momento de su introducción en el mercado, el fabricante seguiría siendo responsable, siempre que la actualización estuviera bajo su control y fuera la causante del defecto.

### *c) Víctimas y daños indemnizables*

En relación con los daños materiales, el nuevo régimen de responsabilidad por productos defectuosos cubre las pérdidas derivadas de la destrucción o el deterioro de cualquier propiedad, siempre que esta no sea el propio producto defectuoso ni un bien dañado por un componente de-

fectuoso del mismo. Como en el régimen anterior, se excluyen las propiedades utilizadas exclusivamente con fines profesionales, así como aquellas que, aunque tengan un **uso mixto**, sean empleadas predominantemente con tales fines.

Este criterio refleja la dificultad creciente de trazar una línea divisoria entre el uso personal y el profesional de muchos dispositivos dotados con IA, cuya versatilidad suele conducir a usos híbridos. Como apunta Jorqui Azofra (2023), el avance de la tecnología digital y los cambios en el mercado laboral han desdibujado los límites entre ambas esferas, haciendo necesaria una interpretación más flexible. Desde esta perspectiva, la Directiva admite que los daños infligidos a bienes de uso mixto queden cubiertos, salvo que el análisis global de su utilización revele un predominio del uso empresarial.

Un avance notable en el ámbito de los daños indemnizables es la inclusión de la **pérdida o corrupción de datos**, según recoge el art. 6. Esta disposición responde a la creciente importancia de los datos en la era digital, reconociendo su valor económico y personal. Por ejemplo, el contenido eliminado de un disco duro o la pérdida de datos esenciales en un sistema de IA pueden dar lugar a una indemnización, siempre que los datos no se utilicen exclusivamente con fines profesionales. Quedan fuera del alcance de esta Directiva los daños relativos a la privacidad, la discriminación o similares, que ya se abordan en marcos normativos específicos.

Como novedad, el art. 1.a de la Directiva (UE) 2024/2853 cubre la indemnización por **daños psicológicos médicamente reconocidos**, reflejando con ello una voluntad de reparación integral del impacto que los productos defectuosos pueden tener sobre las personas. El art. 6 de la Directiva incluye de forma expresa los daños psicológicos como una categoría indemnizable dentro de las pérdidas sufridas por la persona perjudicada. Estos daños son aquellos que afectan a la integridad emocional o mental del individuo y que pueden derivarse de las lesiones corporales o de los efectos secundarios de un producto defectuoso. Por ejemplo, una persona que experimente ansiedad severa o trastorno de estrés postraumático debido al mal funcionamiento de un dispositivo médico de autodiagnóstico basado en inteligencia artificial tiene derecho a reclamar una compensación bajo el régimen establecido por la Directiva. Este reconocimiento resulta particularmente relevante en el ámbito sanitario, donde los productos defectuosos, especialmente los basados en IA, pueden generar situaciones de angustia o impacto emocional significativo en los pacientes.

A diferencia de los daños psicológicos, no están contemplados en el régimen de indemnización previsto por la Directiva los **daños morales de-**

**rivados de la discriminación** en los que un sistema de inteligencia artificial defectuoso perpetúe sesgos discriminatorios, afectando de manera desproporcionada a grupos vulnerables. Por ejemplo, si un sistema de IA utilizado en procesos de selección médica discrimina a un grupo específico por motivos de género, etnia o discapacidad, los daños relacionados con la dignidad, la igualdad de trato o la percepción social del individuo no podrán ser reclamados bajo esta Directiva. Según los Considerandos 20 a 25, los perjuicios debidos a la discriminación son abordados por marcos normativos especializados, como el Reglamento General de Protección de Datos o la legislación antidiscriminación de la Unión Europea. Como mantiene Navas Navarro (2019), es difícil que el defecto que presente un software que no está incorporado a un bien corpóreo (*stand-alone-software*) genere el tipo de daños cubiertos por la normativa de productos defectuosos. En todo caso, podría generar daños económicamente puros como sería el caso de los algoritmos de alta frecuencia que emiten órdenes y contraórdenes al mercado, pero éstos no están comprendidos en el ámbito de aplicación ni de la Directiva ni tampoco de las normas nacionales de trasposición, habida cuenta de que la primera es una norma de armonización máxima. En todo caso, la delimitación entre daños psicológicos y morales es esencial para entender el ámbito objetivo de aplicación de la Directiva. Mientras que los daños psicológicos implican un impacto directo en la salud mental o emocional de la víctima, los daños morales están relacionados con la dignidad, el honor, la reputación o la igualdad.

La exclusión de los daños morales en este ámbito se justifica en parte por la necesidad de evitar una duplicidad normativa y por la dificultad que presenta cuantificar este tipo de perjuicios dentro del marco objetivo de la responsabilidad por productos defectuosos (Rubí Puig, 2024). Además, se pretende garantizar que estas cuestiones sean reguladas mediante legislaciones especializadas que aborden consideraciones éticas y sociales específicas, teniendo en cuenta los matices particulares de cada caso. En nuestra opinión, la exclusión de los daños morales por discriminación responde al enfoque técnico de la Directiva, que se centra en los efectos directos de los productos defectuosos y delega los aspectos más abstractos o vinculados a derechos fundamentales a normativas específicas, destacando así la importancia de una armonización efectiva con estos marcos legales.

En cuanto a las **víctimas legitimadas**, el art. 5 de la Directiva restringe el acceso a la indemnización exclusivamente a personas físicas, excluyendo expresamente a las personas jurídicas sin ánimo de lucro y a las entidades sin personalidad jurídica. Esta limitación ha sido objeto de crítica, pues representa un retroceso respecto a normativas como el art. 128 del

TRLGDCU, que incluye a estas entidades. La exclusión afecta especialmente a sectores como el sanitario, donde profesionales sanitarios, hospitales y centros de salud, que desempeñan un papel fundamental en el uso de productos basados en IA, no podrán reclamar indemnizaciones bajo este régimen (Evangelio Llorca, 2024). No obstante, el art. 2.3.c) de la Directiva permite que, en casos donde concurran los presupuestos adecuados, los perjudicados puedan recurrir a normativas nacionales en materia de responsabilidad contractual o extracontractual por motivos distintos del defecto del producto. De esta forma, se garantiza que puedan subsistir vías alternativas de reclamación basadas en la culpa, la garantía o la responsabilidad objetiva.

Además, el art. 5.2 introduce una excepción importante al permitir que las reclamaciones puedan ser presentadas por personas distintas de la víctima directa en determinados supuestos. Así, se reconoce legitimación a los sucesores o subrogados en el derecho de la víctima, como los herederos en caso de fallecimiento, así como a **quienes actúen en nombre de una o varias personas perjudicadas, de acuerdo con el Derecho nacional o de la Unión**.

### *d) Las expectativas de seguridad en productos con IA*

Según el art. 7.1 de la Directiva, al que se refieren los Considerandos 30 a 35, un producto es defectuoso si no ofrece la seguridad que el público en general tiene derecho a esperar. Con objeto de proteger la salud y la propiedad de las personas físicas, el carácter defectuoso de un producto debe determinarse no por su falta de aptitud para el uso sino por no cumplir las condiciones de seguridad a que tiene derecho una persona o que se exige en virtud del Derecho de la Unión o nacional. Este estándar se evalúa de manera objetiva, tomando en consideración factores como la finalidad prevista, las características del producto, las propiedades específicas y las expectativas legítimas de los consumidores finales.

Según el art. 7.1.h de la Directiva, las expectativas específicas del grupo de usuarios al que se destina el producto deben ser consideradas al evaluar su seguridad. Estas expectativas son particularmente altas en sistemas de IA utilizados en sectores sensibles, como la salud o la movilidad, ya que los consumidores tienen derecho a esperar que estos productos cumplan con los estándares más altos de seguridad antes de su comercialización, de acuerdo con los requisitos legales y técnicos.

Como señala Peña López (2024), la valoración del carácter defectuoso debe incluir un análisis objetivo de la seguridad que el público en general

tiene derecho a esperar y no referirse a la seguridad que espera una persona. La seguridad que el público en general tiene derecho a esperar debe valorarse teniendo en cuenta, entre otras cosas, la finalidad prevista, el uso razonablemente previsto, la presentación, las características objetivas y las propiedades del producto de que se trate, incluido su ciclo de vida previsto, así como las necesidades específicas del grupo de usuarios al que se destina el producto.

Con el fin de reflejar la creciente prevalencia de productos interconectados, la valoración de la seguridad de un producto también debe tener en cuenta los efectos razonablemente previsibles de otros productos en el producto en cuestión, como por ejemplo en un sistema doméstico inteligente o en un hospital en el que se interconectan los aparatos de diagnóstico y los historiales clínicos.

También debe tenerse en cuenta el efecto en la seguridad de un producto de toda capacidad de aprendizaje o de adquisición de nuevas características tras su introducción en el mercado o su puesta en servicio, a fin de reflejar la expectativa legítima de que el programa informático de un producto y los algoritmos subyacentes estén diseñados de manera que se evite un comportamiento peligroso del producto. Por consiguiente, como hemos visto en un apartado anterior, un fabricante que diseñe un producto con la capacidad de desarrollar un comportamiento inesperado debe seguir siendo responsable de todo comportamiento que cause daños.

Un producto también puede considerarse defectuoso debido a su vulnerabilidad en materia de ciberseguridad, por ejemplo, cuando el producto no cumpla los requisitos de ciberseguridad pertinentes (Considerando 32).

Para reflejar el hecho de que, como se ha dicho en el epígrafe anterior, muchos productos permanecen bajo el control del fabricante tras su introducción en el mercado, el momento en que un producto deja de estar bajo el control del fabricante también debe tenerse en cuenta en la valoración de su seguridad.

### *e) El carácter defectuoso de los productos: en especial, de los productos sanitarios de diagnóstico con IA*

La Directiva (UE) 2024/2853 establece un marco flexible para abordar defectos de fabricación, información y diseño en un entorno donde la seguridad depende tanto de estándares técnicos como de las legítimas expectativas de los consumidores y del estado del arte –por lo que se in-

troducen herramientas que permiten exonerar al fabricante en casos de riesgos asociados al desarrollo—.

Tradicionalmente, los defectos de los productos se dividen en tres categorías principales específicas para evaluar la seguridad de estos productos: fabricación, información y diseño. Sin embargo, estas categorías no abarcan los defectos potenciales de los sistemas de IA con capacidad de autoaprendizaje. Por esta razón, la evaluación de la defectuosidad de un producto con IA debe realizarse mediante una mezcla de métodos tradicionales y criterios novedosos ajustados a las características de esta tecnología, como propondremos a continuación.

### Defecto de fabricación

El defecto de fabricación se presenta cuando un producto individual **se aparta del diseño, proyecto o especificaciones previstas** por el fabricante.

Existen distintos medios de probar este defecto. Según el Considerando 30 de la Directiva que nos ocupa, el carácter defectuoso puede considerarse probado si el producto pertenece a la misma serie de producción que otro cuya defectuosidad ya ha sido demostrada.

Algunos productos, como los productos sanitarios de soporte vital, conllevan un riesgo especialmente elevado de daños para las personas y, por lo tanto, generan unas expectativas de seguridad especialmente elevadas. Para tener en cuenta estas expectativas, el órgano jurisdiccional debe poder considerar defectuoso un producto sin que se declare probado su verdadero carácter defectuoso, cuando pertenezca a la misma serie de producción que un producto cuyo carácter defectuoso ya ha sido probado (Considerando 30).

En el caso de los sistemas de diagnóstico con IA, donde los bienes jurídicos comprometidos —la vida y la salud— exigen umbrales de seguridad particularmente elevados, este criterio de comparabilidad con otras IAs de la misma serie o modelo algorítmico se debe aplicar con especial énfasis (Jorqui Azofra, 2023).

### Defecto de información

Los defectos de información abarcan la ausencia de **instrucciones claras de uso o advertencias adecuadas** sobre los riesgos asociados a un producto. En el caso de los sistemas de inteligencia artificial, la interacción con los usuarios presenta retos particulares debido a la complejidad inherente a

estas tecnologías y la dificultad de identificar de manera inmediata todos los posibles riesgos. Por ello, resulta necesario garantizar que la información proporcionada sea exhaustiva, comprensible y suficiente para prevenir malentendidos o un uso indebido.

El art. 13 del RIA y el art. 7.1.h de la Directiva de responsabilidad por productos defectuosos coinciden en exigir que se garantice la seguridad del usuario mediante información adecuada y suficiente. Según ambos preceptos, un producto sanitario basado en inteligencia artificial será considerado defectuoso si no incluye advertencias específicas y detalladas sobre los riesgos asociados con su uso, tanto en condiciones normales como en escenarios previsibles. Este estándar abarca la obligación de identificar y comunicar limitaciones del sistema, como posibles sesgos en los datos de entrenamiento o incertidumbres en los resultados, así como de prever y advertir sobre los usos razonablemente esperados de la IA. Este último aspecto resulta especialmente problemático en sistemas cuyo comportamiento puede ser impredecible o dar lugar a aplicaciones no previstas inicialmente (López Tur, 2023, p. 57; Jorqui Azofra, 2023).

En el caso de los **productos sanitarios que incorporan inteligencia artificial**, la obligación de información y transparencia es distinta según se dirija al personal sanitario o al usuario final.

La **información destinada al médico** debe ajustarse a los estándares profesionales y normativos que permitan un uso seguro y eficaz del producto. Recuérdese que tanto el Reglamento (UE) 2024/1689 sobre inteligencia artificial, como el Reglamento (UE) 2017/745 sobre productos sanitarios establecen la obligación de que los fabricantes proporcionen información técnica suficiente para que los profesionales sanitarios puedan comprender el funcionamiento de los productos, incluyendo aspectos relativos a la fiabilidad de los diagnósticos generados por la IA en función de las métricas de validación clínica como sensibilidad y especificidad de la IA, las posibles limitaciones de los algoritmos y los posibles sesgos inherentes al propio modelo. Este nivel de detalle técnico es indispensable para que el médico tome decisiones clínicas basadas en los resultados proporcionados por el sistema de IA, considerando los riesgos asociados e integrándolos de manera adecuada en el contexto específico de cada paciente.

El **paciente, como destinatario final del diagnóstico**, tiene derecho a recibir información comprensible y adaptada a su nivel de entendimiento, tal y como se establece en el art. 4 de 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. Esta normativa

exige que la información sea clara, veraz y accesible, permitiendo al paciente participar activamente en la toma de decisiones sobre su salud.

Cuando un diagnóstico basado en IA esté totalmente automatizado, el fabricante debe garantizar que el paciente sea informado de manera transparente sobre el papel de la tecnología en el proceso diagnóstico, según lo dispuesto en los arts. 13.2.f y 14.2.g del RGPD. Se trata de explicar, en términos comprensibles, cómo se ha utilizado la IA, sus beneficios y limitaciones. Además, el paciente tiene derecho a que esta información se transmita de manera que le permita comprender su situación clínica y tomar decisiones autónomas, conforme a los principios establecidos en la Ley de Autonomía del Paciente. Este marco normativo se refuerza con el art. 13 del Reglamento de Inteligencia Artificial, que subraya la necesidad de transparencia en el uso de estas tecnologías, y con el art. 7.2.h del Reglamento (UE) 2024/2853, que destaca la importancia de considerar las características específicas del público destinatario del producto.

En el caso de **sistemas de IA que tengan implicaciones sobre la salud del paciente** sin ser productos sanitarios, como sistemas automatizados de triaje de pacientes o decisiones sobre seguros de salud, se aplicará también, como se ha visto en el capítulo 4º de esta obra, el art. 22.1 del RGPD que prohíbe las decisiones "basadas únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzcan efectos jurídicos en el interesado o lo afecten significativamente de modo similar", a no ser que sean aplicables las excepciones previstas en el apartado 2 del art. 22, a saber, consentimiento explícito del interesado, que se trate de una excepción contemplada por el Derecho de la Unión o del Estado miembro, o que sea necesario para cumplir la obligación contraída en un contrato. Dicha disposición establece una prohibición de principio cuya inobservancia no necesita ser invocada proactivamente por el interesado, según la interpretación del TJUE en el asunto SCHUFA (C-634/21, TOL9.882.990). A esto se añade la explicación que tiene derecho a solicitar la víctima de una decisión automatizada al desplegador ex art. 86 del RIA, según el cual, el interesado tendrá derecho a obtener información significativa sobre la lógica aplicada por el algoritmo, tal como hemos visto en el capítulo 6ª. Además, según el art. 22 del RGPD el afectado tendrá derecho a obtener intervención humana por parte del responsable, a expresar su punto de vista y a impugnar la decisión.

