Artículos

Equidad algorítmica y decisiones clínicas basadas en sistemas de soporte basados en inteligencia artificial

Algorithmic fairness and clinical decisions based on artificial intelligence-based support systems

Edison Calahorrano Latorre
Universidad de Tarapacá, Chile
ecalahorranol@academicos.uta.cl

Nuevo Derecho vol. 21 núm. 36 1 17
2025

Institución Universitaria de Envigado
Colombia

Recepción: 23 Octubre 2024
Aprobación: 19 Febrero 2025
Publicación: 04 Abril 2025

Resumen: En este trabajo se desarrollan los conceptos de equidad algorítmica e inclusión de la diversidad como propuesta de mecanismos de prevención de la discriminación en la toma de decisiones clínicas cuando interviene un sistema de soporte basado en inteligencia artificial. A partir de una revisión de la literatura especializada, aplicando los métodos dogmático y analítico, se hace un análisis de los sesgos que pueden generarse en este proceso. En segundo lugar, se profundiza respecto a la manera en que los mismos quebrantan la confianza del paciente y la fiabilidad del sistema. Finalmente, se propone la incorporación de la equidad algorítmica y de la inclusión de la diversidad como elementos transversales en la construcción e implementación de decisiones automatizadas en salud.
Palabras clave: Equidad algorítmica, Inclusión, Discriminación algorítmica, Decisión clínica, Igualdad y no discriminación.
Abstract: This paper develops the concepts of algorithmic fairness and inclusion of diversity as proposed mechanisms for preventing discrimination in clinical decision-making when artificial intelligence-based support systems intervene. Based on a review of specialized literature, applying the dogmatic and analytical methods, an analysis is made of the biases that can be generated in this process. Secondly, an in-depth analysis is made of the way in which these biases undermine patient confidence and the reliability of the system. Finally, the incorporation of algorithmic equity and diversity inclusion as transversal elements in the construction and implementation of automated health decisions is proposed.
Keywords: Algorithmic Equity, Inclusion, Algorithmic Discrimination, Clinical Decision, Equality and Non-discrimination.

Introducción

"Inteligencia artificial" fue la expresión del 2022 según la Fundación de Español Urgente, promovida por la Agencia EFE y la Real Academia Española (Fundéu RAE, 2022). El crecimiento del número de publicaciones científicas en el que consta esta expresión se ha multiplicado exponencialmente en los últimos cinco años, especialmente respecto a sus aplicaciones en el ámbito de la salud.[2]

El crecimiento de las aplicaciones de la inteligencia artificial en salud aborda las más variadas disciplinas, enfocándose ya algún tiempo en el diagnóstico y detección temprana de patologías (Sunarti et al., 2021, p. 567). Esta se ha extendido también a distintas disciplinas de la salud; en dermatología, por ejemplo, el procesamiento de imágenes mediante machine learning (ML) ha permitido una mayor precisión en la identificación de enfermedades e infecciones de la piel. De la misma manera, sistemas predictivos y modeladores de tratamiento han sido utilizados en el ámbito de la oncología, en la que la capacidad de análisis masivo y certeza ha permitido alcanzar un eficiente mecanismo de identificación temprana de casos de alto riesgo para tratamiento temprano. Finalmente, otro avance relevante se ha producido en el manejo de los datos médicos de paciente y en las posibilidades de elaborar una estrategia terapéutica a la medida tras su adecuado registro e interpretación, lo que arroja una de las principales novedades que la aplicación de la inteligencia artificial está desarrollando: la medicina de precisión o personalizada (Hegde y Shenoy, 2021, pp. 146-148).

Al igual que en otras aplicaciones de la inteligencia artificial la aparición de sesgos en el proceso que reproduzcan inequidades es la mayor preocupación en el funcionamiento de los sistemas de soporte de la decisión clínica. Lo señalado ha sido constatado por la doctrina nacional:

> Es preciso, entonces, tener en cuenta que los resultados que arroje un sistema algorítmico —por muy sofisticada que pueda ser la arquitectura de que esté dotado— no necesariamente encarnarán la pretendida imparcialidad que se predicó de ellos durante tanto tiempo y con tan ingenuo entusiasmo. Detrás de todo sistema, sea de inteligencia artificial o un sistema experto, siempre habrá detrás una o más personas que lo diseñen, lo entrenen, lo controlen, lo apliquen y se beneficien de él. El hecho de que no percibamos su presencia no implica en absoluto que hayan dejado de estar detrás de todas esas etapas.Una vez que fue posible constatar que los algoritmos no eran neutrales per se, la comunidad científica y académica comenzó a ocuparse del estudio del sesgo algorítmico. La forma de abordar este problema se ha sostenido, básicamente, en dos grandes premisas: la detección de sesgos y cómo poder evitar que se produzcan y multipliquen en el futuro (Walker, 2023, p. 150).

Por su parte, se ha constatado que el principio de justicia está presente en todas las directrices éticas que se han desarrollado para la

aplicación de la inteligencia artificial, entendiéndose por tal "que los sistemas de IA no generen efectos discriminatorios o injustos en relación con categorías tales como el sexo, la edad, la religión, el origen geográfico, el idioma, etc., previniendo, monitoreando y mitigando los sesgos no deseados en el sistema" (Bedecarratz Scholtz y Aravena Flores, 2023, p. 209).

Con base en lo señalado, en el presente trabajo se pretende indagar respecto de la discriminación algorítmica que puede producirse tras la generación de sesgos en un sistema de inteligencia artificial cuando actúa como soporte o criterio experto adicional en la toma de decisiones clínicas, tanto en su uso para diagnóstico como en la selección de una alternativa terapéutica para el caso concreto.

La estandarización de la medicina a través de guidelines o manuales de buenas prácticas ha generado la necesidad de analizar cómo se debe actuar frente al reconocimiento de protocolos que configuran la actuación de tratante medio, pero que pueden ser muy heterogéneos en su origen (Perin, 2019, pp. 5-10), por lo que la uniformidad en las consideraciones que se deben tener en cuenta para el ejercicio del deber de cuidado requiere el análisis exhaustivo de caso específico.

