

11

BIOÉTICA

Y SIMULACIONES EN LA INTELIGENCIA ARTIFICIAL: TOMA DE DECISIONES EN MEDICINA COMPUTACIONAL

BIOETHICS AND SIMULATIONS IN ARTIFICIAL INTELLIGENCE: DECISION MAKING IN COMPUTATIONAL MEDICINE

Lorraine María Frutos-Morales¹

E-mail: frutosmoralesl@gmail.com

ORCID: <https://orcid.org/0000-0003-4305-2751>

María Caridad González-Alonso¹

E-mail: macaridad62@gmail.com

ORCID: <https://orcid.org/0000-0002-2250-837X>

Marisela Angelis Rivero-Fernández¹

E-mail: mariselaangelisrivero@gmail.com

ORCID: <https://orcid.org/0009-0003-0705-9873>

¹ Universidad de Ciencias Médicas de La Habana. Cuba.

Cita sugerida (APA, séptima edición)

Frutos-Morales, L. M., González-Alonso, M. C., & Rivero-Fernández, M. A. (2026). Bioética y simulaciones en la inteligencia artificial: toma de decisiones en medicina computacional. *Revista Transdisciplinaria de Estudios Sociales y Tecnológicos*, 6(3), 117-134.

Presentación: 04/06/2026

Aceptación: 23/07/2026

Publicación: 01/09/2026



RESUMEN

La vertiginosa integración de la Inteligencia Artificial aplicada a la medicina ha generado desafíos éticos por la falta de transparencia de las herramientas de este tipo de inteligencia. La naturaleza algorítmica y probabilística a escala de estos problemas requiere enfoques cuantitativos que permitan evaluar las consecuencias de las decisiones éticas. El objetivo de este estudio es proponer un marco de evaluación de impacto ético diseñado para sistemas de IA utilizados en medicina, que proporcione una guía estructurada que permita analizar y mitigar los posibles efectos éticos asociados con la integración de estas tecnologías en la práctica clínica. Se realizó una revisión sistemática de literatura científica (2019-2024) en las bases de datos PubMed, Scopus, Philosopher's Index, IEEE Xplore y Google Scholar. Se identifican tres dominios críticos donde la Bioética Cuantitativa resulta esencial, se revela que las simulaciones basadas en agentes, los modelos de optimización multiobjetivo y los sistemas de análisis de impacto ético permiten cuantificar trade-offs entre principios éticos en conflicto y prever consecuencias no intuitivas de políticas sanitarias. La Bioética Cuantitativa representa un paradigma necesario para abordar la complejidad de los dilemas éticos generados por la Inteligencia Artificial en medicina como herramienta práctica para instituciones sanitarias.

Palabras clave:

Algoritmos, bioética, ética médica, inteligencia artificial, simulaciones computacionales.

ABSTRACT

The dizzying integration of Artificial Intelligence applied to medicine has generated ethical challenges due to the lack of transparency of AI tools. The algorithmic and probabilistic nature on the scale of these problems requires quantitative approaches that allow evaluating the consequences of ethical decisions. The study analyzes a methodological framework based on computational simulations from the state of Quantitative Bioethics for the evaluation of bioethical dilemmas in the era of AI in medicine. A systematic review of scientific literature (2019-2024) was conducted in the databases PubMed, Scopus, Philosopher's Index, IEEE Xplore, and Google Scholar. Three critical domains are identified where Quantitative Bioethics is essential; it is revealed that agent-based simulations, multi-objective optimization models, and ethical impact analysis systems allow quantifying trade-offs between conflicting ethical principles and anticipating non-intuitive consequences of health policies. Quantitative Bioethics represents a necessary paradigm to address the complexity of ethical dilemmas generated by AI in medicine as a practical tool for healthcare institutions.

Keywords:

Algorithms, artificial intelligence, bioethics, computational simulations, medical ethics.

