

La inclusión de la diversidad en los sistemas de inteligencia artificial de soporte a decisiones sanitarias como manifestación del enfoque de derechos humanos

The Inclusion of Diversity in Artificial Intelligence Systems of Medical Decision Support as Human Rights Approach in Artificial Intelligence

EDISON CALAHORRANO LATORRE¹ 

RESUMEN

En este trabajo se analiza la relevancia de la inclusión de la diversidad como principio y la manera en que puede aportar en prevenir sesgos en las decisiones sanitarias apoyadas por sistemas basados en inteligencia artificial. En primer lugar, se propone abordar el contenido del principio de inclusión a la diversidad en las distintas fases de diseño e implementación de los sistemas; posteriormente, se propone que la transversalidad del principio permite proteger a los grupos más vulnerables frente a decisiones apoyadas por sistemas de inteligencia artificial en el ámbito sanitario. Para esto se recurrirá a la revisión de literatura especializada y la aplicación del método dogmático. Se plantea la hipótesis de que la inclusión de la diversidad de forma transversal al procesamiento de los datos y diseño del algoritmo para soporte en la toma de decisiones sanitarias permite mitigar riesgos de discriminación algorítmica, evitando la subrepresentación de grupos vulnerables.

Palabras clave: inteligencia artificial, equidad algorítmica, salud digital, discriminación algorítmica, decisión sanitaria.

ABSTRACT

This paper analyzes the relevance of diversity inclusion as a principle and how it can contribute to preventing bias in health care decisions supported by artificial intelligence-based systems. First, it is proposed to address the content of the principle of diversity inclusion in the different phases of the design and implementation of the systems; then it is proposed how the transversality of the principle allows to protect the most vulnerable groups against decisions

¹ Profesor asistente, Universidad de Tarapacá, Iquique, Chile. Correo electrónico: ecalahorrano@academicos.uta.cl



supported by artificial intelligence systems in the health field. For this purpose, a review of specialized literature and the application of the dogmatic method will be used. It is hypothesized that the inclusion of diversity in a transversal way to data processing and algorithm design to support health decision making allows mitigating the risks of algorithmic discrimination, avoiding the underrepresentation of vulnerable groups.

Keywords: artificial intelligence, algorithmic equity, inclusion of diversity, algorithmic discrimination, clinical decision.

1. Introducción

El crecimiento de las aplicaciones de inteligencia artificial ha tenido un especial desarrollo en el ámbito sanitario, específicamente respecto del soporte para la toma de decisiones sanitarias fundamentadas en el procesamiento de datos médicos (Ramón, 2024, pp. 18-20).

El acercamiento paulatino hacia la medicina personalizada y de precisión (Hegde y Shenoy, 2021, pp. 146-148) ha derivado en que procesos como la gestión de imágenes mediante *machine learning* y la implementación de sistemas predictivos y modeladores puedan ser utilizados en disciplinas como la oncología o dermatología (Cobas y Pérez, 2023, p. 129).

De manera simultánea a la multiplicación de estas aplicaciones prácticas de los sistemas de inteligencia artificial surgen también los efectos de la aparición de sesgos que tienen el potencial de perpetuar inequidades. En este sentido, el reconocimiento de la diversidad en la recolección y depuración de los datos, como en el diseño e implementación de los algoritmos es prioritario, por lo que las auditorías de sesgos se consolidan como instrumentos para evitar posibles efectos discriminatorios. Al respecto se ha señalado:

Las auditorías de sesgos son desempeñadas por actores externos e independientes, para verificar de manera exhaustiva cada parte del sistema con medidas cualitativas y técnicas, en cambio, si se requiere una auditoría de prejuicios, se debe aplicar una prueba con un objetivo limitado de un aspecto particular del sistema, por ejemplo, revisar si produce prejuicios raciales en los resultados de una decisión, la cual radica en comparar y observar los resultados que se obtienen y no en evaluar el código fuente del sistema.

No obstante, estas auditorías no pueden proveer una imagen holística del sistema, porque puede que sí demuestren sesgos en el sistema, pero no aseguran que el sistema esté libre de otras formas de problemas, discriminación o impactos en la sociedad (Llamas, Mendoza y Graff, 2022, p. 44)

A pesar de lo señalado, los sistemas de auditoría y monitoreo son aún insuficientes en la medida cuando se ubican en un espacio específico del desarrollo dedicado al control, mas no de forma transversal al diseño e implementación. Al final del día, los sistemas dependen de la manera en que se los alimente con datos y la calidad de los mismos; así como, la forma en que sean entrenados.

La necesidad de alcanzar una aproximación razonable hacia la equidad requiere intensificar los esfuerzos en la identificación de los sesgos, para, de esta manera, evitar su multiplicación con los efectos injustos y discriminatorios que ello conlleva (Calahorrano, 2025, pp. 4-5; Calahorrano 2025b, pp. 409-410; Walker, 2023, pp. 150; Bedecarratz y Aravena, 2023, pp. 209). Se ha señalado como elemento clave en la implementación de los sistemas el asegurarse que se encuentren “siempre al servicio del conjunto de la especie humana y bajo el control ético y jurídico de los ciudadanos, con la intervención de las

instituciones públicas orientadas al bien común. Así, buscar la creación de un marco normativo común para productos dotados de inteligencia artificial y exigir que toda máquina, robot y objeto que posea esta tecnología permanezca siempre bajo el control humano” (Juri, 2021, pp. 9-10).

En atención a lo señalado, en este trabajo se analiza la relevancia de la inclusión de la diversidad como principio, y la manera en que puede aportar en la prevención de sesgos en las decisiones sanitarias apoyadas por sistemas basados en inteligencia artificial. En efecto, el problema jurídico planteado consiste en prevenir la perpetuación de inequidades provocadas por la insuficiente consideración de determinantes sociales en el desarrollo de los algoritmos y la subrepresentación de grupos de mayor vulnerabilidad en la recolección y depuración de datos; así como en el diseño, implementación y ejecución de los algoritmos (Calahorrano, 2025, pp. 11-12; Fosch-Villaronga *et. al.*, 2022, p. 2; Ramón, 2024, pp. 33)

Por otro lado, desde la implementación del enfoque de derechos humanos en el tratamiento de la inteligencia artificial; la consideración de la vulnerabilidad debe estar presente permanentemente (Prictor, 2022, pp. 22-24), por lo que resulta importante discutir las soluciones regulatorias desde una perspectiva holística y participativa que busque activamente proscribir la reproducción de la inequidad (Coddou, Smart y Vivallo 2024, pp. 399-401).