La estandarización se encuentra ahora caracterizada por los sistemas de soporte de decisión clínica (SSDC), por lo que cabe verificar si la deferencia de tratante a los resultados arrojados por el sistema es determinante respecto del ejercicio del deber de cuidado y de su amplitud a la verificación intrínseca de la prestación y extrínseca del modelo de soporte de la decisión clínica (Perin, 2019, p. 11). Lo señalado requiere una reforzada evaluación del tratante tanto respecto de la precisión de la misma como de la corrección de subrepresentaciones que pueden ser reproducidas por sesgos respecto de la diversidad de los pacientes que debe reflejarse en los datos y su procesamiento, específicamente, respecto de grupos de mayor vulnerabilidad (Prictor, 2023, pp. 22-24). Sin embargo, por otro lado, se ha defendido que 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 (Smith y Fotheringham, 2020, pp. 146-154; Verdicchio y Perin, 2022, pp. 23-24).

El problema jurídico planteado radica en la necesidad de establecer mecanismos que permitan prevenir la perpetuación de inequidades, en la medida en que existe consenso en que la salud se encuentra afectada por determinantes sociales (Fletcher et al., 2021, p. 3) o especificidades con base en el género o el sexo (Cirillo et al., 2020 p. 8; Fosch-Villaronga et al., 2022, p .2) que deben ser incorporadas a las distintas fases de desarrollo del algoritmo o red neuronal en la que se basa el SSDC.

Se plantea la hipótesis de que la implementación de SSDC debe establecer mecanismos que prevengan la discriminación algorítmica, especialmente en las fases de almacenamiento, depuración y

aseguramiento de la calidad de los datos, así como en su construcción e implementación, de tal forma que se reduzca la opacidad del proceso que permite la toma de decisiones y la posibilidad de sesgos que pueden vulnerar derechos fundamentales manifestándose en daños con un efecto multiplicador. Es posible afirmar que los mecanismos de prevención se identifican con los conceptos de equidad algorítmica e inclusión de la diversidad; ambos consolidan principios que sirven como referencia para la normativa que pretende regular la inteligencia artificial en el ámbito de salud.

Para comprobar la hipótesis se analizará, desde una perspectiva jurídica, qué es y cómo funciona la inteligencia artificial (IA), cuál es el papel de los algoritmos y de las redes neuronales en su funcionamiento y, posteriormente, se elaborará un concepto de discriminación algorítmica y los riesgos que los sesgos producen en la implementación de los SSDC para, finalmente, construir una propuesta que permita la prevención de la discriminación algorítmica a través de la incorporación de la equidad algorítmica y de la inclusión de la diversidad como parte de la transparencia y la fiabilidad.

Metodología

El método que se utilizará para la presente investigación es el dogmático, recopilándose documentos de los últimos cinco años bajo las palabras clave “Clinical Decision Support System”, “discriminación algorítmica”, “inteligencia artificial en salud”, “algorithmic bias” y “artificial intelligence in healthcare” en las bases de datos PUBMED, SCOPUS, WOS y Scielo Chile. Se aplica además el método analítico para seleccionar aquellos trabajos que aborden exclusivamente la discriminación algorítmica.Con el presente trabajo se pretende contribuir a una revisión sistemática del estado del arte sobre las respuestas de Derecho desde la perspectiva preventiva a la discriminación algorítmica y de redes neuronales en salud.

¿En qué consiste la discriminación algorítmica?

Se ha señalado que la inteligencia artificial tiene el potencial para ser protagonista de la Cuarta Revolución Industrial junto con otras tecnologías emergentes como el blockchain o el internet de las cosas. Las razones que permiten aseverar lo señalado se refieren a su rápida inserción en aplicaciones de la vida cotidiana y a la manera en que interactuamos, producimos o ejercemos el cuidado (Martín-Casals, 2022b, 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 de algoritmos para que ofrecer una asistencia adecuada a los tratantes en la toma de decisiones, afirmándose además que la misma debe estar caracterizada

por la transparencia y por las posibilidades de explicación. Estas características permitirían al tratante interpretar adecuadamente los resultados (Van Baalen et al., 2021, p. 526-527). En este contexto, los SSDC han sido definidos como mecanismos de conocimiento activo que usan datos de pacientes para generar asesoría y apoyo a la decisión clínica en el caso concreto, lo que sirve como insumo para el diagnóstico, la selección del tratamiento o la alternativa terapéutica, el monitoreo, los flujos de trabajo, el almacenamiento y el procesamiento de la información clínica del paciente (Jones et al., 2023, p. 2).

El soporte en la toma de decisión clínica ha sido uno de los ámbitos en los que esta tecnología ha encontrado mayores dificultades para alcanzar la confianza y la fiabilidad de los usuarios (Jones et al., 2023, pp. 2-3), por cuanto el signo característico del ML y su subtipo más reciente, el deep learning (DL), tienen como fin simular y sobrepasar el razonamiento humano (P. Kumar et al., 2023).

El primer concepto se refiere a 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. El segundo concepto es especialmente complicado, ya que 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). Adicionalmente es necesario advertir que estos sistemas pueden ser supervisados cuando se cuenta con la presencia de un controlador humano que introduce los datos y monitorea el análisis; mientras que no son supervisados en los casos en que pueden generar sus propias conclusiones y desarrollar un nuevo set de datos a partir de los originales. Finalmente, existen los sistemas de aprendizaje reforzado, en los que el agente humano introduce incentivos positivos y negativos como recompensas y sanciones que permiten corregir en el entrenamiento mismo aquellos resultados no deseados (Kumar et al., 2023, pp. 5-6).

La preocupación en el tema que nos atañe se configura al introducir en los distintos ámbitos de la salud de la vida cotidiana una tecnología que permite incrementar notablemente la precisión en las decisiones relevantes, como los diagnósticos y la selección de la mejor y más personalizada alternativa terapéutica, pero que, a la vez, en su potencial predictivo y personalizado corre el riesgo de producir daños recurrentes debido a los sesgos en los datos que sirven para su funcionamiento y repercutan en reproducción de inequidades (Agarwal et al., 2023).