INTRODUCCIÓN

La integración de sistemas de inteligencia artificial en la práctica médica constituye una de las transformaciones más relevantes de la atención sanitaria contemporánea. Su incorporación progresiva en procesos de diagnóstico, pronóstico, selección de tratamientos, monitorización de pacientes, gestión hospitalaria y asignación de recursos ha ampliado considerablemente las capacidades de procesamiento y análisis de información clínica. Sin embargo, esta transformación también ha generado una antinomia que trasciende el ámbito estrictamente tecnológico para involucrar cuestiones éticas, profesionales, jurídicas y sociales. La capacidad de los sistemas algorítmicos para participar directa o indirectamente en decisiones relacionadas con la salud obliga a reconsiderar los mecanismos mediante los cuales se evalúan la responsabilidad, la transparencia, la justicia y las consecuencias de las decisiones clínicas.

Los marcos bioéticos tradicionales, centrados en principios como la autonomía, la beneficencia, la no maleficencia y la justicia (Beauchamp & Childress, 2019), han demostrado ser fundamentales para analizar dilemas interpersonales y orientar la toma de decisiones en contextos asistenciales. No obstante, pueden resultar insuficientes cuando deben aplicarse a sistemas caracterizados por una naturaleza algorítmica, probabilística, dinámica y susceptible de operar a gran escala (Char et al., 2018). Mientras que numerosos dilemas bioéticos tradicionales se estructuran alrededor de una relación identificable entre profesionales sanitarios y pacientes, los sistemas de inteligencia artificial pueden intervenir simultáneamente sobre grandes poblaciones, procesar enormes cantidades de información y producir recomendaciones cuyos efectos se distribuyen entre múltiples actores y niveles institucionales.

Esta transformación modifica también la escala del razonamiento ético. La medicina contemporánea enfrenta escenarios en los que las decisiones deben adoptarse considerando millones de puntos de datos, asociaciones estadísticas, complejas cadenas causales y consecuencias poblacionales que pueden escapar a la intuición humana. Una decisión algorítmica aparentemente neutra puede producir efectos diferentes dependiendo de las características demográficas, sociales, económicas o clínicas de las poblaciones sobre las que se aplica. Por esta razón, la evaluación ética de estos sistemas no puede limitarse exclusivamente a determinar si su finalidad es beneficiosa, sino que debe analizar cómo funcionan, qué información utilizan, sobre qué poblaciones fueron desarrollados, qué errores pueden producir y cómo se distribuyen sus beneficios y riesgos.

La magnitud de este desafío se evidencia en casos documentados de sesgos algorítmicos capaces de afectar

desproporcionadamente a determinados grupos poblacionales (Obermeyer et al., 2019). El problema adquiere especial importancia cuando los datos utilizados para desarrollar o validar un modelo no representan adecuadamente la diversidad de la población en la que posteriormente será utilizado. En estas circunstancias, un sistema puede alcanzar indicadores globales aparentemente satisfactorios y, simultáneamente, presentar un rendimiento inferior en determinados grupos. Desde una perspectiva bioética, esta situación obliga a superar la evaluación exclusivamente técnica de la precisión para incorporar criterios relacionados con la equidad, la representatividad y la distribución de los posibles errores.

Otro problema fundamental corresponde a la opacidad de determinados modelos algorítmicos, particularmente aquellos cuyo funcionamiento interno resulta difícil de interpretar incluso para profesionales especializados. Los denominados modelos de caja negra plantean interrogantes sobre la rendición de cuentas cuando una recomendación algorítmica influye en una decisión clínica y posteriormente se produce un resultado adverso (Price, 2019). En tales circunstancias resulta necesario determinar quién debe explicar la decisión, qué grado de comprensión debe poseer el profesional sanitario sobre el funcionamiento del sistema y hasta qué punto una institución puede delegar parte del proceso decisorio en una tecnología cuya lógica interna no es plenamente interpretable.