Desde la perspectiva de la responsabilidad, quien programa el sistema debe hacerse cargo de la calidad de los datos y de la mitigación de la aparición de sesgos en una suerte de corresponsabilidad con el tratante (Calahorrano, 2025, pp. 10-11; Calahorrano, 2025b, pp. 406-409; Vedicchio y Perin, 2022, pp. 23-24). Las discusiones sobre el régimen de responsabilidad extracontractual derivado de daños provocados por sistemas de inteligencia artificial aún se encuentran inacabadas.

Por lo tanto, se plantea la hipótesis de que la inclusión de la diversidad como principio transversal al diseño e implementación del algoritmo a utilizarse como sistema de soporte para las decisiones sanitarias mitiga la aparición y reproducción de sesgos que conlleven la subrepresentación de la población más vulnerable y que puedan derivar en discriminación algorítmica. Los mecanismos planteados, conjuntamente con el abordaje de la regulación e incorporación de control humano en la alimentación de datos y entrenamiento del algoritmo fundamentada en el enfoque de derechos humanos y prevención activa de la discriminación en el desarrollo de sistemas de inteligencia artificial en el ámbito sanitario, son soluciones viables ante la opacidad y poca fiabilidad de los sistemas más complejos.

Para comprobar la hipótesis, se partirá por analizar el contenido del principio de inclusión de la diversidad en las distintas fases de diseño e implementación de los sistemas; posteriormente, se propone demostrar cómo la transversalidad del principio permite proteger a los grupos más vulnerables y se condice con la implementación de un enfoque de derechos humanos en el tratamiento de la inteligencia artificial que sirve como soporte a la toma de decisiones sanitarias.

Para lo señalado se recurrirá a la revisión de literatura especializada y aplicación del método dogmático y al análisis de casuística aplicable, recopilándose documentos de los últimos cinco años bajo las palabras clave “inclusión de la diversidad en la inteligencia artificial”, “discriminación algorítmica”, “inteligencia artificial en salud”, “salud digital” y “*artificial intelligence in healthcare*” en las bases de datos PUBMED, SCOPUS, WOS y Scielo Chile. Se aplicará además el método analítico para desagregar los trabajos que planteen la inclusión de la diversidad como mecanismo para solución de la discriminación algorítmica.

2. El contenido del principio de inclusión a la diversidad en los sistemas de inteligencia artificial que brindan soporte a la decisión sanitaria

La inteligencia artificial se traduce, esencialmente, en un software que busca cumplir con determinados procedimientos como si fuesen realizados por seres humanos, los cuales son desarrollados en base al manejo de macrodatos que son gestionados y procesados buscando la mayor eficiencia (Fosch-Villaronga *et.al*, 2022, pp. 1-2; Ramón, 2021, p. 331).

Los sistemas de inteligencia artificial aplicados al soporte de las decisiones en el ámbito sanitario tienen el potencial de proporcionar mayor precisión y accesibilidad, siempre y cuando se establezcan marcos regulatorios con estándares adecuados de seguridad del paciente dentro de un contexto de monitoreo interdisciplinario permanente (Ramón, 2024, p. 33).

La variedad de aplicaciones de la inteligencia artificial en salud permite afirmar que es inconcebible comprender el desarrollo actual de este sector sin su presencia. Desde la detección y tratamiento temprano de tumores hasta permitir un manejo preciso y descifrar las instrucciones para descubrir la influencia de las variaciones genéticas con distintas patologías (Molnár-Gábor y Giesecke, 2022: pp. 379-380), ejercicio de la medicina predictiva y de precisión.

La inteligencia artificial es una tecnología disruptiva protagonista en el advenimiento de una cuarta revolución industrial, lo cual se ha demostrado por su rápida inserción en aplicaciones de la vida cotidiana, en la manera en que interactuamos, producimos o ejercemos el cuidado a través de ella y en cómo ha tenido una implementación transversal en el ámbito sanitario (Calahorrano, 2025 b, pp. 400-401; Martín Casals, 2022, pp. 101-102).

En el centro de los avances se encuentra el lograr diagnósticos precisos y oportunos que aprovechen la capacidad de análisis de gran cantidad de datos por parte de los sistemas, con el fin de ofrecer una asistencia adecuada a los tratantes en la toma de decisiones. La capacidad señalada debe estar caracterizada por la transparencia y las posibilidades de explicación. Estas características permitirían al tratante interpretar adecuadamente los resultados (Van Baalen *et.al.*, 2020, pp. 526-527).

En este contexto, los sistemas de soporte en decisiones sanitarias son mecanismos de conocimiento activo que usan datos de pacientes para generar brindar soporte a la decisión clínica en el caso concreto, lo que sirve como insumo para el diagnóstico, selección de tratamiento o alternativa terapéutica, monitoreo, flujos de trabajo, almacenamiento y procesamiento de la información clínica de paciente (Calahorrano, 2025, pp. 4-5; Jones, Thornton y Wyatt, 2023, pp. 2.)

El soporte en la toma de decisión clínica ha sido uno de los ámbitos en el que esta tecnología ha encontrado mayores dificultades para alcanzar la confianza y fiabilidad de los usuarios (Jones, Thornton y Wyatt, 2021, pp. 2-3). Se ha señalado que los sistemas de inteligencia artificial pueden ser explicables, semi explicables o estar inmersos en una estructura de caja negra, en las que el costo de la mayor precisión es el menor control y, por ende, conocimiento y posibilidades de explicación respecto del proceso y los resultados; a esta peculiar estructura se le ha llamado *black box medicine*, cuya característica es el grado de inexplicabilidad para el paciente, el tratante o hasta el propio desarrollador (Raposo, 2025, pp. 3-5).

Uno de los desarrollos de la medicina más importantes de los últimos dos siglos consiste en la transición desde una relación clínica paternalista y unidireccional que aborda al paciente como un sujeto pasivo receptor del tratamiento hacia una basada en el consentimiento, bidireccional, y que presupone un paciente informado y plenamente claro de su calidad de sujeto de derechos (Solaiman y Malik, 2025, p. 22).

En este contexto se produce la irrupción de *machine learning* que consiste en la habilidad de las máquinas para aprender de los datos y construir algoritmos, proceso que al automatizarse es capaz de crear modelos que permiten construir patrones y brindar soporte a la toma de decisiones mejorándola y haciéndola más precisa. Por otro lado, la *deep learning* aplica redes neuronales de varias capas, por lo que la opacidad respecto a la manera en que llega a sus conclusiones puede ser mayor (Johnson, 2019, pp. 429-430).