Al respecto, se ha profundizado en la literatura nacional en el estudio crítico de los datos como un esquema epistemológico relevante para evitar una exacerbada deferencia en lo que estos pueden aportar a la toma de decisiones, sobre todo en ámbitos en que hay derechos fundamentales que 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 datos utilizados y de la manera en que se relacionan: además se debe verificar el contexto político estatal o institucional en el que se implementan y los stakeholders involucrados, por ejemplo, para adecuar la mercantilización de los datos o su uso en un contorno que se mantenga en un enfoque de derechos (Coddou Mc Manus y Smart Larraín, 2021, pp. 307-310).

En la misma línea de lo señalado resulta interesante la aproximación de Pfohl et al. (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, como la educación, lo que produce una subrepresentación de determinados grupos vulnerables que se alejan de aquel modelo en base al cual es construido el algoritmo, con determinados atributos demográficos, provocando que los mismos se descontextualicen.

A continuación, se presentará una aproximación a lo que se entiende por inteligencia artificial y su funcionamiento, en la que se incluirán las principales aplicaciones en salud, desde la medicina de precisión por el procesamiento de big data de imágenes en el ámbito radiológico hasta la toma de decisiones asistida o fortalecida por resultados proporcionados por sistemas de machine y deep learning cuya influencia permite considerarlos como un miembro más del equipo médico. Esta aproximación permitirá posteriormente abordar los principales debates ético-jurídicos sobre la discriminación algorítmica en los SSDC.

La inteligencia artificial y sus aplicaciones en el ámbito de la salud

Respecto del primer objetivo planteado para la comprobación de la hipótesis señalada, es necesario constatar que la inteligencia artificial se traduce, esencialmente, en un software que busca cumplir con determinados procedimientos como si fuesen realizados por seres humanos, procedimientos que son desarrollados con base en el manejo de macrodatos que son gestionados y procesados buscando la mayor eficiencia (Alowais et al., 2023, p. 2; Fosch-Villaronga et al., 2022, pp. 1-2; Ramón Fernández, 2021, p. 331)

La función señalada tiene como centro modelos probabilísticos y algoritmos que le permiten intervenir, con distintos grados de autonomía, en predicciones, recomendaciones o toma de decisiones (Parlamento Europeo, 2020).

La inteligencia artificial puede ser simbólica, basada en conocimiento apoyado en reglas de expertos para producir los

resultados, o subsimbólica si incorpora la tecnología ML, por la cual, a partir de determinados datos se genera un aprendizaje automático basado en algoritmos (Silcox, 2020, p. 5). Este proceso tiene distintos grados de claridad respecto de la manera en que se llega a conclusiones, por lo que los daños que puedan producirse ante su uso y la necesidad de lograr su prevención y reparación se convierten en un asunto desafiante al definir quién y bajo qué estatuto es responsable (Comisión Europea, 2020, pp. 15-16; Perin, 2019, pp. 12-14; Prictor, 2023, pp. 4-6). La situación se agrava cuando se constata que es en el ámbito de la salud que esta tecnología ha tenido una gradual consolidación, por lo tanto, los daños que eventualmente puedan provocarse recaen en la integridad física del paciente.

El ML es la subdisciplina de la inteligencia artificial relacionada con el aprendizaje autónomo de las máquinas, mientras que el DL consiste en una especialización del primero que aborda el uso de modelos construidos en redes neuronales profundas capaces de detectar patrones con mínima o ninguna intervención humana. Estas herramientas se han insertado en el cuidado de la salud, especialmente, como soportes para la toma de decisiones en diagnósticos y tratamientos altamente individualizados; sus capacidades predictivas conllevan la necesidad de que los beneficios que proponen sean equitativos para todo tipo de pacientes (Alowais et. al, 2023: pp. 2-3).

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 el 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.

Estos dos últimos enfoques se derivan de la posibilidad de procesar y generar conclusiones autónomas a partir de una gran cantidad de datos, entre ellos, valores y preferencias de los pacientes, lo que permite considerar a los sistemas de inteligencia artificial como herramientas sociotécnicas que permiten tener mayor claridad sobre el uso, la evaluación y la mejora continua de la asistencia que pueden prestar en 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, lo que, a su vez, debe ser incorporado en una categoría más amplia como es la de equidad algorítmica.

El proceso que sigue un modelo de inteligencia artificial en el ámbito de salud ha sido esquematizado (Kumar et al., 2023, pp. 5-6) e inicia con la recolección de datos de diferentes fuentes como registros electrónicos que contienen información sanitaria relevante, las fichas clínicas electrónicas, lo que conlleva desafíos importantes para anonimizar y proteger los datos sensibles; sin embargo, los datos

pueden provenir también de ensayos clínicos y dispositivos portables (wereables). El segundo paso lo constituye la limpieza y el reprocesamiento de los datos para que se encuentren aptos para su uso en el sistema; en esta fase cobran relevancia los mecanismos de control y transparencia para la erradicación de sesgos que puedan generar resultados discriminatorios, por cuanto en la calidad de los datos disponibles radica la precisión de las decisiones automatizadas. Al paso anterior se incorpora la ingeniería del mecanismo que permite construir variables y data points para aplicar como imputs del modelo a ser entrenado, momento en el que se establece el algoritmo que permite al sistema de ML aprender patrones y al control humano realizar ajustes para minimizar la aparición de sesgos. Una vez entrenado el algoritmo se procede a su evaluación, comparándolo con una línea base o modelo alternativo para una posterior implementación real.

La precisión del algoritmo en sus conclusiones depende de la calidad de los datos ingresados y del diseño del mecanismo con el que se produce su interrelación, lo que puede ser un desafío en el diagnóstico médico (Pethig y Kroenung, 2023, p. 638). En este último caso es conocido que un mismo conjunto de hechos puede llevar a varias interpretaciones de diferentes profesionales de la salud, aunque sea una misma situación (Naik et al., 2022, p. 2; Xu et al., 2023, p. 5). Lo señalado conlleva la necesidad de verificar que los datos sean confiables y de estratificar las evidencias para lograr una mayor fiabilidad; así, las opiniones de los expertos ofrecen menor certeza que los datos que pueden ser arrojados por series de casos, estudios con grupo de control y revisiones sistemáticas o meta análisis, lo que ha llevado a la identificación de impedimentos para la incorporación integral de la inteligencia artificial en la toma de decisiones clínica, reconociéndose, entre ellos, los altos costos de los procedimientos que permiten alcanzar datos confiables, la adaptabilidad, los debates éticos sobre la pérdida de control humano y la atribución de responsabilidad por los daños que genere el sistema (Holmes, 2017, pp. 68-70; A. Kumar et al., 2023, p. 18031; Kumar et al., 2023, pp. 7-8).