Para el caso de los **medicamentos que incorporen IA,** como es el caso de productos farmacológicos cuya administración depende de un software inteligente, la Ley 29/2006 establece un modelo de información diferen-

ciada y complementaria que equilibra las exigencias científicas dirigidas a los profesionales sanitarios con los derechos de los pacientes a recibir información comprensible y adaptada, garantizando así tanto la seguridad y eficacia de los tratamientos como la protección de los derechos de los usuarios del sistema sanitario.

En consecuencia, los fabricantes de productos sanitarios que incorporan IA (incluidos los medicamentos) y de sistemas que tengan un impacto en la salud del paciente (como los SIA de triaje), deben desarrollar una estrategia informativa que contemple esta dualidad desplegador/paciente, diseñando documentación técnica detallada y ajustada a los estándares profesionales para el médico, y asegurándose de que esta información se traduzca, en la relación médico-paciente, en una comunicación clara y comprensible que respete plenamente los derechos de autonomía e información del paciente. Este sistema de información dual está dirigido no solo a garantizar la eficacia clínica del producto, sino también el cumplimiento de las exigencias legales de transparencia y la protección de los derechos del paciente.

Cabe destacar que las advertencias, «disclaimers» u otra información proporcionada con el producto no son suficientes para convertir un producto defectuoso en seguro. La determinación del carácter defectuoso se basa en la seguridad que el público general tiene derecho a esperar. De acuerdo con el Considerando 2, esta evaluación incluye también el uso indebido, pero razonablemente previsible del producto, como el comportamiento de un usuario que pierde la concentración al operar maquinaria o el de un niño con un juguete. Asimismo, el consentimiento informado no exime al fabricante de su responsabilidad si el producto resulta defectuoso, ya que las expectativas razonables de seguridad no pueden ser derogadas mediante acuerdos contractuales o información proporcionada al usuario.[6] Esta cuestión es especialmente preocupante en el caso de aplicativos

---

[6] El Tribunal Supremo español en el caso «Ala Octa» (Recurso de Casación núm. 382/2018), ha condenado a fabricantes de productos médicos por defectos relacionados con la falta de información adecuada sobre efectos secundarios, estableciendo la necesidad de informar diligentemente sobre todos los riesgos razonablemente previsibles (art. 137.1 TRLGDCU). En esta sentencia, el Tribunal Supremo abordó la responsabilidad patrimonial derivada del uso de un producto sanitario defectuoso en intervenciones oftalmológicas, específicamente el gas perfluoroctano comercializado como «Ala Octa». El fallo determinó que la responsabilidad por los daños causados recaía en el fabricante del producto y en la Agencia Española de Medicamentos y Productos Sanitarios (AEMPS), eximiendo de responsabilidad al Servicio de Salud que utilizó el producto en sus intervenciones.

de autodiagnóstico en los que el consentimiento informado del paciente es sustituido por el contrato de uso del aplicativo en el que no hay información directa al paciente por parte del profesional (Gerke et al. 2022).

### Defecto de diseño

El defecto de diseño se manifiesta cuando la concepción misma de un producto no asegura un nivel adecuado de seguridad. A diferencia de los defectos de fabricación, donde resulta más sencillo establecer comparaciones, en los defectos de diseño el principal escollo radica en determinar un estándar de referencia. Principalmente, se han empleado dos técnicas para esta evaluación: el test de riesgo-utilidad, que requiere que el demandante demuestre la existencia de alternativas más seguras, y el test de las expectativas legítimas del consumidor, que traslada al fabricante la carga de probar la inexistencia de alternativas viables (Reglero, 2008). Sin embargo, los productos que integran inteligencia artificial plantean una complejidad adicional, ya que su desempeño está influido por múltiples factores que dificultan la delimitación del defecto.

La evaluación de defectos de diseño en sistemas de IA se complica por la naturaleza evolutiva de estos sistemas, así como por la dificultad de prever usos inesperados o inapropiados. En productos basados en IA con aplicaciones imprevisibles, como la IA generativa, la dificultad de detectar defectos antes de su comercialización añade una nueva dimensión al problema.

De acuerdo con el art. 7.e de la Directiva (UE) 2024/2853, los fabricantes pueden eximirse de responsabilidad si demuestran que el defecto era indetectable conforme al estado del conocimiento técnico y científico al momento de la introducción del producto en el mercado. Sin embargo, esta exención, conocida como el **riesgo de desarrollo**, genera tensiones con las expectativas de seguridad del consumidor promedio, especialmente en productos con un impacto crítico en la vida y la salud. En este marco, el citado art. 7 introduce un punto de vista novedoso al considerar la capacidad de los productos para aprender o desarrollar nuevas características tras su lanzamiento. Este aspecto, característico de los sistemas de IA, configura una forma de defectuosidad distinta a las categorías tradicionales (fabricación, información o diseño). En lugar de encajar en estas categorías, se diría que estos sistemas plantean una nueva tipología de defectuosidad, específica para productos con aprendizaje continuo. Esta nueva categoría de defecto reconoce que un producto inicialmente conforme a los estándares de seguridad y calidad puede desarrollar características inesperadas o comportamientos imprevisibles debido a su capacidad autónoma de apren-

dizaje (Peña López, 2024). Tal como referiremos en el epígrafe siguiente, esto implica que la responsabilidad del fabricante no se limita al momento de la comercialización, sino que se extiende a garantizar que los productos continúen siendo seguros durante su ciclo de vida. Así, se refuerza la necesidad de una supervisión continua y de actualizaciones periódicas para mantener la seguridad del producto, incluso después de su introducción en el mercado.

Otro aspecto relevante a la hora de valorar la defectuosidad de los productos que incorporan SIA es el **impacto de los datos defectuosos** utilizados para entrenar y validar los algoritmos, lo que incide directamente en la seguridad y eficacia del producto. Los sesgos y errores derivados de datos de entrenamiento deficientes comprometen la seguridad del producto y generan responsabilidades legales para los actores de la cadena de valor.

La calidad, representatividad y preparación de estos datos son elementos determinantes al evaluar su defectuosidad. En productos sanitarios que emplean IA, la representatividad de los datos de entrenamiento es esencial para cumplir con los estándares de calidad esperados. Según la Directiva (UE) 2024/2853, un producto debe ajustarse a estándares de seguridad específicos, que incluyen la calidad de los datos empleados. Los defectos relacionados con datos de entrenamiento suelen surgir cuando estos no representan adecuadamente a la población destinataria. Lo anterior puede derivar en sesgos que afecten las decisiones del algoritmo, perpetuando desigualdades. Por ejemplo, un dispositivo médico que emita diagnósticos erróneos al omitir datos de ciertas etnias o condiciones médicas puede considerarse defectuoso, especialmente si compromete la seguridad de grupos vulnerables. Por lo tanto, cualquier fallo en la preparación de los datos debe considerarse un defecto de diseño que impacta directamente en la seguridad del producto.

Hemos visto en el capítulo 8º que el proceso de limpieza de datos desempeña un papel fundamental en el desarrollo de algoritmos de IA, ya que pretende garantizar conjuntos de datos representativos, fiables y libres de sesgos. Este proceso incluye la eliminación de errores, inconsistencias y elementos irrelevantes que podrían comprometer el desempeño del sistema. Como destaca Stöger (2023), la calidad del proceso de limpieza influye directamente en el funcionamiento de los sistemas de IA, y cualquier fallo en esta etapa puede resultar en productos defectuosos. Sin embargo, aunque el marco normativo actual reconoce la importancia de la limpieza de datos como parte de la gobernanza de datos, carece de estándares vinculantes específicos para evaluar su calidad. Queda un margen significativo

de interpretación que puede complicar la identificación de negligencias. Stöger (2023) subraya la necesidad de adoptar estándares internacionales que aborden esta laguna normativa, proporcionando guías claras para garantizar la representatividad y calidad de los datos empleados en productos sanitarios con IA, fomentando una mayor responsabilidad entre fabricantes y proveedores. En casos de fallos relacionados con la limpieza de datos, el fabricante podría ejercer acciones de repetición contra el proveedor especializado encargado de esta tarea (Evans y Cabral, 2023).

En conclusión, los defectos en productos sanitarios basados en IA relacionados con errores en la limpieza de datos representan un reto significativo para el Derecho de daños y la regulación de tecnologías avanzadas. Aunque el marco normativo actual presenta vacíos, la tendencia hacia estándares internacionales y modelos de responsabilidad compartida ofrece herramientas prometedoras para garantizar la calidad de los datos y la seguridad de los sistemas de IA. Asegurar la representatividad y fiabilidad de los datos empleados es, en última instancia, un requisito indispensable para proteger a los usuarios finales y mitigar los riesgos asociados a estas tecnologías.

### *f) Operadores económicos responsables de productos defectuosos*

El nuevo marco normativo pretende garantizar que siempre exista un sujeto jurídicamente accesible para responder por los daños causados por productos defectuosos, independientemente de las circunstancias particulares de su comercialización o modificación. Con dicho fin, la Directiva de responsabilidad por productos defectuosos, en su art. 8, redefine el alcance de la responsabilidad, ampliando el círculo de los operadores económicos responsables, adaptándose a las complejidades de las cadenas de suministro actuales y a los problemas específicos que plantean las tecnologías de inteligencia artificial. El elenco de sujetos potencialmente responsables ya no solo abarca a fabricantes, sino también a importadores, distribuidores, prestadores de servicios conexos y plataformas en línea, asegurando así una protección más efectiva para los consumidores.

El art. 8.1.a de la Directiva que nos ocupa establece que el fabricante es el principal responsable de los daños causados por un producto defectuoso. Este término abarca tanto al productor original como a aquellos que realicen modificaciones sustanciales en el producto fuera del control del fabricante y que posteriormente lo comercialice o lo ponga en servicio (art.8.2). Según el art. 4.18, se considera **modificación sustancial** cualquier alteración significativa en el rendimiento, la finalidad o la naturaleza del producto que genere nuevos peligros o aumente los riesgos existentes.

Tania López Tur (2023) subraya que esta disposición es particularmente relevante en el ámbito de los productos sanitarios con IA, donde los dispositivos a menudo se adaptan a usos médicos específicos no contemplados en su diseño original. Por ejemplo, un centro sanitario que modifique el algoritmo de un sistema de IA para un diagnóstico diferente al previsto inicialmente podría ser considerado como un nuevo fabricante.

La Directiva amplía la responsabilidad más allá del fabricante. El nuevo marco responsabiliza también a los desarrolladores de software y algoritmos que sean fundamentales para el funcionamiento del sistema, siempre que sus contribuciones introduzcan defectos que afecten a la seguridad del producto.

Cuando el fabricante no esté establecido en la Unión Europea, la responsabilidad puede recaer sobre el importador, el representante autorizado o, a falta de este, sobre el prestador de servicios de tramitación de pedidos a distancia. Esta previsión asegura que los consumidores tengan acceso a un sujeto responsable dentro de la UE, lo que refuerza la protección jurídica del perjudicado.

Los distribuidores pueden ser responsables si no identifican a un operador económico responsable en el plazo de un mes desde la solicitud del consumidor afectado, tal como establece el art. 7.5 de la Directiva. Este plazo sustituye los tres meses previstos en el art. 138.2 del TRLGDCU, acelerando con ello los procedimientos y garantizando un acceso más ágil a la reparación.

Además, los prestadores de servicios conexos, como las empresas encargadas de actualizaciones de software, pueden considerarse responsables si sus intervenciones introducen defectos en el producto. Según María Jorqui Azofra (2023), estos agentes son especialmente relevantes en los productos basados en IA, donde las actualizaciones o el mantenimiento pueden modificar sustancialmente el comportamiento del sistema y, con ello, su seguridad.

En el contexto de la comercialización digital, las plataformas en línea también pueden ser consideradas responsables si inducen al consumidor a creer que el producto defectuoso es suministrado directamente por ellas o bajo su control. Según el art. 8 de la Directiva (UE) 2024/2853, estas plataformas están exentas de responsabilidad únicamente cuando desempeñen un papel estrictamente pasivo y cumplan con los requisitos establecidos en el Reglamento (UE) 2022/2065 sobre Servicios Digitales (RSD). Evangelio Llorca (2024) destaca que esta disposición es necesaria para proteger a los consumidores frente a la opacidad de las transacciones en línea, especial-

mente en el caso de productos sanitarios con IA, donde la confianza en el proveedor es fundamental para la seguridad del consumidor.

Cuando no sea posible identificar a un operador económico responsable o este sea insolvente, la Directiva de responsabilidad por productos defectuosos permite que los Estados miembros implementen sistemas nacionales de indemnización subsidiaria. Esta disposición garantiza que las víctimas de productos defectuosos tengan acceso a una compensación, incluso en circunstancias excepcionales. No obstante, la Directiva advierte que estos sistemas no deberían depender exclusivamente de recursos públicos, sino que podrían estar cofinanciados por los operadores económicos implicados en la cadena de suministro.

La Directiva, al tiempo que amplia en la atribución de responsabilidad, incorpora salvaguardas para proteger los derechos de los operadores económicos. Por ejemplo, el prestador de servicios de tramitación de pedidos solo será responsable si no identifica al fabricante, y las plataformas en línea quedan exentas si cumplen con sus obligaciones de diligencia. Además, en casos de modificaciones sustanciales, el desplegador puede considerarse fabricante exonerando así al fabricante inicial. En estos casos, López Tur puntualiza (2024) que la evaluación de riesgos debe ser precisa para delimitar si estas alteraciones generan nuevos peligros atribuibles al modificador.

### *g) Responsabilidad de múltiples operadores y regulación de la concurrencia de causas*

Los arts. 12, 13 y 14 de la Directiva (UE) 2024/2853 establecen un marco legal para regular las situaciones en las que concurren diversas causas en la producción de daños, así como la acción de repetición entre operadores económicos responsables y la posible participación del perjudicado. Estas disposiciones tienen por objeto proteger al consumidor y organizar las relaciones entre los agentes implicados en el ciclo de vida de los productos, especialmente aquellos basados en tecnologías avanzadas como la inteligencia artificial. Se pretende una distribución justa de riesgos entre los operadores, considerando la complejidad del mercado actual y las particularidades de sistemas dinámicos como la IA. En estos casos, los defectos pueden deberse a datos de entrenamiento deficientes, la falta de actualizaciones o incluso actos malintencionados de terceros. Por ello, es fundamental contar con herramientas que permitan distribuir responsabilidades de forma proporcional y garantizar la protección de las víctimas.

El art. 12 aborda la responsabilidad en casos de **concurrencia entre defectos del producto y actos u omisiones de terceros**. Según este precepto, la responsabilidad del operador económico no se reduce cuando un daño es causado conjuntamente por un defecto del producto y por la acción de un tercero, salvo que este último haya roto el nexo causal. La Directiva sigue responsabilizando al fabricante, dejando abierta la posibilidad de que este ejercite acciones de repetición contra el tercero que contribuyó al daño. Esta disposición es especialmente relevante en productos de IA, donde la interacción de algoritmos en entornos conectados puede dar lugar a fallos provocados por vulnerabilidades de ciberseguridad o intervenciones maliciosas de terceros. Esta posibilidad refuerza el principio de protección del consumidor al priorizar la reparación de la víctima y desplazar las controversias entre operadores al ámbito de la repetición, lo que evita que el consumidor se vea perjudicado por conflictos entre responsables.

El art. 14 regula el derecho de repetición entre los **operadores económicos responsables solidarios**. Este derecho permite que, una vez satisfecha la indemnización al perjudicado, el operador que haya asumido la carga principal del pago reclame a los demás responsables en función de su participación efectiva en la causa del daño. La Directiva introduce salvaguardas importantes para proteger a las micro y pequeñas empresas, limitando las acciones de repetición contra estas a los casos en que se pruebe negligencia grave o intencionalidad en la generación del daño. La previsión del art. 14 es adecuada para fomentar la participación de pequeñas empresas en mercados de alta tecnología, como el de los productos basados en IA, sin exponerlas a riesgos desproporcionados. No obstante, Evangelio Llorca (2024) advierte de que la falta de armonización en las normativas nacionales sobre repetición puede generar desigualdades en mercados transnacionales, ya que las reglas de imputación y distribución de responsabilidades varían significativamente entre Estados miembros.