En un distinto sentido se diferencian sistemas que pueden ser supervisados, cuando se cuenta con la presencia de un controlador humano que introduce los datos y monitorea el análisis; o no supervisados, en los casos en que son capaces de generar sus propias conclusiones.

Finalmente, pueden existir sistemas de aprendizaje reforzado en el que el agente humano introduce incentivos positivos y negativos como recompensas y sanciones que permiten corregir en su entrenamiento aquellos resultados no deseados (Kumar, Chauhan y Awasti, 2023, pp. 5-6).

La preocupación central en lo señalado es que la aplicación de sistemas de aprendizaje automatizado en salud, en su potencial predictivo y personalizado, corre el riesgo de producir daños recurrentes debido a los sesgos que repercuten en reproducción de inequidades (Agarwal *et. al*, 2023, p. 2.). Los sistemas de inteligencia artificial (IA) en el ámbito sanitario han demostrado el potencial para aprender del procesamiento de una gran cantidad de datos, identificar patrones y llevar a cabo conclusiones precisas sin mayores indicaciones iniciales; los sistemas a los que nos referimos pueden solucionar la toma de decisiones faltas de información que están intrínsecas al conocimiento limitado del tratante no asistido por los mismos; sin embargo, mayor acceso a los datos no significa, necesariamente, mayor explicabilidad, por cuanto esta última se enlaza con la idea de la aptitud para proporcionar explicaciones y puede ser tratada de manera autónoma frente al principio de transparencia (Raposo, 2025, pp. 4-5).

Sin embargo, la explicación sobre las conclusiones del sistema automatizado, especialmente en el ámbito sanitario, requiere, por una parte, obedecer a un determinado enfoque que maximiza la garantía de los derechos de los pacientes y la protección de aquellos que se encuentren en condiciones más susceptibles a la vulnerabilidad y, por otro lado, centrar el debate en la garantía de derechos permita establecer mínimos inclusivos a distintas perspectivas éticas; es así que, se ha señalado:

En efecto, y a diferencia de la ética, que no delimita formalmente su campo de acción, los catálogos de derechos humanos contienen derechos de carácter independiente, con distintos ámbitos de competencia, cuestión que nos permite nombrar o nominar los daños que generan los diversos usos de la IA. Así, por ejemplo, podemos distinguir infracciones que la IA puede generar en el derecho al debido proceso, en el ejercicio igualitario de derechos sociales o, bien, el efecto inhibitorio que las tecnologías digitales pueden generar en el ejercicio de la libertad de expresión. El sistema universal de derechos humanos, tanto en su dimensión convencional (procedimientos de monitoreo y supervisión de cada tratado internacional) como extra-convencional (procedimientos de monitoreo y supervisión que no dependen de un tratado específico), está construido sobre la base de que los derechos humanos son distinguibles conceptualmente, de modo que permiten, a personas que adhieren a diferentes concepciones éticas, consensuar casos o situaciones claras de violaciones a los derechos humanos (Coddou, Smart y Vivallo, 2024, p. 360).

Al respecto se ha profundizado en la literatura nacional el estudio crítico de los datos como un esquema epistemológico para evitar una exacerbada deferencia a lo que pueden aportar en la toma de decisiones. Esto es necesario en ámbitos en que los derechos fundamentales se encuentren involucrados, por lo que la vocación masiva de aplicación de un algoritmo o de una red neuronal en salud requiere no solo de la consistencia y

depuración de los mismos; sino además, se debe verificar el contexto político estatal o institucional en el que se implementan y los *stakeholders* involucrados (Coddou y Smart, 2021, p. 307-310).

En la misma línea de lo señalado resulta interesante la aproximación de (Pfohl, Foryciarz y Shah, 2021, p. 2) quienes señalan que el principal problema para la implementación de la medicina predictiva a través de la consideración de la equidad en los algoritmos, lo constituyen las inequidades estructurales que se presentan en las determinantes sociales de la salud. Esto produce una subrepresentación de determinados grupos vulnerables.

La inclusión a la diversidad surge como principio que incorpora los valores, el contexto y las preferencias de los pacientes en la construcción del algoritmo; y, que se suma al principio de equidad algorítmica centrado en la detección y erradicación de sesgos para proporcionar explicabilidad a las decisiones automatizadas del sistema (Calahorrano, 2025, p. 13). La inclusión de la diversidad en la transversalidad de la construcción del algoritmo permite humanizar la decisión automatizada, contrastando la optimización de la precisión con la solución adecuada al contexto basada en activas correcciones de las subrepresentaciones, en todo caso, se configura como instrumento activo de proscripción de la discriminación, en la medida en que los sistemas concentran datos mayormente seleccionados y procesados desde los actores económicos y geográficos con la capacidad de configurar el algoritmo y equipos de desarrolladores con poca diversidad en su configuración (Coddou, Smart y Vionelli, 2024, pp. 367-368)

Los sistemas de inteligencia artificial utilizados en la toma de decisiones sanitarias automatizadas encajan en aquellos de alto riesgo, según el artículo 6 del Reglamento Europeo que regula la materia, publicada el 24 de julio de 2024. Esta consideración conlleva ciertas previsiones obligatorias en su implementación, entre ellas, la incorporación de estándares de transparencia y la explicabilidad como la aptitud para la comprensión de los motivos que llevan a las conclusiones y la posibilidad de que una persona que se sienta afectada por las mismas pueda acceder de manera clara y precisa cómo el sistema contribuyó a la decisión final (Raposo, 2025, pp. 6-7).

La posibilidad de procesar y generar conclusiones autónomas a partir de una gran cantidad de datos, entre ellos, valores y preferencias de los pacientes, permite considerar a los sistemas de inteligencia artificial como una herramienta sociotécnica que proporciona una mayor claridad sobre el uso, evaluación y mejora continua en la asistencia de decisiones clínicas éticamente complejas (Ferrario, *et al.*, 2023, p. 174). La incorporación de estas variables ha sido referida como incorporación de la diversidad en el diseño e implementación de los algoritmos (Fosch-Villaronga *et al.*, 2022, pp. 9-10), lo que, a su vez, debe ser incorporado en una categoría más amplia como es la equidad algorítmica.