A partir de la comprensión del funcionamiento del ML y del DL —y de las limitaciones que se han señalado respecto de sus posibilidades para abordar la diversidad, el contexto, los determinantes y las especificidades que requiere cada caso—, se puede profundizar respecto de la discriminación algorítmica y de los desafíos de proponer soluciones jurídicas frente a sus efectos en cuanto a la fiabilidad y a la confianza que se requiere para la consolidación de los SSDC.

Discriminación algorítmica en los SSDC: ¿cómo se producen los sesgos?

La maximización de la precisión y de la eficiencia proporcionada por los sistemas que automatizan la toma de decisiones puede verse afectada por factores específicos que producen inequidades. Ante esta dificultad aparece la equidad algorítmica (algorithmic fairness) como un área del ML y del DL que guía el diseño de modelos con el objetivo de prevenir la discriminación que afecta a grupos de protección reforzada como raza, género, religión, diversidad fisiológica, condiciones preexistentes, entre otros (Barda et al., 2021, p. 551; Coddou Mc Manus y Smart Larraín, 2021, pp. 317-319; Grote y Berens, 2020, pp. 9-10; Grote y Keeling, 2022, p. 9-10; Huang et al., 2022, p. 3).

El desarrollo de los mecanismos de equidad algorítmica en la aplicación de SSDC se encuentra en construcción, especialmente por la dificultad de abordar la diversidad de dilemas éticos frente a la necesidad de mantener criterios generalizables para el procesamiento de la información por el algoritmo; es así que este puede incurrir en la perpetuación de inequidades, así como enfocarse en muestras convenientes que dejan de lado a poblaciones vulnerables poco representadas. Lo señalado repercute en la vulneración del principio de no maleficencia. De igual manera, evaluar el funcionamiento del algoritmo en poblaciones más genéricas no permite que el mismo pueda tomar en cuenta las características de subpoblaciones (Chen et al., 2021; p. 19).

La discriminación algorítmica puede ser ejemplificada por casos recientes de decisiones automatizadas en la asignación de recursos como el ocurrido en el caso SyRI, respecto a los beneficios para el cuidado de niños y la generación de perfiles de riesgo de fraude para la seguridad social (Amnistía Internacional, 25 de octubre de 2021; Coddou Mc Manus y Smart Larraín, 2021, p. 305; Lazcoz Moratinos y Castillo Parrilla, 2020, pp. 215-222). También puede ser ejemplificada por los sesgos en el cuidado en salud provocados por algoritmos utilizados por los hospitales basados en características étnicas u otras que provocan una subrepresentación en la asignación de tratamientos personalizados para la atención de patologías complejas (Ledford, 2019).

La aplicación de algoritmos para el diagnóstico de enfermedades se ha logrado consolidar en ámbitos como la radiología y el análisis de imágenes para la identificación de retinopatía diabética; así como, la detección de cáncer a través de análisis masivo de mamografías, en los que ha demostrado mayor precisión que los profesionales especializados (Alowais et al., 2023, p. 3). Se ha implementado con éxito un sistema de detección y tratamiento de apendicitis que ha permitido la prevención de su agravamiento sugiriéndose posibilidades de implementar mecanismos de detección precisos de otras enfermedades como el COVID-19; en este caso, el sistema predictivo basado en ML permite el apoyo y la confirmación del diagnóstico de expertos (Mijwil y Aggarwal, 2022, pp. 7011-23).

La aplicación de inteligencia artificial para el diagnóstico clínico tiene, sin embargo, un importante desafío respecto de la generación de confianza y fiabilidad (Jones et al., 2023, p. 19). Se ha verificado que desde la perspectiva de médicos y profesionales de la salud el mayor aporte de la aplicación de algoritmos y redes neuronales a través de sistemas de machine y deep learning, respectivamente, consiste en la precisión, por lo que representa un recurso valioso para la toma de decisiones (Boden, 2018, pp. 70-72); sin embargo, las reflexiones sobre principios éticos que caracterizan la humanización de la relación clínica y la aplicación de la medicina centrada en el paciente son presentadas como las principales preocupaciones respecto a la protección de la autonomía del paciente (Karimian et al., 2022, pp. 539–551).

Lo último se concreta en la posibilidad de que el SSDC no sea capaz de entregar soluciones que estén centradas en los valores, preferencias y necesidades del paciente, así como en el análisis de su contexto, incluyendo la verificación de determinantes sociales aplicables a su caso y de variables demográficas, económicas y sociales que pueden correr el riesgo de no estar incluidas después de la limpieza de datos y de la construcción del algoritmo provocando discriminación inadmisible (Liyanage et al., 2019, pp. 41-46). El principio por el cual el algoritmo o el funcionamiento de la red neuronal debe ser explicable destaca como otro elemento central con relevancia ética y jurídica que deriva de ello, ya que la opacidad no permite la comunicación transparente con el paciente y genera el cuestionamiento sobre la justificación que debe exhibir el profesional de la salud para alejarse del resultado arrojado por el algoritmo y la eventual responsabilidad por dicha decisión; por otro lado, este principio permite discutir si en el catálogo de derechos de las y los pacientes debe incluirse el control humano y la posibilidad de explicación del algoritmo (Anderson y Anderson, 2019, pp. 5-6; Braun et al., 2020, E125–130).

La privacidad del paciente y la protección de datos son también presupuestos para la fiabilidad y la confianza en los SSDC bajo el presupuesto de que las bases de datos de información clínica y los wereables son fuentes principales (Karimian et al., 2022, pp. 544-547). Finalmente, otros principios discutidos son la equidad (fairness) y la prevención de daño, identificables además con dos principios bioéticos clásicos: la justicia y la no maleficencia.