En relación a la renuncia contractual al derecho de repetición, el art. 12.2 prevé que el fabricante que incorpore un componente de un programa informático en un producto no tendrá derecho de repetición contra el fabricante del programa defectuoso cuando este último sea una microempresa o una pequeña empresa (art. 12.2.a); además, autoriza expresamente la posibilidad de renunciar contractualmente al derecho de repetición entre operadores en el caso de que el fabricante de un programa informático defectuoso sea una microempresa o una pyme (art. 12.2.b). El Considerando 54 explica que dicha previsión está dirigida a fomentar la innovación de las microempresas y las pequeñas empresas que fabrican programas informáticos. Pero también sirve para evitar que los grandes operadores en ca-

denas de suministro complejas descarguen su responsabilidad en proveedores más pequeños a través de cláusulas contractuales, lo que debilitaría el equilibrio en la distribución de los riesgos, en perjuicio de los agentes más vulnerables de la cadena de valor del IA.

El art. 13.2 introduce la previsión de **corresponsabilidad** al establecer que la conducta del perjudicado o de una persona bajo su responsabilidad puede reducir o anular la responsabilidad del operador económico. Habrá que valorar, en el supuesto concreto, si el producto era defectuoso, pero la persona perjudicada contribuyó a la producción del daño, o si este se ocasionó exclusivamente por su actuación negligente.

En el caso de que se acredite una ruptura total del nexo causal entre el defecto y el daño, de modo que este no pueda atribuirse al carácter defectuoso del producto, el operador económico podría quedar exonerado de responsabilidad. Un ejemplo ilustrativo sería el de una persona que, por negligencia, no instale las actualizaciones necesarias para mantener el nivel adecuado de seguridad del sistema de inteligencia artificial, pese a las indicaciones explícitas incluidas en las instrucciones de uso, lo que deriva en la ocurrencia de daños. No obstante, la Directiva 2024/2853 establece que, en ciertos casos, la responsabilidad del operador económico no puede reducirse ni anularse cuando los daños sean atribuibles tanto al defecto del producto como a la actuación u omisión de un tercero. Así, pueden darse escenarios donde los actos u omisiones de una persona ajena al operador económico contribuyan al defecto del producto y, en consecuencia, a los daños sufridos. Por ejemplo, si un tercero explota una vulnerabilidad de ciberseguridad inherente al producto, el operador económico no podrá eludir su responsabilidad bajo el argumento de dicha intervención. En interés de la protección del consumidor, cuando un producto sea defectuoso, como en el caso de una vulnerabilidad que comprometa su nivel de seguridad esperado por el público general, la responsabilidad del operador no debe verse afectada por las acciones u omisiones de terceros. En estas situaciones, se preserva el derecho, conforme al marco del Derecho nacional, de ejercer una acción de repetición contra el tercero responsable.

### *h) Derecho a la exhibición de pruebas*

El art. 9 de la nueva Directiva de responsabilidad por productos defectuosos introduce un derecho específico a la exhibición de pruebas, diseñado para abordar las dificultades probatorias inherentes a los casos relacionados con productos complejos, como los basados en sistemas de inteligencia artificial. Este derecho permite que las personas perjudicadas

soliciten al tribunal nacional que ordene al fabricante o a otras partes implicadas que de acceso a información esencial para determinar la defectuosidad del producto o el vínculo causal entre este y el daño sufrido. Este mecanismo persigue equilibrar las asimetrías informativas que suelen caracterizar estos litigios, especialmente cuando la información técnica y los datos sobre el funcionamiento del producto están bajo el control exclusivo del fabricante. En estos casos, el acceso a la documentación pertinente es esencial para que los perjudicados puedan probar la defectuosidad del producto y el vínculo causal con el daño.

Este derecho a la exhibición adquiere una importancia particular en casos donde la complejidad de los sistemas de IA y la opacidad de su funcionamiento dificultan la prueba de la defectuosidad. Por ejemplo, un producto sanitario basado en IA que emite diagnósticos erróneos podría requerir un análisis profundo de su entrenamiento, diseño y procesos operativos para determinar si el defecto es atribuible a errores en los datos o en los algoritmos. En el caso de productos basados en IA, el acceso a las pruebas incluirá, por tanto, información técnica sobre los algoritmos, datos de entrenamiento, registros operativos y protocolos de funcionamiento. Estas pruebas son esenciales para determinar si el defecto reside en el diseño del sistema, en los datos utilizados para entrenarlo o en su implementación. También para poder identificar a los posibles responsables del defecto que ha causado el daño. El derecho a la exhibición de pruebas incluye tanto documentos existentes como aquellos que el demandado deba generar *ex novo* mediante la compilación o clasificación de información disponible.

Los Considerandos 42 a 45 de la Directiva subrayan que la aplicación de este derecho debe ponderarse cuidadosamente. La Directiva prevé que, cuando se invoque el derecho a la exhibición de pruebas, los tribunales adopten medidas para proteger los secretos comerciales y garantizar la confidencialidad de la información (Considerando 43). Por un lado, se reconoce la necesidad de proteger a los consumidores frente a las dificultades probatorias asociadas al uso de tecnologías complejas como la IA. Por otro lado, se advierte que el derecho a la exhibición de pruebas no debe utilizarse de manera abusiva, evitando la divulgación de información que pueda comprometer secretos comerciales u otros intereses legítimos del fabricante. Se trata, por tanto, de garantizar que el acceso a las pruebas sea justo y proporcional, limitándolo a la información estrictamente necesaria para probar la defectuosidad del producto o el nexo causal.

El derecho a la exhibición de pruebas recogido en el art. 9 guarda ciertas similitudes con la doctrina establecida por el TJUE en el caso Syri (C-

311/18). En este caso, el TJUE, en base al Reglamento General de Protección de Datos, reconoció el derecho de los interesados a acceder a datos personales utilizados en procesos de toma de decisiones automatizadas, incluidos los algoritmos y su lógica subyacente, cuando dichos datos afectaran significativamente a sus derechos. En ambos casos, los derechos de acceso y exhibición permiten a las partes afectadas obtener información esencial para defender sus derechos, ya sea en un litigio sobre productos defectuosos o en una reclamación relacionada con la protección de datos personales. Sin embargo, existen diferencias notables. Mientras que en el asunto Syri el derecho a la información se circunscribe al tratamiento de los datos personales y la transparencia exigida por el RGPD, en cambio, en el art. 9 de la Directiva de responsabilidad por productos defectuosos el ámbito se extiende a toda la información técnica relevante para determinar la defectuosidad del producto, incluyendo datos no personales, como los registros operativos o detalles del diseño del sistema. Por otra parte, con ocasión del caso SCHUFA (C-634/21. TOL9.882.990), el TJUE prioriza los derechos de privacidad y la autodeterminación informativa reconocidos en el marco del RGPD, mientras que el art. 9 de la Directiva pretende equilibrar la necesidad probatoria con la protección de secretos comerciales.

### *i) La carga de la prueba*

El art. 10 de la Directiva (UE) 2024/2853 establece la presunción del carácter defectuoso del producto en determinados supuestos. Estas presunciones se activan cuando el demandado incumple su obligación de exhibir pruebas (art. 10.2.a), o cuando el producto no cumple con los requisitos obligatorios de seguridad establecidos por la normativa aplicable (art. 10.2.b).

La Directiva prevé que, en casos donde la presunción de defectuosidad esté basada en un mal funcionamiento evidente (art. 10.2.c), el demandante solo deba demostrar que el daño no podría haber ocurrido de no ser por un defecto inherente al producto. Por ejemplo, en sistemas de IA que fallan sistemáticamente en circunstancias normales de uso, no sería necesario probar específicamente el defecto en el algoritmo, bastando con evidenciar el mal funcionamiento general.

Además, se presumirá el nexo causal entre el carácter defectuoso del producto y el daño cuando se haya comprobado que el producto es defectuoso y el daño causado sea de un tipo normalmente compatible con el defecto (art. 10.3).

El art. 10.4, por su parte, contempla los casos más complejos, permitiendo a los tribunales nacionales aliviar la carga de la prueba cuando el demandante enfrente dificultades excesivas debido a la complejidad técnica o científica del asunto.No obstante, el demandante debe aportar pruebas suficientes que sugieran que el producto probablemente contribuyó al daño y que es razonable considerar que era defectuoso. Los tribunales deben evaluar estas dificultades caso por caso, basándose en la complejidad del producto y del nexo causal. Se trata de proporcionar mayor flexibilidad a los juzgadores para adaptar las exigencias probatorias a la naturaleza del caso, considerando el grado de opacidad y la falta de acceso a información técnica del demandante. Específicamente, se presume la relación causal entre el defecto del producto y el daño cuando se ha demostrado que el producto es defectuoso y el tipo de daño sufrido es normalmente compatible con ese defecto. Por ejemplo, si un dispositivo médico presenta un defecto de diseño que podría razonablemente conducir a un fallo en el diagnóstico, y el paciente sufre una falta de diagnóstico, se presumiría el nexo causal entre el defecto y el daño.

El fabricante puede demostrar que, aunque exista un defecto, este no fue la causa del daño. Así, aunque la víctima tenga una ventaja probatoria sobre el nexo causal, el fabricante puede refutar la presunción presentando pruebas que desvinculen el defecto del daño.

En nuestra opinión, el art. 10 de la Directiva, junto con las medidas previstas en los arts. 8 y 9, representa un avance significativo en la protección de los consumidores frente a las dificultades probatorias en litigios complejos, especialmente aquellos relacionados con productos que incorporan IA. Sin embargo, pueden presentarse problemas en la implementación práctica de estas medidas, particularmente en lo que respecta a las restricciones de las presunciones y la necesidad de flexibilizar los criterios de aplicación. Para garantizar una protección efectiva, será necesario complementar este marco con un desarrollo normativo y jurisprudencial que permita una aplicación adaptada a la realidad tecnológica y jurídica de cada caso.

Evangelio Llorca (2024) critica que, aunque el alivio de la carga probatoria es un avance positivo, no elimina por completo las barreras para probar la responsabilidad. Según la autora, los criterios establecidos resultan demasiado restrictivos, lo que podría limitar su efectividad en litigios relacionados con productos tecnológicos avanzados. Propone que la regla *res ipsa loquitur*, utilizada habitualmente en el ámbito sanitario, sería más flexible y adecuada para resolver casos complejos, especialmente cuando la

opacidad de sistemas como los de inteligencia artificial dificulta el acceso a pruebas directas. Esta regla permite presumir la existencia de un defecto o de un vínculo causal cuando las circunstancias sugieren que no hay otra explicación razonable para el daño.

Aunque la Directiva introduce medidas para superar estas dificultades probatorias propias de los sistemas de IA, las presunciones que establece no alcanzan la eficacia esperada debido a sus estrictos requisitos, aplicables principalmente a sistemas de alto riesgo. Evangelio Llorca (2024) subraya que el ordenamiento jurídico español ya cuenta con herramientas que alivian la carga probatoria, como los principios de disponibilidad y facilidad probatoria previstos en el art. 217.7 de la Ley de Enjuiciamiento Civil y la mencionada regla *res ipsa loquitur*.

### *j) Exclusión y limitación de la responsabilidad*

El art. 15 de la Directiva de responsabilidad por productos defectuosos regula las causas de exoneración de los operadores económicos, incluyendo fabricantes y otros agentes de la cadena de suministro, cuando los daños no puedan imputarse razonablemente a un defecto previsible del producto en función del estado de los conocimientos científicos y técnicos accesibles en el momento de su introducción en el mercado. Este precepto recoge y amplía la doctrina de las exoneraciones tradicionales recogidas en la Directiva 85/374/CEE, adaptándola a los retos de los productos con inteligencia artificial y otras tecnologías avanzadas.

Una de las causas destacadas de exoneración incluidas en el art. 15 es el **riesgo del desarrollo**, anteriormente mencionado. La exoneración regulada en el art. 7.e de la Directiva permite al fabricante evitar la responsabilidad si demuestra que, en el momento en que el producto fue introducido en el mercado, el estado objetivo de los conocimientos científicos y técnicos más avanzados y accesibles no permitía detectar el carácter defectuoso del producto. Según María Jorqui Azofra (2024), esta causa adquiere especial relevancia en el caso de los sistemas de IA, debido a su capacidad de autoaprendizaje, la rápida evolución del sector y la dificultad inherente de prever todos los riesgos potenciales en el momento del desarrollo.

Un ejemplo paradigmático se encuentra en los sistemas de diagnóstico médico basados en IA. En estos casos, el conjunto de datos empleado para el entrenamiento del sistema puede ser adecuado al momento de su desarrollo, pero quedar obsoleto ante la aparición de nuevas patologías o variaciones genéticas desconocidas en ese entonces. En tales circunstancias, el

fabricante puede exonerarse si demuestra que cumplió con las normas de seguridad y vigilancia postcomercialización establecidas por la normativa europea y nacional, y convence de que no existía forma de prever la deficiencia mientras el producto estaba bajo su control, requiriéndose en todo caso el cumplimiento estricto de los estándares de seguridad y gestión de riesgos por parte de los fabricantes.

La interpretación del estado de los conocimientos científicos y técnicos debe basarse en criterios objetivos, considerando los conocimientos más avanzados y accesibles al momento de la introducción del producto en el mercado. En el ámbito de la IA, esta interpretación incluye la accesibilidad a normas armonizadas, especificaciones comunes y avances técnicos documentados. Sin embargo, la accesibilidad a estos conocimientos puede resultar controvertida, especialmente en sectores donde los avances se concentran en manos de pocos actores, como empresas privadas con capacidad limitada de divulgación.

La Directiva, al aplicar de forma general la exoneración por riesgos del desarrollo, elimina la posibilidad de que los Estados miembros excluyan a categorías específicas de productos, como sucedía bajo la Directiva 85/374/CEE con medicamentos, alimentos o productos sanitarios en algunos países. Para sistemas de IA de alto riesgo, como los utilizados en el ámbito médico, esta exoneración debe interpretarse de manera estricta, atendiendo a la obligación del fabricante de implementar sistemas de gestión de riesgos, documentar procesos y cumplir con los estándares técnicos vigentes.

El art. 9.3 del Reglamento de Inteligencia Artificial refuerza esta lógica al exigir que las medidas de gestión de riesgos consideren el estado actual de la técnica, reflejado en normas armonizadas y especificaciones comunes. Jorqui Azofra (2023) enfatiza que el cumplimiento de estas medidas es esencial para evitar que los fabricantes utilicen la exoneración por riesgos del desarrollo como un escudo injustificado frente a su falta de diligencia en la vigilancia y actualización de sistemas basados en IA.

Aunque la Directiva contempla las causas de exoneración, establece también que el fabricante seguirá siendo responsable si el daño resulta del incumplimiento de obligaciones específicas de seguridad, suministro de actualizaciones de software defectuosas o ausencia de vigilancia postcomercialización. Estas disposiciones pretenden equilibrar la protección de las víctimas con la promoción del desarrollo tecnológico, asegurando que los fabricantes sean responsables solo cuando incumplen obligaciones razonables y previsibles.

Además, el art. 15 recoge otras causas tradicionales de exoneración, como la posibilidad de que un fabricante de componentes se libere de responsabilidad si demuestra que el defecto del producto se debe al diseño del producto final o a las instrucciones proporcionadas por el fabricante del mismo. Este esquema permite deslindar responsabilidades en cadenas de suministro complejas, garantizando un reparto equitativo de riesgos y asegurando que la víctima pueda dirigirse contra el fabricante del producto final sin quedar desprotegida. No obstante, en el contexto de productos basados en IA, Raquel Evangelio Llorca (2024) advierte de que esta disposición puede generar situaciones críticas, dado que la integración de un componente defectuoso en un sistema más amplio puede dificultar la identificación de las causas exactas del daño. Por ejemplo, un fallo en un algoritmo de IA desarrollado por un tercero podría quedar oculto hasta que el sistema sea sometido a pruebas específicas en el entorno del producto final.

### *k) Plazos de prescripción y extinción*

Los arts. 16 y 17 de la Directiva de responsabilidad por productos defectuosos establecen los plazos de prescripción y extinción del derecho para reclamar indemnizaciones por daños derivados de productos defectuosos. Este marco regula tanto el plazo específico para iniciar acciones judiciales como el límite temporal general de la responsabilidad de los operadores económicos.

#### Plazo de prescripción

El art. 16 establece un plazo de prescripción de tres años para la incoación de procedimientos judiciales. Este plazo se computa desde el momento en que la persona perjudicada tiene conocimiento o debería haber tenido razonablemente conocimiento de los siguientes elementos:

1. La existencia del daño.
2. El carácter defectuoso del producto.
3. La identidad del operador económico responsable.

La Directiva adopta así un criterio subjetivo para el inicio del cómputo del plazo, garantizando que este no comience hasta que la víctima tenga conocimiento suficiente para ejercitar su acción.