La interacción tratante-paciente corre el riesgo de caer en una falta de empoderamiento de sus actores para recaer en otros nuevos que protagonizan el desarrollo del sistema, la certificación para su desarrollo y su funcionamiento, aquellos que monitorean y realizan el control externo; así como quienes ofrecen garantías de la calidad de los datos, todos los cuales se encuentran más bien en un contexto mercantilizado en el que los proveedores de los sistemas pueden modificar de manera sustancial el carácter deliberativo que debe comprender la relación clínica (Solaiman y Malik, 2025, pp. 22)

Para verificar si es posible que el modelo de relación clínica intermediado por sistemas de inteligencia artificial que dan soporte a las decisiones sanitarias cumpla con los criterios de legitimidad y no discriminación, es necesario identificar en qué consiste dicho proceso y en qué momentos debe implementarse el control activo de erradicación de sesgos, la verificación de la inclusión de la diversidad en los datos incorporados

y la prevención de la subrepresentación de grupos vulnerables en armonía con la aplicación de un enfoque de derechos humanos transversal al desarrollo y puesta en funcionamiento del algoritmo.

El proceso que sigue un modelo de inteligencia artificial inicia con la recolección de datos de diferentes fuentes como registros electrónicos que contienen información sanitaria, cuya calidad se enlaza directamente con la eficacia en la conservación y control; así como con la interoperabilidad. Las fichas clínicas electrónicas son solo uno de los elementos a tomar en cuenta, multiplicándose las posibilidades de orígenes de los datos, como se da en el caso de los dispositivos portables o alojamiento en repositorios bajo los cuales se almacena la información de las consultas de telemedicina. Lo señalado propone desafíos importantes para anonimizar y proteger los datos sensibles.

Sobre esto último destaca la iniciativa de creación del Espacio Europeo de Datos de Salud implementado a través del Reglamento correspondiente, publicado el 5 de marzo de 2025, en el que se establece como el esfuerzo más importante en la materia, especialmente, respecto del llamado uso secundario de los datos de salud, preocupación fundamental que debe ser resuelta previo a la implementación de sistemas de inteligencia artificial de soporte a las decisiones sanitarias (Recuero, 2024, pp. 547-549).

Continuando con la esquematización del proceso de diseño del algoritmo, corresponde la limpieza y el reprocesamiento de los datos para que se encuentren aptos para su uso en el sistema. Aquí deben implementarse eficaces mecanismos de control y transparencia, en las que contribuye una revisión humana pormenorizada con enfoque de derechos e inclusión de la diversidad (Calahorrano, 2025, p. 7) para la erradicación de sesgos que puedan generar resultados discriminatorios que puede intensificarse al momento de incorporar los datos sesgados al diseño del algorítmica. Este problema ha sido denominado *garbage in, garbage out* y se traduce en que el entrenamiento de sesgos solo puede reportar conclusiones de la misma naturaleza, al igual que la introducción u omisión de determinadas variables por el desarrollador (Raposo, 2025, pp. 10-11).

Al paso anterior se incorpora la ingeniería del mecanismo que permite construir variables y *data points* para aplicar como *inputs* del modelo a ser entrenado, momento en el que se establece el algoritmo que permite al sistema aprender patrones y al control humano realizar ajustes para minimizar la aparición de sesgos. Entrenado el algoritmo debe evaluarse frente a modelos alternativos, fase en la que adquiere importancia la implementación de *sandbox* regulatorios.

Si el algoritmo está destinado al soporte de la decisión en el diagnóstico existe la necesidad de verificar que los datos sean confiables y estratificar las evidencias para lograr una mayor fiabilidad. Una estratificación de los grados de certeza de las fuentes de información puede verificarse desde el más bajo, como son las opiniones de los expertos, posteriormente, las series de casos, estudios con grupo de control hasta los meta análisis. La dificultad en el desarrollo de fuentes confiables de los datos ha impedido la proliferación de sistemas fiables de soporte del diagnóstico en los ámbitos más complejos (Kumar, Chauhan y Awasti, 2023, pp. 7-8).

A partir de la comprensión del funcionamiento de los sistemas y las limitaciones que se han señalado respecto a sus posibilidades para abordar la diversidad, contexto, determinantes y especificidades que requiere cada caso, se puede profundizar respecto de la discriminación algorítmica y los desafíos de proponer soluciones jurídicas respecto a sus efectos.

3. La inclusión de la diversidad como herramienta frente a la discriminación algorítmica en decisiones sanitarias. Una aplicación del enfoque de derechos humanos y consecuencias respecto a la responsabilidad por daños

Una primera aproximación hacia la construcción de algoritmos desde el enfoque de derechos consiste en la identificación exhaustiva de aquellos grupos subrepresentados por distintos factores como el idioma, la raza, el género o cualquiera de aquellos criterios sospechosos de discriminación. Una mirada específica a la consideración de dichos grupos en la obtención y depuración de los datos; así como, en la construcción y diseño del algoritmo permiten romper con la primera barrera consistente en una inadecuada generalización de los casos considerados para la toma de decisiones.

La privacidad del paciente y la protección de datos son también presupuestos para la fiabilidad y confianza de los sistemas, en la medida en que la fase de desarrollo de algoritmo es la que puede originar discriminación, incluso diferenciaciones respecto de la calidad de cuidado generando predicciones injustas a partir de sesgos étnicos, de género o basados en la cobertura de salud. Es frecuente también la discriminación directamente respecto de población vulnerable como las personas con discapacidad, que suelen estar infrarrepresentadas en los datos de entrenamiento (Lillywhite y Wolbring, 2021, pp. 129-135).

La representación excesiva de los grupos culturalmente dominantes y el deficiente análisis del contexto puede afectar la calidad del soporte que brinda el sistema; a lo señalado debe sumarse la necesidad de un punto de control de revisión crítica de los resultados incluyendo el criterio de equidad algorítmica (*algorithmic fairness*), entendida como aquel que guía el diseño de modelos con el objetivo de prevenir la discriminación que afecta a grupos de protección reforzada (Huang *et. al.*, 2022, p. 3; Grote y Keeling, 2022, pp. 9-10; Barda *et. al.*, 2021, p. 551; Grote y Berens, 2019: 9-10). Aplicaciones directas de equidad algorítmica han sido tratadas en el contexto de selección de personal (Coddou y Padilla, 2024, pp. 6-7), acceso a la justicia por parte de los consumidores a través de sistemas de *online dispute resolution* (Barrientos, Bozzo y Jequier, 2023, pp. 15-28). Casos que ejemplifican los efectos nocivos que puede tener la falta de implementación de controles específicos y activos son el del sistema SyRI (Coddou y Smart, 2021, p. 305; Lazcoz y Castillo, 2020, pp. 215-222) respecto de los beneficios para el cuidado de niños y la generación de perfiles de riesgo de fraude para la seguridad social.