Con respecto al primero, se ha comprobado que la fase de desarrollo del algoritmo es la que puede originar discriminación, falta de equidad y no reconocimiento de la diversidad ni de la inclusión e, 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 (Chen et al., 2021, pp. 167-179). Otro caso de discriminación se representa directamente respecto de la población vulnerable, como las personas con discapacidad, que suelen estar infra

representadas al momento de la construcción del algoritmo (Lillywhite y Wolbring, 2021, pp. 129-135).

La prevención de daño mediante la equidad algorítmica y la inclusión de la diversidad

Sobre el caso de los SSDC se plantean dos cuestiones: en primer lugar, cuando el resultado proporcionado por el sistema es considerado un insumo más para la decisión del equipo médico; y en segundo lugar, cuando este resultado es considerado como la opinión de un miembro más de equipo (Ahuja, 2019, pp. 19-20; Perin, 2019, pp. 11-21).

Lo dicho permite profundizar en qué consiste el deber de cuidado de tratante cuando la toma de decisiones en la relación clínica se encuentra afectada por un sistema de soporte basado en inteligencia artificial, específicamente, la posición que debe tomar el tratante cuando existen divergencias entre sus conclusiones y las del sistema. Si bien esta discusión es pertinente para la determinación de la responsabilidad, el debate puede ser reconducido a determinar cuáles son las principales consideraciones a tener en cuenta en la prevención de daño y a aplicar estas consideraciones de manera transversal al proceso de creación e implementación de algoritmos o redes neuronales.

La prevención de daño debe constituirse como el primer deber a tener en cuenta ante una divergencia entre la opinión del tratante y el SSDC, especialmente si se observa el contexto complejo en que muchas decisiones clínicas se toman frente a la falta de evidencia concluyente (Verdicchio y Perin, 2022, p. 9); entonces, a mayor grado de fiabilidad y de transparencia del algoritmo mayor sería la deferencia del tratante hacia el resultado obtenido por su aplicación. Sin embargo, mientras menor sean la confianza en el sistema y la certeza de que la decisión pueda estar libre de sesgos o pueda ser explicable, es posible que el tratante pueda exponer una adecuada justificación para dejar de ser deferente ante la decisión proporcionada por el SSDC. No obstante, las posibilidades de sesgo del tratante para acomodar su opinión a la del SSDC o utilizar esta opinión como mecanismo de medicina defensiva pueden estar presentes y es necesario minimizarlas.

La incorporación de un SSDC en la toma de decisiones clínicas simplifica el procesamiento de la información adaptada al caso y minimiza las posibilidades de errores que provienen de la incompletitud o imprecisión de la información. Sin embargo, la evaluación del tratante se complejiza al concentrarse tanto en la decisión intrínseca como en la extrínseca referente a la adaptación del modelo a la decisión específica (Perin, 2019, p. 11). Lo señalado requiere profundización con respecto a dos funciones del tratante

como son la interpretación de la decisión y una ética e inclusiva implementación.

En una revisión sistemática de la literatura que busca identificar las implicaciones éticas, sociales y jurídicas de la aplicación de inteligencia artificial mediante SSDC, se observa la preocupación por aproximarse a un diseño ético de los algoritmos para la toma de decisiones en salud a través de planes de cumplimiento de la aplicación de principios que permitan sistemas responsables, como la transparencia y la seguridad. Lo señalado ha llevado a afirmar que el uso de los sistemas no explicables y faltos de la adecuada transparencia deben ser prohibidos en el cuidado de la salud, por cuanto la falta de explicación de funcionamiento del algoritmo conduce a una afectación directa de la comunicación en la relación clínica y, por lo tanto, a la gobernanza misma del cuidado en salud y a la humanización de la relación clínica (Čartolovni et al., 2022; pp. 5-7).

La confianza y la fiabilidad, por lo tanto, constituyen los elementos esenciales que permitirían una mayor amplitud en la implementación de los SSDC. Se ha señalado que en la confianza intervienen factores individuales, características del sistema y factores contextuales que se conectan entre sí para la toma de decisiones; los primeros factores pueden ser las características demográficas, las condiciones de salud o el trato individual hacia el paciente; mientras que los segundos se refieren a la transparencia y configuración de caja negra que reducen las posibilidades de desarrollar confianza en el sistema. Menos estudiadas son las características contextuales, como la cultura del cuidado de la salud, las relaciones interpersonales y la gobernanza (Steerling et al., 2023, p. 06-07).

La confianza y la fiabilidad, por lo tanto, constituyen los elementos esenciales que permitirían una mayor amplitud en la implementación de los SSDC. Se ha señalado que en la confianza intervienen factores individuales, características del sistema y factores contextuales que se conectan entre sí para la toma de decisiones; los primeros factores pueden ser las características demográficas, las condiciones de salud o el trato individual hacia el paciente; mientras que los segundos se refieren a la transparencia y configuración de caja negra que reducen las posibilidades de desarrollar confianza en el sistema. Menos estudiadas son las características contextuales, como la cultura del cuidado de la salud, las relaciones interpersonales y la gobernanza (Steerling et al., 2023, p. 06-07).

Con respecto a la primera, resulta gravitante la opinión de expertos y usuarios del sistema y la percepción de riesgo en el ámbito en que se va a aplicar, lo que ha demostrado el éxito en el procesamiento de imágenes, por ejemplo, pero menor presencia en cirugía. Por otro lado, el fortalecimiento de la comunicación y el diálogo en la relación clínica y sobre la implementación de los SSDC ayuda a cimentar la confianza. Finalmente, la supervisión humana como garantía de

fiabilidad surge como elemento recurrente (Datta Burton et al., 2021, p. 6; Steerling et al., 2023, pp. 1-2).

Se toma como referencia el proyecto de marco normativo más adelantado hasta el momento: el Proyecto del Reglamento del Parlamento Europeo, que establece normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y modifica determinados actos legislativos de la Unión (AI Act). La prevención mediante la verificación de posibles riesgos provocados por las aplicaciones de la inteligencia artificial constituye el enfoque principal.