Esta regulación se diferencia de lo previsto en el art. 1.968 del Código Civil y el art. 142.5 del TRLGDCU, que no contemplan de forma explícita este triple mecanismo de seguridad. Esta omisión puede generar inseguridad jurídica, afectando negativamente a los consumidores.

No obstante, la Directiva deja en manos de los Estados miembros la regulación de las causas y efectos de la suspensión o interrupción de este plazo de prescripción, lo que puede dar lugar a variaciones prácticas significativas entre los diferentes ordenamientos nacionales.

**Plazo de extinción**

El art. 17 introduce un plazo de extinción de la responsabilidad que opera como un límite máximo de diez años desde la fecha en que el producto defectuoso fue introducido en el mercado, puesto en servicio o modificado sustancialmente. Este plazo establece un límite temporal razonable para la responsabilidad del fabricante, teniendo en cuenta que los productos envejecen y las normas de seguridad evolucionan con el tiempo.

La Directiva también prevé que este plazo de extinción se reanude tras una modificación sustancial del producto, entendida como una alteración significativa que pueda afectar su conformidad con los requisitos de seguridad aplicables. Por ejemplo, una actualización importante de software en un sistema de IA médico podría activar un nuevo plazo de diez años desde el momento de dicha modificación.

Jorqui Azofra (2024) observa que esta regulación es coherente con el principio de seguridad jurídica y equidad, ya que protege tanto a las víctimas como a los fabricantes frente a demandas que podrían surgir mucho tiempo después de que el producto haya sido comercializado. Sin embargo, la autora critica que la rúbrica del precepto ("plazos de prescripción") no refleje adecuadamente la dualidad entre el plazo de prescripción de la acción y el plazo de extinción de la responsabilidad, lo que puede generar confusión.

Una novedad significativa introducida por la Directiva es la extensión del plazo máximo de responsabilidad a quince años en casos donde, debido a la latencia de un daño corporal, la víctima no haya podido detectar la existencia del daño dentro del plazo ordinario de diez años (art. 17.3). Se trata de un supuesto frecuente en el ámbito sanitario, donde algunos efectos adversos derivados de productos defectuosos, como implantes médicos o sistemas de IA diagnósticos, pueden manifestarse muchos años después de su uso inicial. Se refuerza así la protección de los consumidores en casos excepcionales, garantizando que las víctimas no queden desprotegidas debido a la naturaleza tardía de ciertos daños. Esta ampliación requiere, no obstante, una interpretación precisa de los supuestos de latencia para evitar que esta ampliación se aplique de manera desproporcionada.

Los arts. 16 y 17 de la Directiva de responsabilidad por productos defectuosos establecen, en nuestra opinión, un marco equilibrado que combina el respeto por los derechos de los perjudicados con la necesidad de garantizar la seguridad jurídica para los operadores económicos. El criterio subjetivo adoptado para el inicio del cómputo del plazo de prescripción, junto con la posibilidad de extender el plazo máximo de responsabilidad en casos de latencia del daño, son previsiones adecuadas para afrontar las complejidades de los productos en la era digital. No obstante, parece necesario perfeccionar la redacción y la sistemática de estos preceptos para evitar confusiones y garantizar su correcta aplicación en los distintos Estados miembros.

## 9.2. PROPUESTA DE DIRECTIVA DE RESPONSABILIDAD DE LA INTELIGENCIA ARTIFICIAL

El 28 de septiembre de 2022 la Comisión Europea presentó la Propuesta de Directiva del Parlamento Europeo y del Consejo relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial (PDRIA). Esta iniciativa se presentó junto con la revisión de la Directiva sobre responsabilidad por los daños causados por productos defectuosos, a la que acabamos de hacer referencia, con el objetivo de modernizar y reforzar las normas existentes en materia de indemnización por daños personales, materiales o pérdida de datos causados por productos inseguros.

La propuesta de Directiva sobre responsabilidad en materia de IA planteaba problemas significativos debido a la diversidad de regímenes nacionales de responsabilidad por culpa existentes en los Estados miembros, lo que dificultó la armonización a nivel europeo. Esta complejidad llevó a que la propuesta quedara en un segundo plano mientras se tramitaba la nueva Directiva de responsabilidad por productos defectuosos. Con la entrada en vigor del Reglamento de Inteligencia Artificial el 1 de agosto de 2024, se ha renovado el interés por lograr una armonización, aunque sea mínima, del régimen general de responsabilidad civil en relación con la IA, que complemente las disposiciones del RIA y garantice una protección adecuada para las víctimas de daños causados por sistemas de inteligencia artificial.

### *a) Características generales*

El carácter de Directiva, en lugar de Reglamento, permite a los Estados mantener o incluso adoptar disposiciones más favorables para los demandantes, siempre que no contradigan el Derecho comunitario. Este planteamiento armonizador flexible resulta particularmente relevante en casos

de discriminación algorítmica, ya que posibilita la aplicación de normas de inversión de la carga de la prueba y presunciones de daños morales en contextos específicos, como los previstos en legislaciones nacionales, como la Ley 15/2022 en España.

La propuesta mantiene la culpa como criterio de imputación, identificando como responsables a quienes, en su papel de proveedores o usuarios de sistemas de IA, actúen de manera negligente al incumplir sus deberes de diligencia.

Lejos de establecer un régimen de responsabilidad civil exhaustivo, la Directiva se limita a medidas procesales específicas destinadas a facilitar la prueba de la culpa y la relación de causalidad. Deja fuera elementos esenciales como la definición de culpa, los daños indemnizables y la distribución de responsabilidad entre varios causantes, aspectos que siguen regulándose conforme a los sistemas nacionales o las reglas sectoriales.

La definición de los sujetos responsables está en plena consonancia con el Reglamento de IA, que delimita claramente las obligaciones de proveedores y usuarios según las etapas del ciclo de vida del sistema. Los proveedores tienen la responsabilidad de cumplir con los requisitos durante la fase previa a la comercialización, mientras que los usuarios pueden ser responsables por incumplimientos relacionados con el uso del sistema una vez que este está en el mercado. La meritada distribución de responsabilidades asegura que se cubran todas las etapas asociadas a la IA, desde su diseño y desarrollo hasta su aplicación práctica.

Para aligerar la carga de prueba de las víctimas de daños causados por sistemas de IA, la Directiva introduce dos medidas importantes. Por un lado, otorga a los jueces la facultad de requerir la exhibición de pruebas que estén en posesión de proveedores o usuarios, especialmente cuando se trata de sistemas de alto riesgo. Por otro, establece presunciones *iuris tantum* de causalidad en circunstancias específicas, como la demostración de que un sistema de IA ha generado un resultado perjudicial.

Hacker (2022) critica que la Propuesta presenta dos grandes deficiencias. En primer lugar, señala que la Comisión Europea no ha abordado adecuadamente la hoja de ruta marcada por la Resolución del Parlamento Europeo de 2020 sobre la responsabilidad en materia de IA, omitiendo un análisis profundo de la integración de la responsabilidad objetiva con límites de responsabilidad y descuidando una regulación más amplia de presunciones de negligencia o la inversión plena de la carga de la prueba. En segundo lugar, el autor objeta que, aunque ciertos aspectos de las opciones normativas están razonablemente desarrollados, el análisis de la Comisión

sobre los costes y beneficios de alternativas como la responsabilidad objetiva carece de coherencia y profundidad. Esta falta de exhaustividad limita el potencial del régimen propuesto para abordar eficazmente los desafíos de la IA, por lo que urge a reconsiderar la opción de un Reglamento en lugar de una Directiva, así como incorporar cuestiones sustantivas, y no solo procesales, en la regulación de la responsabilidad civil por daños causados por sistemas inteligentes.

### *b) Ámbito de aplicación*

La Propuesta de Directiva sobre Responsabilidad en Inteligencia Artificial tiene un ámbito amplio, que abarca cualquier tipo de sistema y todo tipo de daños, pero presta especial atención a los sistemas autónomos, ya que su capacidad para operar y tomar decisiones sin intervención humana directa aumenta la dificultad a la hora de atribuir responsabilidades y establecer la causa de posibles daños.

La Directiva excluye expresamente, tal como señala el Considerando 15, los casos en los que el daño resulte de una evaluación humana seguida de una acción u omisión humana, en situaciones en las que el sistema de IA únicamente proporciona información o asesoramiento. En estas circunstancias, la intervención del sistema de IA no interfiere directamente en la cadena causal, permitiendo que la responsabilidad se asigne a la acción u omisión del agente humano, de forma similar a escenarios donde no interviene un sistema de inteligencia artificial.

### *c) Daños cubiertos*

El marco de responsabilidad civil establecido por la PDRIA es notablemente más amplio que el de la Directiva sobre productos defectuosos. Permite reclamar indemnizaciones por cualquier tipo de daño o perjuicio reconocido por las legislaciones nacionales, incluyendo aquellos que afectan a derechos fundamentales como la vida, la intimidad, la igualdad, así como la salud o la propiedad. De este modo, la PDRIA se presenta como un complemento fundamental para los sistemas nacionales de responsabilidad civil, ofreciendo herramientas procesales que facilitan la carga de la prueba a las víctimas y respondiendo a las características particulares de los daños causados por sistemas de IA autónomos (Rubí Puig, 2024).

Por tanto, la Propuesta amplía su ámbito de aplicación a las demandas relacionadas con daños derivados de la discriminación algorítmica, superando las limitaciones de la Directiva sobre responsabilidad por productos

defectuosos. Mientras esta última se centra en defectos de productos y excluye de su regulación los daños morales *stricto sensu* o los daños económicos puros, que son comunes en contextos de discriminación algorítmica, la PDRIA permite reclamar compensaciones por vulneraciones de derechos fundamentales como la privacidad o la no discriminación. Se colma de esta manera una laguna normativa significativa en relación a los daños morales, y se refuerzan los recursos legales para las víctimas en situaciones donde los sistemas de IA desempeñan un papel decisivo.

### *d) Facilitación de la prueba, presunción de culpabilidad y exoneración de responsabilidad*

La Propuesta de Directiva sobre Responsabilidad en materia de Inteligencia Artificial introduce un conjunto de medidas destinadas a facilitar la carga de la prueba en las demandas de responsabilidad civil, especialmente en el contexto de los sistemas de inteligencia artificial de alto riesgo. Los arts. 3 y 4 recogen una batería de medidas para garantizar que las víctimas puedan reclamar indemnización de manera efectiva en casos en los que la complejidad técnica y operativa de los sistemas de IA podría dificultar la prueba de culpa y causalidad.

El art. 3 establece el derecho del demandante a solicitar al juez la **exhibición de pruebas** que obren en poder del proveedor o usuario del sistema de IA, siempre y cuando previamente se haya intentado, sin éxito, obtener dichas pruebas de manera extrajudicial. Esta previsión va más allá de lo dispuesto en los arts. 13, 14 y 22 del Reglamento General de Protección de Datos, y el art. 86 del RIA que reconocen, como se ha dicho anteriormente, el derecho del interesado a obtener información sobre la lógica subyacente a las decisiones automatizadas que le afecten, tal como ha fallado el TJUE en el caso SCHUFA. La PDRIA introduce mecanismos procesales dirigidos a exigir la exhibición judicial de pruebas, incluyendo documentación técnica detallada que puede ser esencial para probar la falta de diligencia del proveedor o usuario del sistema de IA en casos de responsabilidad civil. Esta medida refuerza el marco procesal existente y facilita el acceso a información que de otro modo podría permanecer inaccesible para la víctima, dificultando su capacidad para construir una reclamación sólida.

El art. 4, por su parte, detalla los extremos que debería probar el demandado para **exonerarse de responsabilidad**, dependiendo de si es el proveedor o el desplegador del sistema de IA de alto riesgo. Los proveedores deben acreditar que el sistema de IA fue diseñado y desarrollado conforme a criterios específicos, tales como el uso de conjuntos de datos

de calidad (art. 10 RIA), cumplimiento de requisitos de transparencia (art. 13), mecanismos de vigilancia efectiva (art. 14), niveles adecuados de precisión, robustez y ciberseguridad (arts. 15 y 16), y la adopción de medidas correctoras en caso de incumplimiento (art. 16.g y 21). Por su parte, los desplegadores deben demostrar que emplearon o supervisaron el sistema de acuerdo con las instrucciones de uso y que expusieron al sistema a datos de entrada pertinentes para la finalidad prevista (art. 29).

En relación con los sistemas de inteligencia artificial, la PDRIA **presume la culpa del desplegador** cuando el demandante demuestra que aquel no supervisó ni suspendió el uso del sistema, pese a tener motivos para considerar que podía generar riesgos o defectos graves conforme al art. 26 del Reglamento de Inteligencia Artificial; o expuso al sistema a datos de entrada inadecuados para su finalidad prevista. En otras palabras, si el uso del producto es negligente o no se siguen las instrucciones proporcionadas, también podría atribuirse responsabilidad al desplegador si el daño es consecuencia de una implementación incorrecta o de datos introducidos que alteren el funcionamiento del sistema.

El Considerando 15 de la PDRIA aclara que las previsiones de esta Directiva se aplican exclusivamente a daños causados directamente por sistemas de IA, excluyendo los derivados de decisiones humanas adoptadas tras el uso de información proporcionada por el sistema. Esta exclusión responde a la lógica de que dichos casos no requieren regulación específica, ya que la causalidad puede atribuirse directamente a la acción u omisión humana, como sucede en escenarios sin IA.

Sin embargo, aun cuando los sistemas de IA reemplacen en ciertos contextos a la decisión humana, esto no exime automáticamente de responsabilidad al profesional o usuario del sistema. El art. 14 del RIA establece que los sistemas de IA deben diseñarse y desarrollarse para permitir una vigilancia efectiva durante su uso, y el art. 86 reconoce al afectado por una decisión automatizada el derecho a una explicación de la lógica subyacente al sistema de IA. Por tanto, los profesionales que interactúan con el sistema siguen siendo responsables de supervisar su funcionamiento y, si detectan riesgos evidentes, interrumpir su utilización. La responsabilidad también puede extenderse a situaciones donde el profesional no verificó la idoneidad de los datos de entrada o no garantizó que el sistema se utilizara dentro de los límites de su finalidad prevista.

En relación art. 4, el Considerando 22 del PDRIA refuerza esta obligación al indicar que el incumplimiento de las restricciones de uso impuestas al sistema puede fundamentar una imputación de culpa. En el ámbito mé-

dico, por ejemplo, si un profesional utiliza un sistema diseñado para analizar imágenes pediátricas en pacientes adultos, generando diagnósticos erróneos, el daño será atribuible al uso indebido. Asimismo, el Considerando 25 subraya que la falta de medidas para limitar el ámbito de aplicación del sistema puede justificar una presunción de causalidad, como en el caso de sistemas de IA diseñados para identificar enfermedades específicas que se emplean fuera de su propósito inicial debido a la ausencia de restricciones técnicas.

En este contexto, la PDRIA no solo configura un marco procesal especial para facilitar la prueba en casos de responsabilidad civil, sino que también establece que la responsabilidad no se extingue automáticamente por la mera sustitución de la acción humana por sistemas de IA. Las obligaciones de supervisión, el uso diligente y el manejo adecuado de los sistemas de alto riesgo siguen siendo esenciales para garantizar la protección efectiva frente a posibles daños.

### *e) Revisión de la Directiva*

La Propuesta de Directiva sobre Responsabilidad en materia de Inteligencia Artificial adopta un criterio temporal escalonado para su aplicación. Además, prevé su revisión, reconociendo de esta manera la necesidad de adaptarse a un entorno tecnológico en evolución.

El art. 2.3 de la Propuesta confirma que las normas nacionales sobre responsabilidad extracontractual por culpa siguen siendo aplicables, permitiendo que los Estados miembros adapten estas reglas a sus sistemas jurídicos específicos. Asimismo, el art. 1.4 concede a los Estados miembros la facultad de adoptar normas más favorables para los demandantes, incluyendo la posibilidad de introducir regímenes de responsabilidad objetiva en sus legislaciones. El hecho de que cada Estado miembro pueda desarrollar medidas que amplíen la protección de las víctimas, siempre que estas normas sean compatibles con el Derecho de la Unión, pone de manifiesto el carácter de «armonización mínima» de la Directiva.