La aplicación de la inteligencia artificial para el diagnóstico clínico tiene, sin embargo, importantes desafíos respecto a la generación de confianza y fiabilidad (Jones, Thornton y Wyatt, 2023, p. 19; Raposo, 2025, pp. 11-12). Las reflexiones sobre la humanización de la relación clínica y la aplicación de la medicina centrada en el paciente toman especial relevancia respecto del ejercicio de su autonomía (Calahorrano, 2025, pp. 9-10; Karimian, Petelos y Evers, 2022, pp. 539-551; Raposo, 2025, p.12).

En este punto se han revisado los mecanismos por los cuales un sistema de inteligencia artificial que brinda soporte a la decisión sanitaria puede ingresar al mercado, señalándose como punto de partida la certificación de conformidad que se aplica, por ejemplo, en el contexto europeo. A pesar de que se sostiene la necesidad de independencia de aquellos certificadores autorizados, adquiere importancia la manera en que estos procesos determinan un control fiable y no solamente directrices autoimpuestas por los propios proveedores de los sistemas como parte de las dinámicas de mercado; por lo señalado, el desafío consiste en que el *compliance* respecto de la previsión de riesgos y eliminación temprana de sesgos en los datos; así como en la construcción y diseño del algoritmo no se convierte en una mera *checklist* en que los intereses de los proveedores sean reproducidos en los controladores (Solaiman y Malik, 2025, pp. 12-13).

Una adecuada prevención de riesgos sobre posibles afectaciones a derechos fundamentales de los pacientes o sobre la aparición de sesgos discriminatorios debe continuar una vez implementado el algoritmo y solo puede sostenerse a través de un sistema abierto de reportes de posibles incidencias, en las que el esquema sobrepase la información generada únicamente por los proveedores del sistema o los usuarios autorizados; se hace necesaria la intervención de los directos afectados, razón por la cual el Reglamento Europeo que regula estos sistemas señala que la interacción del proveedor con los *stakeholders*, la explicabilidad y la posibilidad de que el algoritmo pueda ser auditado por una persona natural son esenciales (Smuha *et.al*, 2021, pp. 44-47)

Sin un adecuado *compliance* se corre el riesgo de que el sistema no sea capaz de entregar soluciones que consideren valores, preferencias y necesidades de paciente, incluyendo la verificación de determinantes sociales y variables demográficas, económicas y sociales (Liyanage *et. al.*, 2019, pp. 41-46).

Se ha señalado que para los sistemas de soporte a las decisiones sanitarias resulta fundamental verificar quiénes tienen la facultad de evaluar cuándo existe un riesgo y cuáles son los criterios de determinarlo como tal, lo que en la práctica se traduce en quién ejerce el control respecto del análisis de los reportes de incidentes y quién proporciona el entrenamiento adecuado al encargado del control humano; esto es determinante para la configuración de la relación clínica, en la medida en que proveedores y controladores pueden acaparar la detección de riesgos y auditoría de equidad algorítmica, dejando al tratante y a los pacientes como agentes poco empoderados, a pesar de ser el núcleo en el que se desenvuelve la garantía de los derechos fundamentales de los segundos y el respeto a sus valores y preferencias, desde la perspectiva de una relación clínica deliberativa (Solaiman y Malik, 2025, pp. 13-17).

Ante lo señalado se resalta, finalmente, la necesidad de incorporar un efecto irradiador del enfoque de derechos humanos al abordar los mecanismos de implementación de equidad algorítmica desde la depuración de los datos y construcción del algoritmo, con la transversalidad del principio de inclusión de la diversidad; y, procurar que los mismos permanezcan en los controles *ex post* que lleven a cargo las herramientas de mitigación de riesgos y el agente humano de control. El principio de inclusión de la diversidad coincide plenamente con el enfoque de derechos humanos al priorizar la protección de los grupos con mayor vulnerabilidad y detectar de manera activa la subrepresentación de los mismos, lo que debe estar presente en el tratamiento de datos, diseño y construcción del algoritmo, implementación de *sandbox* regulatorios, implementación de planes de cumplimiento y mecanismos de *compliance*, entrenamiento del agente de control humano y procesamiento de reportes de alertas de riesgo. En el caso de los sistemas que brindan soporte a las decisiones sanitarias, estas medidas deben ser respetuosas del carácter deliberativo de la relación clínica.

Finalmente, respecto a los sistemas señalados se plantean dos escenarios con distinto potencial de afectación a derechos fundamentales del paciente; en primer lugar, cuando el resultado proporcionado por el sistema es considerado un insumo más para la decisión del equipo médico; y, la segunda, cuando el sistema es considerado como un miembro más de equipo y sus conclusiones son ponderadas en iguales condiciones que los criterios de los tratantes (Calahorrano, 2025, p. 10; Perin, 2019, pp. 11-21.)

En este contexto tiene vital importancia verificar los efectos de la deferencia que puede existir por parte del tratante respecto de las conclusiones emitidas por el sistema de inteligencia artificial que da soporte a la decisión sanitaria.

La regla se configuraría de la siguiente manera: a mayor grado de fiabilidad y transparencia del algoritmo mayor sería la deferencia del tratante hacia el resultado obtenido por su aplicación; sin embargo, mientras

menor sea la confianza en el sistema, es posible que el tratante pueda exponer una adecuada justificación para no tomar en cuenta la conclusión del sistema. El objetivo de este razonamiento es evitar la configuración del sesgo de confirmación por el cual el tratante adapta sus propias consideraciones a las conclusiones proporcionadas por el sistema o lo utiliza como mecanismo de medicina defensiva (Calahorrano, 2025, p. 10; Grote y Berens, 2020, p. 208).

Lo señalado ha llevado a afirmar que los sistemas no explicables y faltos de la adecuada transparencia deben ser prohibidos para el uso en cuidado de la salud, por cuanto, la falta de explicación de funcionamiento del algoritmo conduce a una afectación directa a la comunicación en la relación clínica (Čartolovni *et.al*, 2022, pp. 5-7.).

A pesar de esta afirmación, hay quienes argumentan que un sistema de inteligencia artificial de caja negra que brinda soporte o, inclusive, se suma como un criterio adicional al del tratante, no necesariamente debe ser excluido, en la medida en que la relación clínica está basada, fundamentalmente, en la confianza; señalándose también que la falta de explicabilidad ya es una característica general de la relación clínica, por lo que la explicación detallada al paciente no siempre es buscada ni requerida al tratante, como en el caso del funcionamiento de los medicamentos, en el que su fiabilidad está basada en estudios clínicos con márgenes de aceptabilidad y presencia permanente de efectos secundarios; adicionalmente, la aleatoriedad intrínseca de la medicina, el funcionamiento del cuerpo humano y el cerebro conllevan la necesidad de buscar siempre explicaciones razonables mas no agotadas en dichos ámbitos (Raposo, 2025, pp. 13-18).