Sin duda, para la prevención de la responsabilidad de los SSDC, 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 proyecto de 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. Otras medidas establecidas en este capítulo se refieren a la seudonimización o cifrado cuando la anonimización pueda afectar al objeto de la implementación del sistema.

Incorporándose a las herramientas preventivas, además de las señaladas, el artículo 14 aborda la vigilancia humana en los sistemas de alto riesgo que, como hemos dicho, abarcarían los SSDC.

Se señala la obligatoriedad de una interfaz humano-máquina que garantizará la reducción de riesgos. La necesidad de la supervisión humna se ha planteado como un requerimiento necesario como elemento preventivo en el caso de empleo de sistemas de toma de decisiones automatizadas, estableciéndose la referencia del artículo 22 del Reglamento Europeo de Protección de Datos (Viollier y Fischer, 2023, p. 155); en efecto, se señala en dicha norma que los sujetos tienen el derecho a no ser objeto de una decisión basada únicamente en el tratamiento automatizado de datos que produzca efectos jurídicos en él o le afecte significativamente de modo similar.

Adicionalmente, en los casos excepcionales que se permite la toma de decisiones basada exclusivamente en el tratamiento automatizado de datos personales, se consagra el derecho del titular a obtener una intervención humana por parte del controlador, expresar su punto de vista y la posibilidad de impugnar la decisión (Viollier y Fischer, 2023, p. 156).

La manera en que se plantea el funcionamiento de la vigilancia humana como parte de la equidad algorítmica puede ser de gran utilidad para los SSDC, especialmente si los controladores y supervisores pueden fiscalizar el sistema con una herramienta de las mismas capacidades que la implementada pero que se concentre en entender sus limitaciones y capacidades, si pueden, además, ser concientes de la tendencia a confiar en exceso en los resultados del

sistema, es decir, de los sesgos de automatización, lo cual es fundamental para la implementación de los SSDC. Aún más importante es la posibilidad de interpretar la información emanada del sistema, no utilizar, desestimar o hasta revertir la información generada por el mismo.

La consagración normativa de estos mecanismos permitiría que los tratantes puedan tener mayor seguridad con respecto a la no deferencia con los resultados emanados por el SSDC, de tal forma que se mantenga como una herramienta auxiliar o, en el caso en que la fiabilidad y la transparencia lo permitan, como parte de un cuerpo colectivo que tome la decisión final.

Finalmente, cabe señalar que la norma en análisis además incluye obligaciones específicas a los fabricantes, a los distribuidores y a los importadores, obligaciones que se manifiestan, especialmente, en los deberes de información, de presentación de documentación técnica, de instrucciones de uso, de transparencia de la información, de claridad de las instrucciones de uso —asemejánse esta obligación a la que conllevan los productos riesgosos que pueden presentar defectos —, así como el registro de sus sistemas en una base de datos de la UE (Añón Roig, 2022, pp. 32-49; Navas, 2021, pp. 43-50; 2022, p. 18-23).

Conclusiones

La discriminación algorítmica se presenta cuando los sistemas de inteligencia artificial de ML o DL perpetúan inequidades debido a sesgos originados en la etapa de recolección, limpieza o depuración de datos, así como en la de diseño y entrenamiento del sistema mediante los mecanismos de ensayo.Las inequidades que se perpetúan y que se reproducen por el funcionamiento sesgado no toman en cuenta el contexto ni las especificidades de los casos concretos de grupos subrepresentados que son susceptibles a una mayor vulnerabilidad. En la aplicación de los SSDC en el cuidado de la salud, la fiabilidad es un elemento esencial, por lo que es necesario incorporar una acción positiva para la erradicación de sesgos tanto en la recolección de los datos como en el diseño del algoritmo y en la implementación de la decisión, todo ello, mediante los mecanismos adecuados de supervisión humana.En los SSDC, la principal ventaja proporcionada por el ML o el DL es la precisión y la solidez como criterio auxiliar para la toma de decisiones sobre el diagnóstico, el tratamiento o la alternativa terapéutica. Sin embargo, el potencial para provocar daños proviene de la mala calidad de los datos o de la construcción defectuosa del algoritmo al no tomar en cuenta las especificidades de cada caso y los valores, deseos y preferencias de paciente.La equidad algorítmica y la inclusión de la diversidad aparecen como mecanismos transversales al proceso de configuración de SSDC, la primera, entendida como mecanismo transversal a la construcción e

implementación de sistemas focalizados en la erradicación de sesgos; y la segunda entendida como la incorporación del contexto, de los valores y de las preferencias de los pacientes en el caso específico, evitando de esta manera que grupos subrepresentados sean excluidos injustificadamente.Para evitar la aparición de sesgos que perpetúen inequidades, la AI Act ha previsto mecanismos preventivos frente a los sistemas que ha catalogado como altamente riesgosos, entre ellos, la supervisión humana, la entrega de información e instrucciones de uso, y el sistema de gestión de riesgos.

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), 100702. https://doi.org/10.1016/j.hlpt.2022.100702

Ahuja, A. S. (2019). The impact of artificial intelligence in medicine on the future role of the physician. *PeerJ*, *7*, e7702. https://doi.org/10.7717/peerj.7702

Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., Aldairem, A., Alrashed, M., Bin Saleh, K., Badreldin, H. A., Al Yami, M. S., Al Harbi, S., y Albekairy, A. M. (2023). Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Medical Education*, *23*(1), 689. https://doi.org/10.1186/s12909-023-04698-z

Amnistía Internacional. (25 de octubre de 2021). *Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal*. Amnesty International. https://www.amnesty.org/en/documents/eur35/4686/2021/en/

Anderson, M., y Anderson, S. L. (2019). How Should AI Be Developed, Validated, and Implemented in Patient Care? *AMA Journal of Ethics*, *21*(2), E125-130. https://doi.org/10.1001/amajethics.2019.125

Añón Roig, M. J. (2022). Desigualdades algorítmicas. *DERECHOS Y LIBERTADES: Revista de Filosofía Del Derecho y Derechos Humanos*, *47*, 17–49. https://doi.org/10.20318/dyl.2022.6872

Araya, C. (2020). Desafíos legales de la inteligencia artificial en Chile. *Revista Chilena de Derecho y Tecnología*, *9*(2), 257. https://doi.org/10.5354/0719-2584.2020.54489

Barda, N., Yona, G., Rothblum, G. N., 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*, *28*(3), 549–558. https://doi.org/10.1093/jamia/ocaa283

Bedecarratz Scholtz, F., y Aravena Flores, M. (2023). Principios y directrices sobre inteligencia artificial. In M. (Coord.) Azuaje Pirela (Ed.), *Introducción a la ética y el derecho de la Inteligencia Artificial*. Wolters Kluwer España. https://elibro-net.utalca.idm.oclc.org/es/ereader/utalca/229859?page=210.