La Comisión Europea justifica la opción por la responsabilidad subjetiva argumentando que, aunque algunas voces, especialmente del ámbito de la protección al consumidor, abogan por un régimen de responsabilidad objetiva acompañado de seguro obligatorio, las empresas consideran que tal medida sería desproporcionada. Este desacuerdo se combina con el hecho de que los sistemas de IA de alto riesgo aún no están ampliamente difundidos, representando solo el 5% del total de sistemas de IA según

la Comisión (COM, 2021). En este contexto, la Propuesta establece un sistema escalonado o por fases, diseñado para equilibrar los intereses de protección y desarrollo tecnológico.

El art. 5 de la Propuesta establece un primer período de implementación de cinco años, durante el cual se exige una armonización de mínimos centrada en la obligación de exhibición de pruebas y la utilización de presunciones *iuris tantum*. Estas herramientas procesales están dirigidas aliviar la carga de prueba para los demandantes, facilitando la imputación de responsabilidad sin imponer cargas excesivas a los operadores de sistemas de IA. El establecimiento de un período inicial tiene como objetivo recopilar información sobre la eficacia de estas medidas y evaluar si el marco actual es suficiente para garantizar una adecuada protección de las víctimas.

Tras estos cinco años, la Comisión llevará a cabo una revisión integral del régimen normativo establecido por la PDRIA. Este proceso de reevaluación analizará, entre otros aspectos, la necesidad de introducir un régimen de responsabilidad objetiva y un sistema de seguro obligatorio para cubrir los daños ocasionados por sistemas de IA. Esta posibilidad se contempla como una fase futura que dependerá del nivel de adopción y difusión de los sistemas de IA de alto riesgo, así como de la evolución de los mercados y las prácticas aseguradoras.

El despliegue escalonado refleja una estrategia prudente y adaptable, diseñada para equilibrar la protección de las víctimas con el impulso a la innovación y la competitividad en el campo de la inteligencia artificial (Atienza Navarro, 2024). Asimismo, la flexibilidad que ofrece la Propuesta, al facultar a los Estados miembros para implementar medidas más estrictas si lo consideran necesario, garantiza que el marco regulatorio pueda ajustarse a las particularidades de cada contexto nacional y a las necesidades de los actores involucrados.

## 9.3. PARTICULARIDADES DEL DERECHO ESPAÑOL DE DAÑOS APLICABLE A LA IA SANITARIA

En el supuesto de que una IA autónoma emitiera un diagnóstico erróneo debido a un sesgo en los datos de entrenamiento, tanto el **daño físico** —por el retraso en el diagnóstico que perjudica la condición médica del paciente–, como el **psicológico** y el **moral**, derivado de la discriminación subyacente que vulnera derechos fundamentales como la igualdad de trato, podrían ser objeto de reclamación ante un tribunal español, aunque se requerirían vías normativas diferenciadas para abordar cada uno de ellos.

El régimen específico de **responsabilidad por producto defectuoso** será aplicable a los fabricantes de sistemas de IA cuando estos puedan calificarse como productos defectuosos, lo cual abre la puerta a reclamaciones contra los productores por daños físicos y morales derivados de defectos del sistema.

En los casos no cubiertos por los regímenes específicos de responsabilidad objetiva, como los daños puramente económicos, la discriminación, o aquellos que afectan a derechos de la personalidad, así como los derivados de la responsabilidad profesional médica en general, será necesario recurrir al **régimen general de responsabilidad civil extracontractual**, regulado en el art. 1.902 del Código Civil. Este régimen podrá ser complementado en el futuro con las normas procesales y sustantivas introducidas por la Propuesta de Directiva sobre Responsabilidad en materia de Inteligencia Artificial en casos en los que el daño sea causado por sistemas de IA de alto riesgo y cuando el régimen general o especial no prevea una norma más favorable —ya que la inversión de la carga de la prueba de la culpa prevista en el Derecho español, será mejor que las reglas de PDRIA que pretenden facilitar al demandante la prueba de la negligencia—.

La coexistencia de estas dos vías normativas no solo es viable, sino necesaria, ya que permite abordar de manera integral las diferentes dimensiones del daño causado por la IA autónoma. Mientras que la Directiva 2024/2853 garantiza la reparación del daño físico en términos de responsabilidad objetiva por producto defectuoso; las normas nacionales y, en el futuro, las disposiciones de la PDRIA ofrecerán un marco adecuado para abordar las vulneraciones de derechos fundamentales y los perjuicios de carácter moral. Esta combinación asegura que las víctimas no se vean limitadas a una única vía de reclamación y que los diversos aspectos del daño puedan ser reconocidos y reparados en toda su complejidad.

En lo que respecta al uso de sistemas inteligentes de alto riesgo aplicados al diagnóstico médico, resulta fundamental distinguir entre aquellos que asisten al médico y los que sustituyen su decisión. En el primer caso, la responsabilidad del facultativo se determinará en función de su adherencia a la *lex artis ad hoc*. No será considerado negligente si, tras un análisis diligente, confía en un diagnóstico o tratamiento prescrito por el sistema, aunque este resulte erróneo. Sin embargo, si actúa sin el debido control o revisión de las decisiones del sistema, podría incurrir en responsabilidad por negligencia (Luquín Bergareche). Asimismo, en caso de que el daño sea causado por un defecto inherente al sistema, la responsabilidad recaerá en el productor del sistema conforme a las normas de responsabilidad

por productos defectuosos, o en el centro sanitario. En los casos de SIA autónoma de diagnóstico, hemos visto más arriba que deberá exigirse el consentimiento informado del paciente, cuya omisión también podrá dar lugar a responsabilidad.

Por otro lado, si un médico decide apartarse del diagnóstico o tratamiento sugerido por un sistema inteligente y ello resulta en daños al paciente, podría surgir la cuestión de una posible negligencia por pérdida de oportunidad. Este supuesto introduce un estándar de conducta innovador que compara la decisión del profesional con la que habría adoptado el sistema de IA, planteando una reinterpretación del concepto de negligencia médica en el Derecho español. Sin embargo, esta evaluación debe abordarse con precaución y considerando las circunstancias específicas de cada caso, ya que apartarse de las recomendaciones de un sistema de IA no constituye, por sí mismo, una conducta negligente.

Téngase en cuenta que la responsabilidad médica varía según el tipo de centro en el que se presten los servicios. En centros públicos, la responsabilidad recae en la Administración, que solo podrá repercutir contra el médico en casos de dolo o culpa grave, de conformidad con el art. 36 de la Ley 40/2015. En centros privados, la responsabilidad del facultativo por negligencia es personal, y solo podría reclamarse contra el centro si se demuestra su culpa en la organización o supervisión del servicio, según el art. 1903.5 del Código Civil.

En definitiva, el uso de IA en el ámbito médico español debe integrarse en el marco normativo existente, combinando los principios de la *lex artis ad hoc* con las nuevas disposiciones sobre responsabilidad por productos defectuosos e inteligencia artificial.

# *Conclusiones*

Los sesgos algorítmicos, especialmente los derivados de datos de entrenamiento deficientes, pueden generar exclusiones, diagnósticos erróneos y tratamientos inadecuados que afectan a la equidad y también a la precisión de los diagnósticos. La regulación vigente ha supuesto un avance en esta dirección, pero, debe ser más ambiciosa en garantizar que la inteligencia artificial aplicada a la sanidad no perpetúe desigualdades estructurales ni amplifique los riesgos inherentes a los sistemas algorítmicos.

El Reglamento de Productos Sanitarios (Reglamento (UE) 2017/745), el Reglamento de Inteligencia Artificial (Reglamento (UE) 2024/1689) y el Reglamento General de Protección de Datos (Reglamento (UE) 2016/679) constituyen pilares fundamentales en la prevención e indemnización de daños, pero presentan vacíos que deben ser abordados para garantizar el despliegue de una IA ética y fiable en el sector sanitario.

El éxito de la inteligencia artificial en sanidad depende, en primer lugar, de un compromiso firme con la calidad y representatividad de los datos. Es esencial desarrollar estándares internacionales vinculantes para la depuración de datos (art. 10 RIA y art. 5 RGPD) y exigir que los sistemas de IA puedan adaptarse a nuevos datos y contextos sin comprometer su fiabilidad. En este sentido, es necesario implementar guías normalizadas para un diseño ético de los algoritmos de inteligencia artificial que incorpore desde el inicio requisitos estrictos de inclusión de datos representativos y de depuración de datos históricos.

Además, es imprescindible armonizar las disposiciones sobre las evaluaciones de impacto en derechos fundamentales (art. 35 RGPD y art. 27 RIA) y sobre las medidas de mitigación derivadas de dicha evaluación, con el fin de evitar redundancias y asegurar una protección más efectiva de los derechos individuales. La implementación de mecanismos de supervisión humana activa durante todo el ciclo de vida del producto, como exige el art. 14 del RIA, puede consolidar un sistema integral de vigilancia y mitigación de los sesgos.

El derecho a recibir explicaciones significativas sobre la lógica algorítmica que subyace en las decisiones automatizadas (art. 86 del RIA y art. 22 RGPD) debe fortalecerse mediante directrices prácticas que aseguren tanto la comprensión técnica de los operadores como la información acce-

sible para los pacientes. La desconexión entre el requisito técnico de transparencia y comunicación de la información a los responsables del despliegue (art. 13 RIA) con el derecho del afectado a comprender las decisiones automatizadas (art. 86 del RIA, art. 22 RGPD) exigen una interpretación conjunta que garantice explicaciones comprensibles tanto para operadores como para pacientes.

El Reglamento de Productos Sanitarios (Reglamento (UE) 2017/745) ha representado un avance importante en la regulación de los dispositivos médicos que incorporan inteligencia artificial, al establecer requisitos específicos en materia de seguridad y eficacia. Entre estos destacan las disposiciones relativas a la trazabilidad y transparencia en los procesos de diseño, fabricación y evaluación de dispositivos, que resultan especialmente relevantes para abordar los riesgos asociados a los sistemas de IA. No obstante, es necesario un esfuerzo adicional para garantizar que estas disposiciones se adapten a las características particulares de los sistemas de IA sanitaria, especialmente en lo que respecta a la supervisión de su desempeño tras la puesta en servicio o introducción en el mercado. Se incluyen mecanismos adecuados para identificar y mitigar riesgos emergentes en tiempo real, como la aparición de sesgos algorítmicos o fallos en las actualizaciones de software que podrían comprometer la seguridad de los pacientes.

La implementación efectiva del Reglamento exige también una armonización con otros marcos normativos, como el Reglamento de Inteligencia Artificial y el Reglamento General de Protección de Datos, para evitar duplicidades y garantizar una supervisión integral. En este sentido, resulta prioritario reforzar las capacidades de las autoridades competentes en la evaluación de dispositivos médicos con componentes de IA, mediante la formación especializada y la colaboración transfronteriza. Además, deben impulsarse estándares técnicos claros y vinculantes que orienten el diseño ético de los sistemas de IA médica desde su concepción hasta su uso final.

En materia de responsabilidad por los daños generados por la IA, la inclusión del software y los servicios digitales conexos en la definición de 'producto' según la Directiva (UE) 2024/2853 sobre Productos Defectuosos representa un avance significativo al establecer nuevas obligaciones para fabricantes, desplegadores y actualizadores. Este planteamiento amplía las responsabilidades más allá de la comercialización inicial, abarcando actualizaciones y modificaciones significativas, especialmente en productos con capacidades de autoaprendizaje (art. 7.2.e de la Directiva 2024/2853).

El nuevo marco normativo ofrece mayor protección a los consumidores frente a tecnologías emergentes como la IA, como los derivados de actuali-

zaciones mal diseñadas o comportamientos imprevisibles, y refuerza la rendición de cuentas a lo largo de toda la cadena de suministro. No obstante, el reto actual radica en la implementación efectiva de estas disposiciones, incluyendo el desarrollo de herramientas normativas para evaluar la seguridad tras la comercialización o puesta en servicio conforme a lo exigido por la Directiva sobre Responsabilidad de Productos Defectuosos y el Reglamento de Productos Sanitarios.

La interpretación sistemática de la Directiva de Productos Defectuosos, la Propuesta de Directiva de Responsabilidad en IA y los marcos nacionales exige soluciones capaces de permitir adaptaciones locales sin comprometer los estándares europeos.

En España, la combinación del régimen de responsabilidad objetiva por productos defectuosos (Directiva 2024/2853) —que exige una actualización del TRLGCU— y las disposiciones del Código Civil (art. 1.902) debe articularse de forma que permita abordar la diversidad de daños físicos, psicológicos y morales asociadas a la IA sanitaria. La incorporación de herramientas probatorias como la regla *res ipsa loquitur* y la ampliación de mecanismos como los establecidos en el art. 9 de la Propuesta de Directiva sobre exhibición de pruebas favorecería a las víctimas a la hora de reclamar justicia en litigios complejos.

Por otra parte, es imprescindible avanzar hacia una regulación más adaptativa que contemple revisiones periódicas del marco normativo, permitiendo su actualización frente a los continuos avances tecnológicos y las necesidades emergentes de los sistemas de salud. De *lege ferenda*, sería deseable desarrollar instrumentos jurídicos que fortalezcan los mecanismos de transparencia y rendición de cuentas en todas las etapas del ciclo de vida de los productos sanitarios con IA, como, por ejemplo, auditorías algorítmicas obligatorias, realizadas por organismos independientes, y la creación de un registro europeo de incidentes relacionados con sistemas de IA sanitaria, para facilitar el análisis y la corrección de fallos sistémicos.

En definitiva, el éxito de la inteligencia artificial en el ámbito sanitario dependerá de la capacidad de los legisladores y reguladores para garantizar un equilibrio entre la innovación tecnológica y la protección de los derechos fundamentales, promoviendo un modelo de gobernanza que combine exigencias técnicas rigurosas con principios éticos ampliamente aceptados. Una regulación integradora debería contribuir a mitigar los riesgos inherentes a los sistemas de IA tanto como a impulsar una inteligencia artificial inclusiva, fiable y orientada al servicio de la salud pública.

# *Referencias bibliográficas*

**Normativa citada**


Comisión Europea (2021). Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión. COM(2021) 206 final.

Comisión Europea (2022). Comunicación de la Comisión "Guía azul" sobre la aplicación de la normativa europea relativa a los productos. https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=OJ:C:2022:247:FULL

Consejo de Europa. Convenio para la protección de las personas con respecto al tratamiento automatizado de datos de carácter personal (Convenio 108+). Estrasburgo, 18 de mayo de 2018. Disponible en: https://www.coe.int/en/web/data-protection/convention108.

Decisión nº 768/2008/CE del Parlamento Europeo y del Consejo, de 9 de julio de 2008, sobre un marco común para la comercialización de los productos y por la que se deroga la Decisión 93/465/CEE del Consejo.

Directiva (UE) 2000/43/CE del Consejo, de 29 de junio de 2000, relativa a la aplicación del principio de igualdad de trato de las personas independientemente de su origen racial o étnico.

Directiva (UE) 2000/78/CE del Consejo, 27 de noviembre de 2000, relativa al establecimiento de un marco general para la igualdad de trato en el empleo y la ocupación.

Directiva (UE) 2004/113/CE del Consejo, de 13 de diciembre de 2004, por la que se aplica el principio de igualdad de trato entre hombres y mujeres al acceso a bienes y servicios y su suministro

Directiva (UE) 2006/54/CE, del Parlamento Europeo y del Consejo, de 5 de julio de 2006, relativa a la aplicación del principio de igualdad de oportunidades e igualdad de trato entre hombres y mujeres en asuntos de empleo y ocupación (refundición).

Directiva (UE) 2019/882 del Parlamento Europeo y del Consejo, de 17 de abril de 2019, sobre los requisitos de accesibilidad de los productos y servicios

Directiva (UE) 2022/2555 del Parlamento Europeo y del Consejo, de 14 de diciembre de 2022, relativa a medidas destinadas a garantizar un elevado nivel común de ciberseguridad en la Unión (Directiva NIS 2).

Directiva (UE) 2024/2853 del Parlamento Europeo y del Consejo, de 23 de octubre de 2024, sobre responsabilidad por los daños causados por productos defectuosos y por la que se deroga la Directiva 85/374/CEE del Consejo.

Directiva (UE) 97/80/CE, del Consejo, de 15 de diciembre de 1997, relativa a la carga de la prueba en los casos de discriminación por razón de sexo.

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.

Ley 29/2006, de 26 de julio, de garantías y uso racional de los medicamentos y productos sanitarios.

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

Ley 11/2023, de 8 de mayo, de trasposición de Directivas de la Unión Europea en materia de accesibilidad de determinados productos y servicios, migración de personas altamente cualificadas, tributaria y digitalización de actuaciones notariales y registrales; y por la que se modifica la Ley 12/2011, de 27 de mayo, sobre responsabilidad civil por daños nucleares o producidos por materiales radiactivos.