Adicionalmente, más allá de que las legislaciones hayan incorporado en el deber de información auto determinativa el de proporcionar información respecto del uso de algún mecanismo tecnológico en el tratamiento, como es el caso de la reforma introducida por la Ley 21.541 en el artículo 14 de la Ley 20.584 en Chile, la sobreinformación respecto al funcionamiento de los mecanismos de salud digital también pueden vulnerar el derecho del paciente al consentimiento informado, por lo que la proporcionalidad en la entrega de información debe tenerse en cuenta.

Bajo las previsiones anteriores, es necesario señalar que en la confianza que pueda caracterizar a un sistema de soporte a la decisión sanitaria intervienen factores individuales, características del sistema y factores contextuales que se conectan entre sí. Los primeros factores pueden ser las características demográficas, condiciones de salud o trato individual hacia el paciente; mientras que los segundos se refieren a la transparencia. Menos estudiadas son las características contextuales como la cultura del cuidado de la salud, relaciones interpersonales y la gobernanza (Steerling *et al.*, 2023, pp. 6-7), que como se ha señalado, deben ser transversales para lograr una razonable explicación y garantía de los derechos fundamentales del paciente.

En conclusión, retomar los componentes deliberativos de la relación clínica que permiten la toma de decisiones sanitarias centradas en el paciente, el fortalecimiento de la comunicación y el diálogo sobre la implementación de los sistemas de soporte basados en inteligencia artificial ayuda a cimentar la confianza (Steerling *et al.*, 2023, p. 6)

El Reglamento de Parlamento Europeo 2024/1689, que establece normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión (AI Act), publicado el 13 de junio de 2024 en el Diario Oficial de la Unión Europea, incorpora un enfoque regulatorio basado en el riesgo y la magnitud de los daños que pueden provocarse, combinado con el nivel de autonomía que puede llegar a desarrollar el sistema respecto de la toma de decisiones prevé la incorporación de un seguro obligatorio complementado por un régimen de responsabilidad que suaviza la prueba de la culpa y la causalidad (Madrid, 2023, p. 97; Martín Casals, 2022, pp. 9-10).

Sin duda, para la prevención de la responsabilidad de los sistemas de soporte a la decisión sanitaria, como se ha señalado en el apartado anterior, la gobernanza de datos es esencial, por lo que los numerales 3 y 4 de artículo 10 de la AI Act se refieren a la obligatoriedad de que los datos de entrenamiento, validación y prueba sean pertinentes y representativos, carezcan de errores y estén completos, además de tener en cuenta el contexto geográfico, conductual o funcional específico que se pretende usar en el sistema (Calahorrano, 2025, pp. 11-12).

Finalmente cabe señalar que la norma en análisis además incluye obligaciones específicas a los fabricantes, distribuidores, importadores que se manifiestan, especialmente en los deberes de información (Añón, 2022, pp. 32-49; Navas, 2021, pp. 43-50; Navas, 2022, pp. 18-23). A lo anterior se suma la reciente publicación de la Directiva 2024/2853 de 23 de octubre del año pasado que deberá ser implementada por los estados miembros hasta el 9 de diciembre de 2026; cuyas principales motivaciones entre varios motivos, por la insuficiencia de la primera frente al advenimiento de los productos digitales de mayor complejidad y sistemas de inteligencia artificial, además de las consecuencias de la economía circular y estructuras cada vez más intrincadas para encontrar una respuesta al daño en el proceso de toma de decisiones; la autonomía de los sistemas de inteligencia artificial, la dificultad en la determinación de ámbito de control de fabricante respecto de la toma de decisiones, la capacidad de aprendizaje que estos presentan, la opacidad y la interconectividad con componentes que se manifiestan como un desafío que obligó a la actualización de régimen de responsabilidad por productos defectuosos. El autor resalta, especialmente, las dificultades que este entorno complejo pueden provocar en el demandante respecto de la prueba de nexo causal y la determinación de operador económico responsable frente al daño (Martín Casals, 2024, pp. 37-38).

Respecto a la responsabilidad civil extracontractual generada por daños provocados por sistemas de inteligencia artificial, que se encuentren fuera de los presupuestos del estatuto de responsabilidad por productos defectuosos, se estaba discutiendo 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 (Directiva sobre responsabilidad en materia de IA) que preveía un sistema de responsabilidad por culpa con sistemas de flexibilidad probatoria de la culpa y presunción de la causalidad a partir de la misma, sumado a deberes especiales de aportación de información para los desarrolladores (Martín Casals, 2024, pp. 71-74); sin embargo, este proyecto se encuentra actualmente suspendido en su discusión, a la espera de los primeros resultados de la entrada en vigencia del Reglamento Europeo sobre Inteligencia Artificial y la renovada Directiva sobre el régimen de productos defectuosos.²

4. Conclusiones

Los sistemas de inteligencia artificial de soporte a las decisiones sanitarias o que se incorporan como un criterio adicional al tratante requieren tener en cuenta el principio de inclusión de la diversidad en la recolección y en el procesamiento de datos, así como, en el diseño del algoritmo. La complejidad de la relación clínica, su fundamento en la confianza y los riesgos de vulneración de los pacientes como a la salud, integridad física y psíquica requiere también emitir su consentimiento libre e informado.

² Kroet, Cynthia (2025). The AI Liability directive was taken off the Commission's 2025 work program due to the lack of progress in negotiations. Disponible en: <https://bit.ly/46kvPw9>. Revisado en: 14-04-2025.

La inclusión de la diversidad consiste en abordar la construcción misma del sistema desde el control activo respecto de la aparición de sesgos por subrepresentación de grupos con mayor susceptibilidad a la vulnerabilidad, por lo que la implementación del algoritmo puede reproducir inequidades. La consideración transversal del principio de inclusión de la diversidad se extiende al control posterior que la auditoría algorítmica debe realizar, misma que debe estar a cargo de una revisión humana.

La consideración de los sistemas de inteligencia artificial para el soporte de la decisión sanitaria como herramienta sociotécnica que debe tener en cuenta las determinantes sociales y el contexto son necesarias para que los dispositivos regulatorios de la inteligencia artificial respondan a un enfoque de derechos, al coincidir el centro de la protección en los grupos más desaventajados, lo que se manifiesta también en la prevención del riesgo de discriminación como centro del Reglamento Europeo que regula la materia, los deberes de información reforzados de los proveedores, y las presunciones incorporadas en la reciente a la Directiva Europea que regula el régimen de productos defectuosos, todos mecanismos centrados el fortalecer la transparencia y favorecer a la explicabilidad de los sistemas. Queda, sin embargo, como ámbito para profundizar la investigación el análisis profundo del régimen de responsabilidad por productos defectuosos aplicable a sistemas de alto riesgo, como es el caso de aquellos que sirven de soporte a las decisiones sanitarias; así como las propuestas discutidas sobre las reglas de responsabilidad extracontractual en la materia.