Boden, M. (2018). *Artificial Intelligence. A very short introduction*. Oxford University Press.

Braun, M., Hummel, P., Beck, S., y Dabrock, P. (2020). Primer on an ethics of AI-based decision support systems in the clinic. *J. Med. Ethics*, *47*(3), 1-8.

Čartolovni, A., Tomičić, A., y Lazić Mosler, E. (2022). Ethical, legal, and social considerations of AI-based medical decision-support tools: A scoping review. *International Journal of Medical Informatics*, *161*, 104738. https://doi.org/10.1016/j.ijmedinf.2022.104738

Chen, I. Y., Pierson, E., Rose, S., Joshi, S., Ferryman, K., y Ghassemi, M. (2021). Ethical Machine Learning in Healthcare. *Annual Review of Biomedical Data Science*, *4*(1), 123–144. https://doi.org/10.1146/annurev-biodatasci-092820-114757

Cirillo, D., Catuara-Solarz, S., Morey, C., Guney, E., Subirats, L., Mellino, S., Gigante, A., Valencia, A., Rementeria, M. J., Chadha, A. S., y Mavridis, N. (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. *Npj Digital Medicine*, *3*(1), 81. https://doi.org/10.1038/s41746-020-0288-5

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), 301. https://doi.org/10.5354/0719-2584.2021.61034

Comisión Europea. (2020). *Libro Blanco sobre la inteligencia artificial- un enfoque europeo orientado a la excelencia y la confianza*. Parlamento Europeo. https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_ES.html

Datta Burton, S., Mahfoud, T., Aicardi, C., y Rose, N. (2021). Clinical translation of computational brain models: understanding the salience of trust in clinician–researcher relationships. *Interdisciplinary Science Reviews*, *46*(1–2), 138–157. https://doi.org/10.1080/03080188.2020.1840223

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), 165–174. https://doi.org/10.1136/jme-2022-108371

Fletcher, R. R., Nakeshimana, A., y Olubeko, O. (2021). Addressing Fairness, Bias, and Appropriate Use of Artificial Intelligence and Machine Learning in Global Health. *Frontiers in Artificial Intelligence*, *3*. https://doi.org/10.3389/frai.2020.561802

Fosch-Villaronga, E., Drukarch, H., Khanna, P., Verhoef, T., y Custers, B. (2022). Accounting for diversity in AI for medicine. *Computer Law y Security Review*, *47*, 105735. https://doi.org/10.1016/j.clsr.2022.105735

FundéuRAE. (2022, December 19). *Inteligencia artificial es la expresión del 2022 para la FundéuRAE*. FundéuRAE. https://www.fundeu.es/recomendacion/inteligencia-artificial-es-la-expresion-del-2022-para-la-fundeurae/

Grote, T., y Berens, P. (2020). On the ethics of algorithmic decision-making in healthcare. *Journal of Medical Ethics*, *46*(3), 205–211. https://doi.org/10.1136/medethics-2019-105586

Grote, T., y Keeling, G. (2022). On Algorithmic Fairness in Medical Practice. *Cambridge Quarterly of Healthcare Ethics*, *31*(1), 83–94. https://doi.org/10.1017/S0963180121000839

Hegde, P. R., y Shenoy, M. M. (2021). Artificial Intelligence in Medicine and Health Sciences. *Archives of Medicine and Health Sciences*, *9*(1), 145–150. https://doi.org/10.4103/amhs.amhs_315_20

Holmes, D. (2017). *Big Data. A very short introduction*. Oxford University Press.

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), e36388. https://doi.org/10.2196/36388

Johnson, S. (2019). AI, Machine Learning, and Ethics in Health Care. *Journal of Legal Medicine*, *39*(4), 427–441. https://doi.org/10.1080/01947648.2019.1690604

Jones, C., Thornton, J., y Wyatt, J. C. (2023). Artificial intelligence and clinical decision support: clinicians' perspectives on trust, trustworthiness, and liability. *Medical Law Review*, *31*(4), 501–520. https://doi.org/10.1093/medlaw/fwad013

Karimian, G., Petelos, E., y Evers, S. M. A. A. (2022). The ethical issues of the application of artificial intelligence in healthcare: a systematic scoping review. *AI and Ethics*, *2*(4), 539–551. https://doi.org/10.1007/s43681-021-00131-7

Kumar, A., Aelgani, V., Vohra, R., Gupta, S. K., Bhagawati, M., Paul, S., Saba, L., Suri, N., Khanna, N. N., Laird, J. R., Johri, A. M., Kalra, M., Fouda, M. M., Fatemi, M., Naidu, S., y Suri, J. S. (2023). Artificial intelligence bias in medical system designs: a systematic review. *Multimedia Tools and Applications*, *83*(6), 18005–18057. https://doi.org/10.1007/s11042-023-16029-x

Kumar, P., Chauhan, S., y Awasthi, L. K. (2023). Artificial Intelligence in Healthcare: Review, Ethics, Trust Challenges yamp; Future Research Directions. *Engineering Applications of Artificial Intelligence*, *120*, 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), 207. https://doi.org/10.5354/0719-2584.2020.56843

Ledford, H. (2019). Millions of black people affected by racial bias in health-care algorithms. *Nature*, *574*(7780), 608–609. https://doi.org/10.1038/d41586-019-03228-6

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*, *33*(3), 129–135. https://doi.org/10.1080/10400435.2019.1593259

Liyanage, H., Liaw, S.-T., Jonnagaddala, J., Schreiber, R., Kuziemsky, C., Terry, A. L., y de Lusignan, S. (2019). Artificial Intelligence in Primary Health Care: Perceptions, Issues, and Challenges. *Yearbook of Medical Informatics*, *28*(01), 041–046. https://doi.org/10.1055/s-0039-1677901

Madrid, R. (2023). Las máquinas y la agencia moral. In M. (Coord.) Azuaje Pirela (Ed.), *Introducción a la ética y el derecho de la Inteligencia Artificial*. Wolters Kluwer España.