Propuesta de DIRECTIVA DEL PARLAMENTO EUROPEO Y DEL CONSEJO relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial (Directiva sobre responsabilidad en materia de IA)

Propuesta de DIRECTIVA DEL PARLAMENTO EUROPEO Y DEL CONSEJO relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial (Directiva sobre responsabilidad en materia de IA), Bruselas, 28.9.2022, COM (2022) 496 final, 2022/0303 (COD).

Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y por el que se modifican determinados actos legislativos de la Unión – Evaluación de Impacto, SWD (2021) 85 final, Comisión Europea, Bruselas, 21 de abril de 2021.

Proyecto de Real Decreto por el que se regulan los procedimientos de financiación y precio de los medicamentos. https://www.sanidad.gob.es/normativa/docs/CPP_RD_PRECIO_MED.pdf

Real Decreto 1345/2007, de 11 de octubre, por el que se regula el procedimiento de autorización, registro y condiciones de dispensación de los medicamentos de uso humano fabricados industrialmente.

Real Decreto 192/2023, de 21 de marzo, por el que se regulan los productos sanitarios.

Real Decreto 729/2023, de 22 de agosto, por el que se aprueba el Estatuto de la Agencia Española de Supervisión de Inteligencia Artificial. Boletín Oficial del Estado (BOE), núm. 210, de 2 de septiembre de 2023, páginas 122289 a 122316.

Real Decreto 729/2023, de 22 de agosto, por el que se aprueba el Estatuto de la Agencia Española de Supervisión de Inteligencia Artificial.

Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos)

Reglamento (UE) 2017/745 del Parlamento Europeo y del Consejo, de 5 de abril de 2017, sobre los productos sanitarios, por el que se derogan las Directivas 90/385/CEE y 93/42/CEE del Consejo. Diario Oficial de la Unión Europea, L 117, 5.5.2017, p. 1-175. Reglamento de Productos Sanitarios (RPS):

Reglamento (UE) 2017/746 del Parlamento Europeo y del Consejo, de 5 de abril de 2017, relativo a los productos sanitarios para diagnóstico in vitro y por el que se deroga la Directiva 98/79/CE del Consejo y la Decisión 2010/227/UE de la Comisión.

Reglamento (UE) 2019/881 del Parlamento Europeo y del Consejo, de 17 de abril de 2019, relativo a ENISA (Agencia de la Unión Europea para la Ciberseguridad) y a la certificación de la ciberseguridad de las tecnologías de la información y la comunicación y por el que se deroga el Reglamento (UE) nº 526/2013 ("Reglamento sobre la Ciberseguridad").

Reglamento (UE) 2021/2282 del Parlamento Europeo y del Consejo de 15 de diciembre de 2021 sobre evaluación de las tecnologías sanitarias y por el que se modifica la Directiva 2011/24/UE.

Reglamento (UE) 2022/1925 del Parlamento Europeo y del Consejo de 14 de septiembre de 2022 sobre mercados disputables y equitativos en el sector digital y por el que se modifican las Directivas (UE) 2019/1937 y (UE) 2020/1828 (Reglamento de Mercados Digitales)

Reglamento (UE) 2022/2065 del Parlamento Europeo y del Consejo de 19 de octubre de 2022 relativo a un mercado único de servicios digitales y por el que se modifica la Directiva 2000/31/CE (Reglamento de Servicios Digitales)

Reglamento (UE) 2022/868 del Parlamento Europeo y del Consejo de 30 de mayo de 2022 relativo a la gobernanza europea de datos y por el que se modifica el Reglamento (UE) 2018/1724 (Reglamento de Gobernanza de Datos)

Reglamento (UE) 2023/988 del Parlamento Europeo y del Consejo, de 10 de mayo de 2023, sobre la seguridad general de los productos y por el que se deroga la Directiva 2001/95/CE y la Directiva 87/357/CEE del Consejo. Diario Oficial de la Unión Europea, L 135, 23.5.2023, p. 1-47. Reglamento General de Seguridad de los Productos (RGSP)

Reglamento (UE) 2023/988 relativo a la seguridad general de los productos, por el que se modifican el Reglamento (UE) nº 1025/2012 y la Directiva (UE) 2020/1828, y se derogan la Directiva 2001/95/CE y la Directiva 87/357/CEE.

Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) nº 300/2008, (UE) nº 167/2013, (UE) nº 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial).

Reglamento (UE) 2024/903 del Parlamento Europeo y del Consejo, de 13 de marzo de 2024, por el que se establecen medidas a fin de garantizar un alto nivel de interoperabilidad del sector público en toda la Unión (Reglamento sobre la Europa Interoperable).

Reglamento (UE) 2025/327 del Parlamento Europeo y del Consejo de 11 de febrero de 2025 relativo al Espacio Europeo de Datos de Salud, y por el que se modifican la Directiva 2011/24/UE y el Reglamento (UE) 2024/2847

**Jurisprudencia citada**

Audiencia Nacional, Sala de lo Contencioso-Administrativo (Sección Séptima). Sentencia núm. 2013/2024, de 30 de abril de 2024. TOL10.005.554

Juzgado Central de lo Contencioso-Administrativo n.º 8, de fecha 30 de diciembre de 2021, en el Procedimiento Ordinario 18/2019,

Tribunal de Apelación de Ámsterdam, Sentencia del de 4 de abril de 2023, Asunto 200.295.742/01, ECLI:NL:GHAMS:2023:793.

Tribunal de Justicia de la Unión Europea (Gran Sala), sentencia del de 7 de diciembre de 2023, Asunto C-634/21, OQ contra Land Hessen, con intervención de SCHUFA Holding AG. TOL9.882.990

Tribunal de Justicia de la Unión Europea, (Gran Sala), sentencia de 16 de julio de 2020 Asunto C-311/18 Data Protection Commissioner contra Facebook Ireland Limited y Maximillian Schrems Petición de decisión prejudicial planteada por la High Court (Irlande), TOL9.908.235

Tribunal de Justicia de la Unión Europea, sentencia 19 de abril de 2016, Asunto C-441/14 (Dansk Industri (DI), acting on behalf of Ajos A/S v Estate of Karsten Eigil Rasmussen. ECLI:EU:C:2016:278;

Tribunal de Justicia de la Unión Europea, sentencia de 10 de junio de 2021, Asunto C-65/20 (VI contra Krone), TOL9.907.951

Tribunal de Justicia de la Unión Europea, sentencia de 16 de febrero de 2017, Asunto C-219/15 (Schmitt/TÜV Rheinland), TOL5.958.920

Tribunal de Justicia de la Unión Europea, sentencia de 16 de febrero de 2017, Asunto C-219/15 (Schmitt contra TÜV Rheinland LGA Products GmbH), ECLI:EU-:C:2017:128.

Tribunal de Justicia de la Unión Europea, sentencia de 17 de abril de 2018, Asunto C-414/16, Vera Egenberger contra Evangelisches Werk für Diakonie und Entwicklung eV, TOL6.573.837

Tribunal de Justicia de la Unión Europea, sentencia de 17 de octubre de 1989, Asunto C-109/88, Handels- og Kontorfunktionærernes Forbund I Danmark (HK) contra Dansk Arbejdsgiverforening, actuando por Danfoss, ECLI:EU:C:1989:383.

Tribunal de Justicia de la Unión Europea, sentencia de 19 de enero de 2010, Asunto C-555/07 (Seda Kücükdeveci v Swedex GmbH & Co. KG. ECLI:EU:C:2010:21;

Tribunal de Justicia de la Unión Europea, sentencia de 22 de noviembre de 2005, Asunto C-144/04 (Werner Mangold v Rüdiger Helm), ECLI:EU:C:2005:709;

Tribunal de Justicia de la Unión Europea, sentencia de 26 de abril de 2022, Asunto C-401/19 (República de Polonia contra Parlamento Europeo y Consejo de la Unión Europea.), TOL8.914.622

Tribunal de Justicia de la Unión Europea, sentencia de 4 de julio de 2023, Asunto C-252/21 (Meta Platforms y otros), TOL9.908.883

Tribunal de Justicia de la Unión Europea, sentencia de de 7 de diciembre de 2017, Asunto C-399/16, Syndicat national de l'industrie des technologies médicales (Snitem) y Philips France contra Premier ministre y Ministre des Affaires sociales et de la Santé. TOL9.913.342

Tribunal de Justicia de la Unión Europea, sentencia del 19 de abril de 2012, Asunto C-415/10 (Galina Meister/Speech Design Carrier Systems GmbH, TOL9.916.431

Tribunal Ordinario de Bolonia, sentencia de 31 de diciembre de 2020, n.º 2949/2020, Caso Deliveroo

Tribunal Supremo, Sala de lo Contencioso-Administrativo, Sección Cuarta, Sentencia de 3 de diciembre de 2020 conocida como _Ala Octa_ Recurso de Casación núm. 382/2018

Verwaltungsgericht Wien (Austria), Cuestión prejudicial presentada ante el TJUE el 16 de marzo de 2022 — CK (Asunto C-203/22), https://curia.europa.eu/juris/document/document.jsf?docid=260303&doclang=es

**Bibliografía citada**

AEMPS (2022) *Instrucciones de la AEMPS para la realización de investigaciones clínicas con productos sanitarios en España,* https://www.aemps.gob.es/productos-sanitarios/investigacionclinica-productossanitarios/instrucciones-de-la-aemps-para-la-realizacion-de-investigaciones-clinicas-con-productos-sanitarios-en-espana/

AEPD (2024) *Synthetic data and data protection*. Disponible en: https://www.aepd.es/en/prensa-y-comunicacion/blog/synthetic-data-and-data-protection#:~:text=Synthetic%20data%2C%20as%20many%20other,the%20real%20personal%20data%20for.

Agencia de los Derechos Fundamentales de la Unión Europea y del Consejo de Europa (2018) *BigData: Discrimination in data-supported decision making*. https://fra.europa.eu/en/publication/2018/bigdatadiscrimination-data-supported-decision-making.

Agencia de los Derechos Fundamentales de la Unión Europea y del Consejo de Europa (2021). *Construir correctamente el futuro. La inteligencia artificial y los derechos fundamentales*. Luxemburgo: Oficina de Publicaciones de la Unión Europea.

AI Watch (2023) *Global regulatory tracker – Spain, White & Case LLP*. Disponible en https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-spain#:~:text=Currently%2C%20there%20are%20no%20specific,sectoral%20AI%20legislation

AlgoSoc. (2024). CK v Dun & Bradstreet Austria (C-203/22): *The case that ends the GDPR right to explanation debate?* https://algosoc.org/results/ck-v-dun-bradstreet-austria-c-203-22-the-case-that-ends-the-gdpr-right-to-explanation-debate

Alkorta Idiakez I. (2022-a) *El Espacio Europeo de Datos Sanitarios: Nuevos Enfoques de la Protección e Intercambio de Datos Sanitarios,* Cizur Menor: Thomson Reuters-Aranzadi.

Alkorta Idiakez, I. (2022-b), "Five crucial challenges for regulation of Medical artificial intelligence", *Revista internacional de los estudios vascos, RIEV*, Vol. 67, Nº. 2, 31-40.

Alkorta Idiakez, I. (2024). "La discriminación algorítmica en el sector sanitario". En J. A. Moreno y P.J. Femenía López, *Inteligencia Artificial y Derecho de Daños. Cuestiones Actuales: Acorde al Reglamento (UE) 2024/1689,* Madrid: Dykinson. 1-30.

ALTAI (2020). *Assessment List for Trustworthy Artificial Intelligence*, https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

Altman, D. G., y Bland, J. M. (1994). "Diagnostic tests 1: Sensitivity and specificity". *BMJ*, 308(6943), 1552.Disponible en: https://doi.org/10.1136/bmj.308.6943.1552

ANCEI (2021), "Novedades legislativas en investigación con productos sanitarios ", *Boletín ANCEI*, vo. IV, nº. 1

Angwin, J., Larson, J., Mattu, S., y Kirchner, L. (2016). "Machine Bias". *ProPublica*. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Asociación de Maquinaria Computacional (ACM) (1972) "Guidelines for Professional Conduct in Information Processing", ACM, https://ethics.acm.org/code-of-ethics/previous-versions/1966-acm-code/

Atienza Navarro, M.L. (2024) "La responsabilidad civil por daños causados por la inteligencia artificial. Estado de la cuestión", *APPDC, Derecho de contratos, responsabilidad extracontractual e inteligencia artificial*, Cizur Menor: Aranzadi, 341-386.

Barba, V. (2023) *Principio de no discriminación y contrato*. Madrid: Colex.

Barnard, C., Peers, S. (2014). *European Union Law*. Oxford University Press, 674-683.

Barocas, S. y Selbst, A. D. (2016). "Big Data's Disparate Impact ". *California Law Review*, 104(3), 671-732.

Barocas, S., Hardt, M. y Narayanan, A. (2019) *Fairness and Machine Learning: Limitations and Opportunities*, Ed. fairmlbook.org,

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

Beam, A. L., y Kohane, I. S. (2018). "Big data and machine learning in health care ". *JAMA*, 319(13), 1317-1318.

Beauchamp, T. L., y Childress, J. F. (1979). *Principios de ética biomédica*. Oxford University Press.

Berenguer Albaladejo, C. (2024), "Transparencia y la explicabilidad en el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, de protección de datos (RGPD): especial referencia a las decisiones automatizadas del art. 22 ", (Eds.) J.A. Moreno y P.J. Femenía López, *Inteligencia Artificial y Derecho de Daños: Acorde al Reglamento (UE) 2024/1689,* Madrid: Dykinson, 49-111.

Binns, R. (2020). "Fairness in Machine Learning: Lessons from Political Philosophy". *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAccT),* 149-159. https://doi.org/10.1145/3351095.3372839

Boix Palop, A. (2022) "Transparencia en la utilización de inteligencia artificial por parte de la Administración"*, El Cronista del Estado Social y Democrático de Derecho, (Ejemplar dedicado a: Inteligencia artificial y derecho),* Nº 100 (SeptiembreOctubre) 90-105

Bosoer, L., Cantero Gamito, M. y Rubio-Marin, R. (2023). "Non-Discrimination and the AI Act" Casadei (ed.*), Law and Digitalization*. Arazandi. https://ssrn.com/abstract=4666071

Buolamwini, J. (2019). "AJL Report on Algorithmic Bias in Big Tech Companies". *Algorithmic Justice League*. https://www.ajl.org

Buolamwini, J. y Gebru, T. (2018) "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification ". In *Proceedings of Machine Learning Research*, 81, 77–91.

Burrell, J. (2016) "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms." *Big Data & Society,* vol. 3, no. 1, 1-12.

Bygrave, L. A. (2014). *Data Privacy Law: An International Perspective*. Oxford University Press.

Cavallo, P. "The AI Act: What will the impact be on the medical device industry?"; https://portolano.it/en/blog/life-sciences/ai-act-impact-medical-device-industry#:~:text=Specifically%2C%20Article%2010%20of%20the,system%20is%20intended%20to%20operate

Carrasco Perera, A. (1993), «STS 14 de diciembre de 1993. Propiedad intelectual. Derecho moral de autor. Indemnización del daño moral», Cuadernos Civitas de Jurisprudencia Civil, núm. 33, 1105-1119.

Carrasco Perera, Á. (2019-a) "El cártel de los camiones. Presunción y prueba del daño", en *Revista de Derecho de la Competencia y la Distribución,* núm. 25, 5-15.

Carrasco Perera, A. (2019-b), "A propósito de un trabajo de Gunter Teubner sobre la personificación civil de los agentes de inteligencia artificial avanzada (Digitale Rechtssubjekte? Zum privatrechtlichen Status autonomer Softwareagentem, Archiv für die Civilistische Praxis), 218, 2018, 155-205, *Centro de Estudios de Consumo CESCO*, 9-21.

Castilla Barea, M. (2024) "Vigilancia postcomercialización, códigos de conducta y directrices. Capítulo v." En M. Castilla Barea et al. (ed.) *El Reglamento de Inteligencia Artificial*, Tirant lo Blanch, 2024, pp. 177 a 180.

Channa R, Wolf R, Abramoff M. (2021) "Autonomous artificial intelligence in diabetic retinopathy: from algorithm to clinical application". *J Diabetes Sci Technol.* 15 695–698.

Chung, C.T. et al. (2022) "Clinical significance, challenges and limitations in using artificial intelligence for electrocardiography-based diagnosis" *Int J Arrhythmia*, 23:24.