Referencias

- Agarwal, R.; Bjarnadottir, M.; Rhue, L.; Dugas, M.; Crowley, K.; Clark, J.; y Gao, G. (2023). Addressing algorithmic bias and the perpetuation of health inequities: An AI bias aware framework. *Health Policy and Technology*, 12(1): pp. 100702. <https://doi.org/10.1016/j.hlpt.2022.100702>.
- Añón Roig, M. J. (2022). Desigualdades algorítmicas: Conductas de alto riesgo para los derechos humanos. *Derechos y Libertades: Revista de Filosofía del Derecho y derechos humanos*, 47, pp. 17-49. <https://doi.org/10.20318/dyl.2022.6872>.
- Barda, N., Yona, G., Rothblum, G., Greenland, P., Leibowitz, M., Balicer, R., Bachmat, E. y Dagan, N. (2021). Addressing Bias in Prediction Models by Improving Subpopulation Calibration. *Journal of the American Medical Informatics Association: JAMIA*, 28(3), pp. 549-58. <https://doi.org/10.1093/jamia/ocaa283>.
- Barrientos, F., Bozzo, S. y Jequier, E. (2023). Nuevas tecnologías para el acceso a la justicia del consumidor: Diagnóstico de la situación en Chile y una propuesta a la luz de la plataforma europea online de resolución de conflictos. *Revista Chilena De Derecho Y Tecnología*, 12: pp. 1-35. <https://doi.org/10.5354/0719-2584.2023.69963>.
- Bedecarratz Scholtz, F., Aravena Flores, M. (2023). Principios y directrices sobre inteligencia artificial”, en: Azuaje Pirela, Michelle. (Coord.) *Introducción a la ética y el derecho de la Inteligencia Artificial*. Wolters Kluwer.
- Calahorrano Latorre, Edison (2025a). Equidad algorítmica y decisiones clínicas basadas en sistemas de soporte basados en inteligencia artificial. *Nuevo Derecho*, 21(36), 1-17. <https://doi.org/10.25057/2500672X.1691>.
- Calahorrano Latorre, Edison (2025b). Responsabilidad respecto de los daños provocados por dispositivos médicos basados en inteligencia artificial. Argumentos para su tratamiento en sede de derechos de Consumo, en Alberto Pino y Esteban Pereira, *Estudios de Derecho del Consumo*. Thomson Reuters.

- Cobas Cobiella, M. E. y Pérez Cobas, A. E. (2023). Medicina personalizada. Retos para el derecho, en *Derecho y medicina: desafíos tecnológicos y científicos*. Dykinson.
- Čartolovni, A., Tomičić, A., y Lazić Mosle, E. (2022). Ethical, legal, and social considerations of AI-based medical decision-support tools: A scoping review. *International Journal of Medical Informatics*, 161: pp. 104738. <https://doi.org/10.1016/j.ijmedinf.2022.104738>.
- Coddou Mc Manus, A., y Smart Larraín, S. (2021). La transparencia y la no discriminación en el Estado de bienestar digital. *Revista chilena de derecho y tecnología*, 10(2). <https://doi.org/10.5354/0719-2584.2021.61034>.
- Coddou Mc Manus, A., y Padilla Parga, R. (2024). Discriminación algorítmica en los procesos automatizados de reclutamiento y selección de Personal. *Revista Chilena de Derecho y Tecnología*, 13: pp. 1-34. <https://doi.org/10.5354/0719-2584.2024.71312>.
- Coddou Mc Manus, A., Smart, S. y Vivallo, V. (2024). Derechos Humanos e Inteligencia Artificial: el caso chileno; en: Schönsteiner, Judith; Fuentes, Claudio; Barrera, Senead; *Informe Anual de Derechos Humanos*: Universidad Diego Portales.
- Ferrario, A., Gloeckler, S. y Biller-Andorno, N. (2023). Ethics of the Algorithmic Prediction of Goal of Care Preferences: From Theory to Practice. *Journal of Medical Ethics*, 49 (3): pp. 165-74. <https://doi.org/10.1136/jme-2022-108371>.
- Fosch-Villaronga, E., Drukarch, H., Khanna, P., Verhoef, T., Custers, B. (2022). Accounting for diversity in AI for medicine. *Computer Law & Security Review*, 47: p. 105735. <https://doi.org/10.1016/J.CLSR.2022.105735>.
- Grote, T., y Keeling, G. (2022). On Algorithmic Fairness in Medical Practice. *Cambridge Quarterly of Healthcare Ethics: CQ: The International Journal of Healthcare Ethics Committees*, 31(1): pp. 83-94. <https://doi.org/10.1017/s0963180121000839>.
- Grote, T. y Berens, P. (2020). On the ethics of algorithmic decision-making in healthcare. *Journal of Medical Ethics*, 46(3): pp. 205-211. <https://doi.org/10.1136/medethics-2019-105586>.
- Hegde, P. Shenoy, M. M. (2021). Artificial Intelligence. *Medicine and Health Sciences. Archives of Medicine and Health Sciences*, 9(1), pp. 145-150. https://doi.org/10.4103/amhs.amhs_315_20.
- Huang, J., Galal, G., Etemadi, M. y Vaidyanathan, M. (2022). Evaluation and mitigation of racial bias in clinical machine learning models: Scoping review. *JMIR Medical Informatics*, 10(5): p. e36388. <https://doi.org/10.2196/36388>.
- Johnson, S. (2019). AI, Machine Learning, and Ethics in Health Care. *Journal of Legal Medicine*, 39(4): pp. 427-441. <https://doi.org/10.1080/01947648.2019.1690604>.
- Jones, C., Thornton, J. y Wyatt, J. (2023). Artificial intelligence and clinical decision support: clinicians' perspectives on trust, trustworthiness, and liability. *Medical Law Review*, 31(4): pp. 501-520. <https://doi.org/10.1093/medlaw/fwad013>.
- Juri, Yamila (2021). Inteligencia artificial y dignidad humana: Los desafíos para el derecho. *Revista Justicia & Derecho*, 4(2), 1-12. <https://doi.org/10.32457/rjyd.v4i2.1350>.
- Karimian, G., Petelos, E. y Evers, S. (2022). The Ethical Issues Of The Application Of Artificial Intelligence In Healthcare: A Systematic Scoping Review. *Ai And Ethics*, 2(4), pp. 539-551. <https://doi.org/10.1007/S43681-021-00131-7>.
- Kumar, P., Chauhan, S. y Awasthi, L. (2023). Artificial intelligence in healthcare: Review, ethics, trust challenges and future research directions. *Engineering Applications of Artificial Intelligence*, 120: p. 105894. <https://doi.org/10.1016/j.engappai.2023.105894>.