Martín-Casals, M. (2022a). An approach to some EU initiatives on the regulation of liability for damage caused by AI-Systems. *Ius et Praxis*, *28*(2), 3–24. https://doi.org/10.4067/S0718-00122022000200003

Martín-Casals, M. (2022b). Desarrollo tecnológico y responsabilidad extracontractual. A propósito de los sistemas de inteligencia artificial. In J. Pérez y F. SanJuan (Eds.), *La Cultura Jurídica en la era digital, Cizur Menor, Aranzadi* (pp. 101-138.).

Mijwil, M., y Aggarwal, K. (2022). A diagnostic testing for people with appendicitis using machine learning techniques. *Multimedia Tools and Applications*, *81*(5), 7011–7023. https://doi.org/10.1007/s11042-022-11939-8

Molnár-Gábor, F., y Giesecke, J. (2022). Medical AI Key Elements at the International Level. In P. H. Silja Voeneky, M. O. Oliver, y W. Burgard (Eds.), *The Cambridge Handbook of Responsible Artificial Intelligence*. Cambridge University Press.

Naik, N., Hameed, B. M. Z., Shetty, D. K., Swain, D., Shah, M., Paul, R., Aggarwal, K., Ibrahim, S., Patil, V., Smriti, K., Shetty, S., Rai, B. P., Chlosta, P., y Somani, B. K. (2022). Legal and Ethical Consideration in Artificial Intelligence in Healthcare: Who Takes Responsibility? *Frontiers in Surgery*, *9*. https://doi.org/10.3389/fsurg.2022.862322

Navas, S. (2021). Salud Electrónica e Inteligencia Artificial. In S. (Dir.) Navas (Ed.), *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.

Parlamento Europeo. (2020). *Régimen de Responsabilidad Civil en materia de inteligencia artificial, Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un régimen de responsabilidad civil en materia de inteligencia artificial (2020/2014(INL)*. Parlamento Europeo. https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_ES.html

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), 3. https://doi.org/10.5354/0719-2584.2019.52560

Pethig, F., y Kroenung, J. (2023). Biased Humans, (Un)Biased Algorithms? *Journal of Business Ethics*, *183*(3), 637–652. https://doi.org/10.1007/s10551-022-05071-8

Pfohl, S. R., Foryciarz, A., y Shah, N. H. (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

Prictor, 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), 1–24. https://doi.org/10.1093/medlaw/fwac022

Ramón Fernández, F. (2021). Inteligencia artificial en la relación médico-paciente: Algunas cuestiones y propuestas de mejora. *Revista Chilena de Derecho y Tecnología*, *10*(1), 329. https://doi.org/10.5354/0719-2584.2021.60931

Silcox, C. (2020). *La inteligencia artificial en el sector salud: Promesas y desafíos*. Inter-American Development Bank. https://doi.org/10.18235/0002845

Smith, H., y Fotheringham, K. (2020). Artificial intelligence in clinical decision-making: Rethinking liability. *Medical Law International*, *20*(2), 131–154. https://doi.org/10.1177/0968533220945766

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

Sunarti, S., Fadzlul Rahman, F., Naufal, M., Risky, M., Febriyanto, K., y Masnina, R. (2021). Artificial intelligence in healthcare: opportunities and risk for future. *Gaceta Sanitaria*, *35*, S67–S70. https://doi.org/10.1016/j.gaceta.2020.12.019

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), 520–528. https://doi.org/10.1111/jep.13541

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

Viollier, P., y Fischer, E. (2023). La intervención humana como resguardo ante la toma automatizada de decisiones. Implicancias éticas y jurídicas. In M. (Coord.) Azuaje Pirela (Ed.), *Introducción a la ética y el derecho de la Inteligencia Artificial*. Wolters Kluwer España.

Walker, N. (2023). El problema de sesgo algorítmico y, en particular, el sesgo de género. In C. Droguett y N. Walker (Eds.), *Derecho Digital y Privacidad en América y Europa* (pp. 148-161.). Tirant Lo Blanch.

Xu, Q., Xie, W., Liao, B., Hu, C., Qin, L., Yang, Z., Xiong, H., Lyu, Y., Zhou, Y., y Luo, A. (2023). Interpretability of Clinical Decision Support Systems Based on Artificial Intelligence from Technological and Medical Perspective: A Systematic Review. *Journal of Healthcare Engineering*, *2023*(1). https://doi.org/10.1155/2023/9919269

Notas

[2]

La base de datos PUBMED, especializada en la investigación en el ámbito de la salud y la medicina, arroja 55.880 resultados cuando se incorpora el término “inteligencia artificial” a su motor de búsqueda, colocando solamente un rango temporal desde el 2018 al 2023. Por otro lado, en la base de datos SCOPUS existen 284 revistas científicas especializadas exclusivamente en este tema, como se puede constatar en https://www.scimagojr.com/journalrank.php?category=1702.

Información adicional

redalyc-journal-id: 6697

Edison Calahorrano Latorre

Equidad algorítmica y decisiones clínicas basadas en sistemas de soporte basados en inteligencia artificial
Algorithmic fairness and clinical decisions based on artificial intelligence-based support systems

Nuevo Derecho
vol. 21, núm. 36, p. 1 - 17, 2025
Institución Universitaria de Envigado, Colombia
nuevo.derecho@iue.edu.co

ISSN: 2011-4540
ISSN-E: 2500-672X

DOI: https://doi.org/10.25057/2500672X.1691