Citron, D, Pasquale, F. (2014) "The Scored Society: Due Process for Automated Predictions." *Washington Law Review*, vol. 89, no. 1, 1-33.

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). *Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. Annals of Internal Medicine,* 162(1), 55-63.: https://doi.org/10.7326/M14-0697.

Collins et al. (2024) TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ, 2024, 385.

Comision Europea (COM) (2020-a) *Report from the Commission to the European Parliament, the Council and the European Economic and Social Committee. Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics COM*(2020) 64 final.

Comisión Europea (COM) (2020-b). *Libro Blanco sobre la Inteligencia Artificial: un enfoque europeo hacia la excelencia y la confianza.* COM(2020) 65 final. Bruselas, 19 de febrero de 2020. https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52020DC0065

Comisión Europea (COM) (2021) *Estudio de apoyo a la evaluación de impacto del Reglamento sobre la IA.*

Comité Europeo de Protección de Datos, (ECPD) (2024) *Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models*

Comité Europeo de Protección de Datos, (ECPD) (2021). *Statement 05/2021 on the Data Governance Act.*

Consejo de Europea (2020). *Recomendación CM/Rec(2020)1 del Comité de Ministros a los Estados miembros sobre los impactos humanos y éticos de la inteligencia artificial.* Estrasburgo: Consejo de Europa.

Cotino Hueso, L. (2024-a), "Caso Bosco, a la tercera tampoco va la vencida. Mal camino en el acceso los algoritmos públicos", *Sección Ciberderecho, Diario LA LEY*, Nº 84, 23-45

Cotino Hueso, L. (2024-b). *Derechos Digitales y Transparencia Algorítmica: Desafíos para el Derecho Español y Europeo.* Valencia: Tirant Lo Blanch.

Craig, P., De Búrca, G. (2020). *EU Law: Text, Cases, and Materials.* Oxford University Press, 98-108.

Crenshaw, K. (1989), "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics." *University of Chicago Legal Forum*, vol. no. 1, 139-167.

Cueto Pérez, M. (2022), "Jurisprudencia en el caso ALA OCTA: responsabilidad patrimonial por la utilización de productos defectuosos en el ámbito sanitario", *Revista de Administración Pública*, 217, 172-203.

De Miguel Beriain, I. y Mursuli Yanes, Y. Sesgos (2023) "IA y biomedicina: un comentario desde la ética y el Derecho", *Revista de Derecho y Genoma Humano: genética, biotecnología y medicina avanzada*, Nº 59, 129-148

De Mol, M. (2011) "The Novel Approach of the CJEU on the Horizontal Direct Effect of the EU Principle of Non-discrimination: (Unbridled) Expansionism of EU Law?" *Maastricht Journal of European and Comparative Law*, 18(1-2), 109-135.

De Schutter, O. (2010). Proving Discrimination: The Role of Situation Testing. European Commission.

Demkova, S. (2024) "The AI Act's Right to Explanation: A Plea for an Integrated Remedy" Law and Policy of the Media in a *Comparative Perspective October,* 31, 43-78

Dignum, V. (2009) *Handbook of Research on Multi-agent Systems: Semantics and Dynamics of Organizational Models. Information Science Reference.*

Dignum, V. (2019) *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way.* Springer.

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., y Zemel, R. (2012), "Fairness Through Awareness ". In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference ACM,* 214-226.

EFF (2020). *Why Google Firing an Ethical AI Researcher Harms the Quest for Accountable AI.* https://www.eff.org/deeplinks/2020/12

Elmore, J. G., Armstrong, K., Lehman, C. D., & Fletcher, S. W. (2005). "Screening for breast cancer ". *JAMA*, 293(10), 1245-1256.

Emiroglu, M, et al. (2022) "National study on use of artificial intelligence in breast disease and cancer" *Bratisl Lek Listy*, vol 123, 191–196.

EMA (2024), *Reflection-paper-use-artificial-intelligence-ai-medicinal-product*, Disponible en: https://www.ema.europa.eu/system/files/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle-en.pdf

ENISA. (2023). *Artificial Intelligence Cybersecurity Challenges: Threat Landscape for Artificial Intelligence*. European Union Agency for Cybersecurity. https://www.enisa.europa.eu/

Escobar Mimbrera, Á. (2021). *Métricas y guías para el desarrollo de algoritmos de inteligencia artificial explicable*. Universidad de Jaén, 2021.

Eubanks, V. (2018), *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press.

European Data Protection Board (EDPB) (2024), Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the contexto f AI models, 2024.

European Digital Rights (EDRi) (2021). *An EU Artificial Intelligence Act for Fundamental Rights: A Civil Society Statement*. https://edri.org/wp-content/uploads/2021/12/Political-statement-on-AI-Act.pdf.

European Medicines Agency (EMA) (2024). Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle. https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle_en.pdf

European Law Institute (ELI) (2023a). *Guiding Principles and Model Rules on Algorithmic Contracts. Interim Report*. https://www.europeanlawinstitute.eu

European Medicines Agency (EMA) (2023b). Multi-annual Artificial Intelligence Workplan 2023-2028: HMA-EMA Joint Big Data Steering Group

European Parliament Research Service (EPRS) (2021). *The Artificial Intelligence Act: Ensuring a balance between innovation and fundamental rights*. https://epthinktank.eu.

Evangelio Llorca, R. (2024). "Responsabilidad civil e inteligencia artificial en el ámbito sanitario: posibles vías de reclamación". En J. A. Moreno Martínez & P. J. Femenía López (Coords.), *Inteligencia artificial y derecho de daños: Cuestiones actuales: Acorde al Reglamento (UE) 2024/1689,* Madrid: Dykinson, 149-214.

Evans, R. y Cabral, P. (2023). *Product Liability in the context of AI-driven medical devices.*

Fernández, C. (2022). "Supervisión humana en sistemas de IA: Retos y perspectivas ". *Revista de Derecho y Tecnología,* 14(4), 101-125.

Floridi, L., y Cowls, J. (2022). *AI Ethics: A Revaluation*. Springer.

Floridi, L. (2021) *Ethics, Governance, and Policies in Artificial Intelligence*. Springer

Floridi, L. (2023) *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*. Oxford University Press.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., y Schafer, B. (2021). "AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations". *Minds and Machines*, 31(1), 1–17

Fredman, S. (2011). *Discrimination Law* (2.ª ed.). Oxford University Press.

Fredman, S. (2016). *Intersectional Discrimination in EU Gender Equality and Non-Discrimination Law*. European Commission

Friedler, S. A., Scheidegger, C., y Venkatasubramanian, S. (2016) "On the (im)possibility of fairness ". arXiv preprint arXiv:1609.07236.

Friedman, L. M., Furberg, C. D., y DeMets, D. L. (2015). *Fundamentals of Clinical* Trials (5ª ed.). Springer.

Future of Life Institute (2017). *Asilomar AI Principles*. https://futureoflife.org/open-letter/ai-principles/

Future-AI (2023) *Best practices for trustworthy AI in medicine*, https://future-ai.eu/

Gallifant, J. et al. (2025) "The TRIPOD-LLM reporting guidelines for the studies using large language models". *Nature Medicine*, VOL. 31, 2025, 60-69.

Gal, M., Lynskey, O. (2023) "Synthetic Data: Legal Implications of the Data-Generation Revolution", LSE Legal Studies Working Paper No. 6/23

García Micó, T.G. (2024) *Robótica quirúrgica y Derecho de daños*, Barcelona: Col.legi Notarial de Catalunya.

García, J. y Martínez, L. (2023). "Regulación de la Inteligencia Artificial en la Unión Europea: Análisis del RIA". *Revista de Derecho y Tecnología*, 15(2), 45-67.

Gebru, T., Bender, E. M., McMillan-Major, A., & Shmitchell, S. (2021). "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT),* 610–623. https://dl.acm.org/doi/10.1145/3442188.3445922

Gerke, S., Minssen, T., Cohen, G., (2022), *Chapter 12 - Ethical and legal challenges of artificial intelligence-driven healthcare*, 301-322.

Gil Membrado, C (2021), "Una asignatura pendiente: la regulación de la prestación sanitaria a través de telemedicina", en *El nuevo marco legal del teletrabajo en España. Presente y futuro. Una aproximación multidisciplinar*", (ed.) Atienza Macías, E., Madrid: Bosch.

Gil Membrado, C (2021), *Bioderecho y Retos. M-Health, Genética, IA, Robótica y Criogenización*, Madrid: Dykinson.

Gil Membrado, C. (2023). *Derecho y Medicina: Desafíos Tecnológicos y Científicos*, Madrid: Dykinson.

Global Partnership on Artificial Intelligence (GPAI) (2020). *Global Partnership on Artificial Intelligence*. https://gpai.ai/

Global Partnership on Artificial Intelligence (GPAI) (2023-a). Report on Bias in AI Systems ". *GPAI Annual Report 2023*. https://gpai.ai/reports/bias-in-ai-systems.

Global Partnership on Artificial Intelligence (GPAI) (2023-b). *White Paper on Ethical AI*. https://gpai.ai/

Goodfellow, I., Bengio, Y., y Courville, A. (2016). *Deep Learning*. Cambridge: MIT Press.

Greser, J. (2023) "Synthetic data and Medical AI – where do we stand" *LSTS*; https://lsts.research.vub.be/synthetic-data-and-medical-ai-where-do-we-stand

Grimalt Servera, P. (2023), "La responsabilidad (civil o patrimonial) del personal médico, de los centros sanitarios y de las administraciones sanitarias por los daños causados por un product sanitario defectuoso", *Bioderecho.es*, Núm. 17, 56-87.

Grimes, D. A., & Schulz, K. F. (2002). "Uses and abuses of screening tests ". *The Lancet*, 359 (9309), 881-884.

Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial (EGTAI) (2019). *Directrices éticas para una IA digna de confianza*. Brussels: European Commission.

Grupo Europeo de Ética de la Ciencia y de las Nuevas Tecnologías (EGE) (2018). *Declaración sobre inteligencia artificial, robótica y sistemas "autónomos*". https://ec.europa.eu/info/publications/ege-statement-artificial-intelligence-robotics-and-autonomous-systems_en

Hacker, P. (2021) "Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies against Algorithmic Discrimination under EU Law" *European Review of Private Law*, 26(4), 579-602.

Hacker, P. (2022) "The European AI Liability Directives—Critique of a Half-Hearted Approach and Lessons for the Future", *Computer Law & Security Review,* Vol. 51, November 2023, 105871

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

Hildebrandt, M. (2015). *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology*. Edward Elgar Publishing.

Hoffmann, A. L. (2019), "Where Fairness Fails: Data, Algorithms, and the Limits of Antidiscrimination Discourse". *Information, Communication & Society*, vol. 22, no. 7, 900-915.

Hsieh, W., et al. (2024) "A Comprehensive Guide to Explainable AI: From Classical Models to LLMs". https://arxiv.org/abs/2412.00800

IEEE (2017). "Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems". Global Initiative on Ethics of Autonomous and Intelligent Systems IEEE.

International Medical Device Regulators Forum (IMDRF) (2014). *Software as a Medical Device": Possible Framework for Risk Categorization and Corresponding Considerations*, https://www.gmp-compliance.org/files/guidemgr/imdrf-tech-140918-samd-framework-risk-categorization-141013.pdf.

International Medical Device Regulators Forum (IMDRF) (2025). IMDRF Good Machine Learning Practice (GMLP) for Medical Devices (N88, 2025). Published January 29, 2025. https://www.imdrf.org/sites/default/files/2025-02/IMDRF_AIML%20WG_GMLP_N88%20Final.pdf.

Innerarity, D. (2024). "Defensa y crítica de la gobernanza algorítmica ", *CIDOB d'Afers Internacionals*, n.º 138, p. 11-25

Innerarity, D. (2025). *Crítica de la razón algorítmica*, Barcelona: Galaxia Guttenberg. En imprenta.

Johnson, D. G. (1985) *Computer Ethics*. Prentice-Hall.

Johnson, L. (2023). "Explainable AI: Bridging the Gap Between AI and Human Understanding." *International Journal of AI Research*, 11(4), 45-60.

Jorqui Azofra, M. (2023). *Responsabilidad por los daños causados por productos y sistemas de inteligencia artificial*. Madrid: Dykinson

Jorqui Azofra, M. (2024). "Encaje del sistema de Inteligencia Artificial utilizado con determinados fines médicos en algunas de las cuestiones suscitadas al amparo del

régimen de responsabilidad por productos defectuosos". En Moreno Martínez, J. A., y Femenía López, P. J., *Inteligencia Artificial y Derecho de Daños: Acorde al Reglamento (UE) 2024/1689*, Madrid: Dykinson, 149-214.

Kahan, B. C., Jairath, V., Doré, C. J., y Morris, T. P. (2014). "The risks and rewards of covariate adjustment in randomized trials: an assessment of 12 outcomes from 8 studies" *Trials*, 15(1), 139. https://doi.org/10.1186/1745-6215-15-139

Kaushik, P., Venkatesh, M., Raza, J., Kvedar, C. (2022) "Health digital twins as tools for precision medicine: Considerations for computation, implementation, and regulation", *NPJ Digital Medicine* 5/23.

Kiseleva, A. (2020) "AI as a Medical Device: Is It Enough to Ensure Performance Transparency and Accountability in Healthcare?" *European Pharmaceutical Law Review*, 1/28.

Kiskinov, V. (2024). "Preliminary Comments on European Union's Artificial Intelligence Act ". Academia.edu

Kleinberg, Ludwig, Mul-Lainathan y Sunstein, M. (2018) "Discrimination in the Age of Algorithms", *Journal of Legal Analysis*, 10, 174.

Kuner, C., Bygrave, L. A., y Docksey, C. (Eds.). (2020). *The EU General Data Protection Regulation (GDPR): A Commentary*. Oxford University Press.

Lagioia, F., Rovatti, R., y Sartor, G. (2022). "Algorithmic fairness through group parities? The case of COMPAS-SAPMOC". *AI & Society*, 38(2), 459-478.

Lazcoz Moratinos, G. (2023) *Sistemas de inteligencia artificial en la asistencia sanitaria: cómo garantizar la supervisión humana desde la normativa de protección de datos*. https://www.aepd.es/documento/premio-emilio-aced-2022-guillermo-lazcoz.pdf

Lekadir, K., Feragen, A., Fofanah, A. J., Frangi, A. F., Buyx, A., et al. (2024). *FUTURE-AI: International consensus guideline for trustworthy and deployable artificial intelligence in healthcarear*. https://arxiv.org/abs/2309.12325

Lenaerts, K. (2012) "Exploring the Limits of the EU Charter of Fundamental Rights", *European Constitutional Law Review*, 8(3) 375-403.

Liu, X., Cruz Rivera, S., Moher, D., Calvert, M. J., Denniston, A. K., & SPIRIT-AI and CONSORT-AI Working Group. (2020). Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nature Medicine, 26(9), 1364–1374. https://doi.org/10.1038/s41591-020-1034-x

López Tur, T. (2023). "Inteligencia artificial y responsabilidad civil en el ámbito sanitario". En López Sánchez, C., & Ortiz Fernández, M., *Derecho y salud: retos jurídicos actuales*. Cizur menor: Aranzadi.

López, M. (2022). "Derechos Fundamentales y Regulación de la IA: Un Estudio del Reglamento de Inteligencia Artificial". *Journal of European Law*, 28(3), 123-140.

Luquín Bergareche, R. (2024), "Responsabilidad contractual por el uso de la IA en la prestación de servicios de telemedicina y aplicaciones de salud digital ", *APPDS, Derecho de Contratos, Responsabilidad extracontractual e Inteligencia Artificial*, Cizur Menor: Aranzadi, 265-295.

Maner, W. (1993), "Computer Ethics." In *Encyclopedia of Computer Science*, edited by A. Ralston and E. D. Reilly, 3rd ed. New York: Van Nostrand Reinhold, 268-272.

Mantelero, A. (2018). "AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment". *Computer Law & Security Review*, 34(4), 754–772.

Martín-Casals, M. (2011), "La modernización del Derecho de la responsabilidad extracontractual ", *Cuestiones actuales en materia de responsabilidad civil*, Universidad de Murcia, Servicio de Publicaciones, 11-112.

Medical Device Coordination Group (MDCG). (2019). *Guía MDCG 2019-11. Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 – MDR, and Regulation (EU) 2017/746 – IVDR.*

Medical Device Coordination Group (MDCG). (2021). *Guía MDCG 2021-24: Calificación y clasificación del software en el Reglamento (UE) 2017/745 y el Reglamento (UE) 2017/746.* Disponible en: https://health.ec.europa.eu/system/files/2021-10/mdcg_2021-24_en_0.pdf.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., y Galstyan, A. (2021). "A survey on bias and fairness in machine learning ". *ACM Computing Surveys (CSUR)*, 54(6), 1-35.