- Lazcoz Moratinos, G. y Castillo Parrilla, J. A. (2020). Valoración algorítmica ante los derechos humanos y el Reglamento General de Protección de Datos: el caso SyRI. *Revista Chilena de Derecho y Tecnología*, 9(1): pp. 207-225. <https://doi.org/10.5354/0719-2584.2020.56843>.
- Lillywhite, A. y Wolbring, G. (2021). Coverage of ethics within the artificial intelligence and machine learning academic literature: The case of disabled people. *Assistive Technology: The Official Journal of RESNA*, 33(3): pp. 129-135. <https://doi.org/10.1080/10400435.2019.1593259>.
- Liyanage, H., Liaw, S.-T., Jonnagaddala, J., Schreiber, R., Kuziem Sky, C. y Terry, A. (2019). Artificial intelligence in primary healthcare: perceptions, issues, and challenges. *Yearb. Med. Inform*, 28: pp. 41-46.
- Llamas, J., Mendoza, O., y Graff, M. (2022). Enfoques regulatorios para la inteligencia artificial (IA). *Revista chilena de derecho*, 49(3): pp. 31-62. <https://dx.doi.org/10.7764/r.493.2>.
- Madrid, R. (2023). Las máquinas y la agencia moral, en: Azuaje Pirela, Michelle. (Coord.) *Introducción a la ética y el derecho de la Inteligencia Artificial*. Wolters Kluwer España, pp. 87-101
- Martín-Casals, Miquel (2022). Desarrollo tecnológico y responsabilidad extracontractual. A propósito de los sistemas de inteligencia artificial, en: Pérez, José y Sajuán, Francisco, *La Cultura Jurídica en la era digital* (pp. 101-138). Aranzadi.
- Martín-Casals, M. (2022b). An approach to some EU initiatives on the regulation of liability for damage caused by AI-Systems. *Ius et Praxis*, 28(2): pp. 3-24. <https://doi.org/10.4067/s0718-00122022000200003>.
- Martín Casals, M. (2024). Líneas generales de la nueva Directiva Europea de Responsabilidad por los daños causados por productos defectuosos, *Revista Actualidad Jurídica*, 50: pp. 35-76.
- Molnár-Gábor, F. y Giesecke, J. (2022). Medical AI Key Elements at the International Level, en Silja Voenekey, Phillip Hellmeyer; Oliver, Mueller y Wolfram, Burgard, *The Cambridge Handbook of Responsible Artificial Intelligence*. Cambridge University Press pp. 379-396.
- Navas, S. (2021). Salud Electrónica e Inteligencia Artificial. En *Salud e Inteligencia Artificial desde el Derecho Privado. Con especial atención a la pandemia SARS COv2 COVID-19* (pp. 1-48). Comares.
- Navas, S. (2022). *Daños ocasionados por sistemas de inteligencia artificial: Especial atención a su futura regulación*. Comares.
- Perin, A. (2019). Estandarización y automatización en medicina: El deber de cuidado del profesional entre la legítima confianza y la debida prudencia. *Revista Chilena de Derecho y Tecnología*, 8(1): pp. 3-28. <https://dx.doi.org/10.5354/0719-2584.2019.52560>.
- Pfohl, S., Foryciarz, A., y Shah, N. (2021). An empirical characterization of fair machine learning for clinical risk prediction. *Journal of Biomedical Informatics*, 113: 103621. <https://doi.org/10.1016/j.jbi.2020.103621>.
- Pricitor, M. (2023). Where does responsibility lie? Analysing legal and regulatory responses to flawed clinical decision support systems when patients suffer harm. *Medical Law Review*, 31(1): pp. 1-24. <https://doi.org/10.1093/medlaw/fwac022>.
- Ramón, F. (2021). Inteligencia artificial en la relación médico-paciente: Algunas cuestiones y propuestas de mejora. *Revista chilena de derecho y tecnología*, 10(1): pp. 329-351. <https://dx.doi.org/10.5354/0719-2584.2021.60931>.
- Ramón, F.(2024). Diagnóstico Médico por robots y responsabilidad civil. *Lex Medicinae*, 21(42): pp. 17-34. Disponible en <https://bit.ly/4ntG2gc>.
- Recuero, M. (2024). El uso secundario de datos de salud electrónicos: El futuro Reglamento del Espacio Europeo de Datos de Salud y su interacción con la protección de datos personales. *InDret*, 2: pp. 525-51. <https://doi.org/10.31009/InDret.2024.i2.13>.

- Raposo, V. (2025). The fifty shades of black: about black box AI and explainability in healthcare, *Medical Law Review*, 33(1): fwaf005, <https://doi.org/10.1093/medlaw/fwaf005>.
- Smith, H., y Fotheringham, K. (2020). Artificial Intelligence in Clinical Decision-Making: Rethinking Liability. *Medical Law International*, 20(2):pp. 131-54. <https://doi.org/10.1177/0968533220945766>.
- Smuha, N. Ahmed Renger, E.; Harkens, A., *et.al* (2021). How Can the EU Achieve a Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act, *Leeds Law University of Birmingham*, 1.
- Solaiman, B.; Malik, A. (2025). Regulating algorithmic care in the European Union: Evolving doctor-patient models through the Artificial Intelligence Act (AI-Act) and the liability directives. *Medical Law Review*, 33(1): fwae033. <https://doi.org/10.1093/medlaw/fwae033>.
- Steerling, E. Siira, E., Nilsen, P., Svedberg, P y Nygren, J. (2023). Implementing AI in healthcare: the relevance of trust: a scoping review. *Frontiers in health services*, 3. <https://doi.org/10.3389/frhs.2023.1211150>.
- Van Baalen, S. Boon, M., y Verhoef, P. (2021). From clinical decision support to clinical reasoning support systems. *Journal of evaluation in clinical practice*, 27(3), pp. 520-528. <https://doi-org.atalca.idm.oclc.org/10.1111/jep.13541>.
- Walker, N. (2023). El problema de sesgo algorítmico y, en particular, el sesgo de género, en: Droguett, Carmen; Walker, Nathalie, *Derecho Digital y Privacidad en América y Europa* (pp. 148-161). Tirant Lo Blanch.

Normas citadas

Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 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).