Ministerio de Asuntos Económicos y Transformación Digital Gobierno de España (2021) *Carta de Derechos Digitales.* https://espanadigital.gob.es/lineas-de-actuacion/carta-de-derechos-digitales] (https://espanadigital.gob.es/lineas-de-actuacion/carta-de-derechos-digitales)

Mökander, J., Sheth, M., Gersbro-Sundler, M., Blomgren, P., y Floridi, L. (2024). "Challenges and Best Practices in Corporate AI Governance: Lessons from the Biopharmaceutical Industry." arXiv preprint arXiv:2407.05339.

Moor, J. H. (1985) "What is Computer Ethics?" *Metaphilosophy*, vol. 16, no. 4, 266-275

Morley, J., Shears, P., y Floridi, L. (2023). *Data Governance and AI: From Compliance to Competitive Advantage.* Oxford University Press.

Morondo Taramundi, D., Eguiluz Castañeira, J. y Álvarez, T. (2022) "La discriminación algorítmica en España: límites y potencial del marco legal ". *Digital Future Society.*

Navas Navarro, S. (2019) "Responsabilidad civil del fabricante y tecnología inteligente. Una mirada al futuro", *Diario La Ley*, Nº 35, Sección Ciberderecho, 27 de diciembre de 2019, LA LEY 15349/2019.

Nicolas Jimenez, P. (2019) "Los derechos sobre los datos utilizados con fines de investigación médica ante los nuevos retos tecnológicos y científicos", *Revista de derecho y genoma humano: genética, biotecnología y medicina avanzada*, Nº Extra 1 (Ejemplar dedicado a: Uso de datos clínicos ante nuevos escenarios tecnológicos y científicos. Oportunidades e implicaciones jurídicas), 129-167.

Nikolenko, S., (2021) *Synthetic Data for Deep Learning*, Springer

Nnawuchi, U. y George, C. A(2024), Grand Entrance Without a Blueprint: A Critical Analysis of the Right to Explanation in Article 86 of the European Union Artificial Intelligence Act. Lexxion, Volume 1 Issue 4, 402 - 419

Novelli, C., Hacker, P., Morley, J., Trondal, J., y Floridi, L. (2024). "A Robust Governance for the AI Act: AI Office, AI Board, Scientific Panel, and National Authorities" *European Journal of Risk Regulation.*, 0[10.1017/err.2024.57]

O'Neil, C. (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* Crown Publishing Group.

Obermeyer, Z., Powers, B., Vogeli, C., y Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447-453.

OCDE (2019). *Recommendation of the Council on Artificial Intelligence*. Paris: OECD Publishing.

OMS (2021). *Ética y gobernanza de la inteligencia artificial para la salud: directrices de la OMS*. Geneva: World Health Organization.

ONU (2020). *AI For Good*. https://aiforgood.itu.int/

Orwat, C., et al. (2022). *Normative Challenges of Risk Regulation of Artificial Intelligence and Automated Decision-Making*. https://arxiv.org/abs/2211.06203

Papernot, N., McDaniel, P., Goodfellow, I., Jha, S., Celik, Z. B., y Swami, A. (2017). "Practical Black-Box Attacks against Machine Learning". *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security (AsiaCCS)*, 506-519.

Parissi-Poumian, Y. (2021) "Apuntes para el análisis del sesgo en la investigación en Ciencias de la Salud ", *Revista de Medicina de la Universidad Veracruz*, 1, 67-89.

Parlamento Europeo, el Consejo y la Comisión (2023). *Declaración Europea sobre los Derechos y Principios Digitales para la Década Digital 2023/C 23/01. PUB/2023/89*. Disponible en: https://eur-lex.europa.eu/legal-content/ES/TXT/?toc=OJ%3AC%3A2023%3A023%3ATOC&uri=uriserv%3AOJ.C_.2023.023.01.0001.01.SPA

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

Pecqueux M, et al. (2022): "The use and future perspective of Artificial Intelligence-A survey among German surgeons" *Front Public Health*. 10: 982335.

Peña Lopez F. (2024), "Responsabilidad objetiva y subjetiva en las propuestas legislativas europeas sobre responsabilidad civil aplicables a la inteligencia artificial", *APP-DC, Derecho de contratos, responsabilidad extracontractual e inteligencia artificial,* Cizur Menor: Aranzadi, 341-386.

Pereira, I. G., da Silva Junior, A. G., Aragão, D. P., de Oliveira, E. V., Bezerra, A. A., Pereira, F. de A., Costa, J. G. F. S., Cuno, J. S., dos Santos, D. H., Guerin, J. M., Conci, A., Clua, E. W. G., Distante, C., & Gonçalves, L. M. G. (2022). "Epidemiology Forecasting of COVID-19 Using AI—A Survey ". En *Computational Intelligence for COVID-19 and Future Pandemics*, Springer, 89-120

Pérez, A., y Sánchez, R. (2023). "La supervisión humana en la IA: Un análisis desde el Derecho Internacional y la normativa europea". Revista *de Estudios Jurídicos*, 19(1), 89-112.

Piantadosi, S. (2005). *Clinical Trials: A Methodologic Perspective* (2nd ed.). Wiley.

Pontifical Academy for Life (2020). *Rome Call for AI Ethics*. Vatican City: Pontifical Academy for Life.

Popejoy, A. B., y Fullerton, S. M. (2016). "Genomics is failing on diversity ". *Nature*, 538(7624), 161–164.

Presno Linera, M. Á. y Meuwese, A. (2024). "The regulation of artificial intelligence in Europe". *Teoria y Realidad Constitutional*, (54), 131-161. https://doi.org/10.5944/trc.54.2024.43310

Rajesh, A. E., Davidson, C., Aaron, S., Lee, Y. (2023), "Artificial Intelligence and Diabetic Retinopathy: AI Framework, Prospective Studies, Head-to-head Validation, and Cost-effectiveness ". *Diabetes Care* 1 October; 46 (10): 1728–1739.

Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., Liu, P. J., Liu, X., Marcus, J., Sun, M., Sundberg, P., Yee, H., Zhang, K., Zhang, Y., Flores, G., Duggan, G. E., Irvine, J., Le, Q., & Dean, J. (2018). "Scalable and accurate deep learning with electronic health records". npj Digital Medicine, 1(1), 18.

Ramírez, J. D., Herrera, G., Martínez-Salazar, S., y Vallejo, G. A. (2017). The impact of standardization and training in data quality: A Colombian case study. *Journal of Clinical Epidemiology*, 85, 15-21. https://doi.org/10.1016/j.jclinepi.2016.12.001

Red de Sanidad Electrónica. Comisión Europea (2022). *Principios Europeos para la Ética en la Salud Digital*. Adoptado el 26 de enero de 2022. https://sante.gouv.fr/IMG/pdf/220131_european_ethical_principles_for_digital_health_fr_eng.pdf

Reglero, Fernando, *Tratado de Responsabilidad Civil* (3ª ed.) Tomo III, Cizur Menor: Aranzadi, 2008

Renda, A. (2021) *Artificial Intelligence Ethics, governance and policy challenges Report of a CEPS Task Force*, CEPS, https://www.ceps.eu/ceps-publications/artificial-intelligence-ethics-governance-and-policy-challenges/

Richter, J. Vogel, S. (2020). "Illustration of Clinical Decision Support System Development Complexity", In *The Importance of Health Informatics in Public Health during a Pandemic, Studies in Health Technology and Informatics*, Volume 272, 261-264.

Ritter, Z. et al. (2022) "Using Explainable Artificial Intelligence Models (ML) to Predict Suspected Diagnoses as Clinical Decision Support,", B. Séroussi et al. (Eds.*) Challenges of Trustable AI and Added-Value on Health European Federation for Medical Informatics (EFMI) and IOS Press*

Rivera, S. C., Liu, X., Chan, A. W., Denniston, A. K., Calvert, M. J., & SPIRIT-AI and CONSORT-AI Working Group. (2024). Directrices para los protocolos de ensayos clínicos de intervenciones con inteligencia artificial: la extensión SPIRIT-AI. Revista Panamericana de Salud Pública, 48, e13. https://doi.org/10.26633/RPSP.2024.13

Rubi Puig, A. (2024) "Inteligencia artificial y daños indemnizables ". *APPDC, Derecho de contratos, responsabilidad extracontractual e inteligencia artificial,* Cizur Menor: Aranzadi, 621-688

Rühlicke, S. (2024) *Entwicklung einer Bewertungsmetrik für klinische Entscheidungsunterstützungssysteme (CDSS) in der Kardiologie*, Tesis Doctoral, Universidad de Götinghem, https://ediss.uni-goettingen.de/handle/11858/15266

Schiek, D. (2016-b), "Revisiting Intersectionality for EU Anti-Discrimination Law in an Economic Crisis: A Critical Legal Studies Perspective." *Sociologia del Diritto*, vol. 43, no. 1, 9-27.

Schiek, D. (2016-a), *European Union Non-Discrimination Law and Intersectionality: Investigating the Triangle of Racial, Gender and Disability Discrimination*. Routledge.

Shneiderman, B. (2020), *Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy*. arXiv preprint arXiv:2022.04087.

Smith, J. (2024), "Regulation (EU) 2024/1689: Key Aspects of the European Artificial Intelligence Regulation" *European Journal of Law and Technology*, 12(3), 45-67.

Smith, J. (2024), "Transparency and Accountability in AI Systems." *Journal of Legal Technology*, 15(2), 89-102.

Soriano Soriano, A. (2020) *Algoritmos y discriminación. La importancia de calificar correctamente los supuestos de discriminación algorítmica directa.* Universidad de Valencia, Valencia, 56 y ss.

Stadler, T., Oprisanu, B., Troncoso, C., (2022) "Synthetic Data – Anonymisation Groundhog Day", https://www.usenix.org/conference/usenixsecurity22/presentation/stadler

Stöger, K., Kletecka-Pulker, M., y Wahlmüller, S. (2020), "Legal aspects of data cleansing in medical AI ", in C. Gasteiger & A. F. Holzinger (Eds.), *Digital Transformation in Healthcare*. Springer, 203-218

Stöger, K., Schneeberger, D. M. y Kieseberg, P. (2021) "Legal aspects of data cleansing in medical AI", *Computer Law & Security Review*, vol. 42, 105587.

Suresh, H., y Guttag, J. (2021). "A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle ". *Equity and Access in Algorithms, Mechanisms, and Optimization*. ACM. arXiv:1901.10002 [cs.LG]

SYNTHIA (2023), *Project initiated to progress use of synthetic data for personalized medicine*, Becaris Publishing). https://www.ihi-synthia.eu/

Sweeney, L. (2013) "Discrimination in Online Ad Delivery." *Communications of the ACM*, vol. 56, no. 5, 44-54.

Tahernejad, A., Sahebi, A., Abadi, A.S.S. et al. (2024) "Application of artificial intelligence in triage in emergencies and disasters: a systematic review". *BMC Public Health* 24, 3203.

Tejani, A. S., Klontzas, M. E., Gatti, A. A., Mongan, J. T., Moy, L., Park, S. H., & Kahn, C. E. Jr. (2024). Checklist for Artificial Intelligence in Medical Imaging (CLAIM): 2024 Update. Radiology: Artificial Intelligence, 6(4), e240300. https://doi.org/10.1148/ryai.240300

Trstenjak, V., Beysen, E. (2013) "European Law on Discrimination: How It Is Applied by the Court of Justice of the EU", *Common Market Law Review*, 50(6), 1527-1564.

UNESCO (2021). Recomendación sobre la Ética de la Inteligencia Artificial. https://unesdoc.unesco.org/ark:/48223/pf0000380455_spa

UNI Global Union (2020). *10 Principles for Ethical Artificial Intelligence*. https://uniglobalunion.org/report/10-principles-for-ethical-artificial-intelligence/

Unión Internacional de Telecomunicaciones (UIT) (2023). *Informe sobre la Gobernanza de la IA*. Ginebra: UIT. Unión Internacional de Telecomunicaciones (UIT).

Université de Montréal (2017). *Déclaration de Montréal pour un développement responsable de l'intelligence artificielle*. https://declarationmontreal-iaresponsable.com

Vasan, R. S. (2018). "The Framingham Heart Study: A historical perspective". *Lancet*, 391(10135), 1200–1211.

Veale, M., & Binns, R. (2017). "Fairer Machine Learning in the Real World: Mitigating Discrimination Without Collecting Sensitive Data". *Big Data & Society*, 4(2).

Wachter, S., Mittelstadt, B., y Floridi, L. (2017). "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation ". *International Data Privacy Law*, 7(2), 76–99.

Walters, J., Dey, D., Bhaumik, D., y Horsman, S. (2023). "Complying with the EU AI Act." arXiv preprint arXiv:2307.10458.

Wang, X., Zhao, J., Marostica, E. et al. (2024) "A pathology foundation model for cancer diagnosis and prognosis prediction ". Nature 634, 970–978.

Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F., y Kale, D. C. (2019). "Do no harm: A roadmap for responsible machine learning for health care ". Nature Medicine, 25(9), 1337-1340.

Williams, M., Brooks, T. y Shmargad, G. (2018) "How Algorithms Discriminate Based on Data They Lack" Journal of Information Policy, 8, 78.

Wynants, L., Van Calster, B., Collins, G. S., Riley, R. D., Heinze, G., Schuit, E., y Van Smeden, M. (2020). "Prediction models for diagnosis and prognosis of COVID-19: Systematic review and critical appraisal." *BMJ*, 369, m1328.

Xenidis, R. y Senden, L. (2020), "EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination" Bernitz et al. (eds), *General Principles of EU law and the EU Digital Order*, Kluwer Law International, 67 y ss.

Zarsky, T.Z. (2016) "The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making" Science, Technology, & Human Values, vol. 41, no. 1, 118-132.

Zhou, X.H., Obuchowski, N. A., y McClish, D. K. (2011). *Statistical Methods in Diagnostic Medicine. Wiley-Interscience*, 2011.

Žliobait y Custers, (2016) "Using Sensitive Personal Data May be Necessary for Avoiding Discrimination in Data-driven Decision Models", *AI and Law*, 24, 183.

Zuiderveen Borgesius, F. J. (2020), "Strengthening legal protection against discrimination by algorithms and artificial intelligence". *International Journal of Human Rights*, vol. 24, no. 10, 1572-1593.

La inteligencia artificial (IA) está transformando de manera radical el ámbito sanitario, generando oportunidades sin precedentes para mejorar la atención médica, desde el diagnóstico precoz hasta la personalización de tratamientos. Sin embargo, estos avances conllevan riesgos éticos y sociales que exigen un análisis profundo y una regulación adecuada. Esta obra examina la compleja intersección entre tecnología, derecho y medicina, con una atención particular en los sesgos algorítmicos y sus implicaciones para la equidad en el acceso y la calidad de la salud.

A lo largo de sus capítulos, el libro analiza los marcos normativos que regulan el uso de la IA en sanidad. El análisis se centra en identificar fortalezas y limitaciones de los principales instrumentos normativos europeos, como el Reglamento General de Protección de Datos (RGPD), el Reglamento de Inteligencia Artificial (RIA) y el Reglamento de Productos Sanitarios (RPS).

Los temas tratados abarcan desde la problemática de los sesgos algorítmicos y la ética computacional, pasando por la regulación específica del software sanitario, hasta la responsabilidad derivada de los daños morales y materiales ocasionados por la discriminación algorítmica. Con un enfoque crítico y propositivo, se abordan cuestiones clave como la transparencia, la explicabilidad, la supervisión humana activa y la rendición de cuentas en casos de decisiones automatizadas. Además, se exploran las implicaciones éticas y prácticas de la discriminación algorítmica, con ejemplos concretos y estrategias destinadas a mitigar sus efectos.

La obra invita al lector a reflexionar sobre el equilibrio necesario entre la innovación tecnológica y la protección de los derechos fundamentales, proponiendo un marco regulatorio inclusivo y sostenible que garantice un uso responsable y equitativo de la inteligencia artificial en el ámbito sanitario. Con esta perspectiva, aspiramos contribuir al desarrollo de una inteligencia artificial que no solo sea eficaz, sino también ética, justa y centrada en las personas, consolidando así su papel como una herramienta inclusiva en un sector tan esencial como la salud.

tirant
lo blanch

978-84-1095-863-0
9 788410 958630