La transparencia, por consiguiente, no debe entenderse únicamente como la posibilidad técnica de acceder al código o conocer la arquitectura de un modelo. Desde una perspectiva ética, también comprende la capacidad de explicar de manera comprensible los factores que influyeron en una recomendación, identificar las limitaciones del sistema y establecer procedimientos para cuestionar o revisar sus resultados. La existencia de mecanismos de supervisión humana adquiere especial importancia cuando las recomendaciones algorítmicas pueden afectar decisiones relacionadas con diagnósticos, tratamientos, priorización asistencial o acceso a determinados recursos sanitarios.

Los dilemas se intensifican cuando los sistemas de inteligencia artificial intervienen en la distribución de recursos escasos. Los algoritmos de priorización pueden utilizarse para ordenar listas de espera, identificar pacientes con mayor riesgo, distribuir recursos hospitalarios o determinar qué personas requieren una intervención prioritaria. Estas aplicaciones pueden mejorar la eficiencia organizativa, pero también incorporan decisiones normativas sobre qué variables deben valorarse y qué resultados deben considerarse prioritarios. La asignación de recursos sanitarios implica necesariamente decisiones relacionadas con equidad, necesidad, beneficio esperado y protección de poblaciones vulnerables (Emanuel et al., 2020). Cuando estas decisiones son parcialmente

automatizadas, los criterios éticos deben incorporarse explícitamente al diseño, validación y supervisión del sistema.

Estos problemas comparten una característica fundamental: su análisis requiere no solo reflexión filosófica, sino también modelado matemático y evaluación cuantitativa de sus consecuencias. La ética aplicada a sistemas algorítmicos necesita herramientas capaces de determinar cómo diferentes decisiones modifican la distribución de beneficios y riesgos entre grupos poblacionales. Esto supone evaluar no solamente qué decisión parece éticamente justificable en términos abstractos, sino también qué consecuencias previsibles genera cuando se aplica de manera reiterada sobre cientos, miles o millones de personas.

La transformación tecnológica de la atención sanitaria también repercute directamente en las funciones de los profesionales encargados del cuidado. En el ámbito de la enfermería, las nuevas tecnologías ofrecen oportunidades para mejorar la calidad asistencial, fortalecer el seguimiento del paciente, facilitar el acceso a información relevante y optimizar determinados procesos vinculados con el cuidado (Albán-Sabando, 2024). Estas posibilidades muestran que la innovación tecnológica puede contribuir a una atención más eficiente y oportuna cuando se integra adecuadamente en las dinámicas profesionales.

Sin embargo, la incorporación tecnológica no debe concebirse como un proceso exclusivamente instrumental. El cuidado sanitario comprende dimensiones humanas, comunicativas, emocionales y relacionales que no pueden reducirse a la eficiencia del procesamiento de información. En este sentido, las innovaciones tecnológicas utilizadas en la práctica enfermera deben integrarse de manera que fortalezcan, y no sustituyan, los componentes humanos del cuidado (Albán-Sabando, 2024). Esta consideración adquiere una relevancia particular en el contexto de la inteligencia artificial, debido a que una mayor capacidad de automatización no implica necesariamente una mejor calidad ética de las decisiones.

La relación entre innovación tecnológica y cuidado permite, además, ampliar el análisis bioético más allá de los resultados clínicos inmediatos. Una herramienta puede mejorar determinados indicadores de eficiencia y, al mismo tiempo, incrementar la carga tecnológica de los profesionales, modificar la comunicación con los pacientes o generar nuevas formas de dependencia respecto de los sistemas automatizados. Por ello, la evaluación ética debe considerar tanto los resultados cuantificables como las transformaciones que la tecnología introduce en las relaciones entre pacientes, profesionales, instituciones y sistemas de información.

Los enfoques tradicionales para el análisis conceptual presentan limitaciones significativas cuando se enfrentan

a dilemas propios de la era de la inteligencia artificial. Una primera limitación corresponde a la abstracción excesiva, debido a que los principios generales ofrecen una orientación operativa limitada para implementar sistemas algorítmicos específicos (Juca et al., 2025; Mittelstadt, 2019). Reconocer que un sistema debe ser justo, transparente o beneficioso constituye un punto de partida necesario, pero no determina automáticamente qué métricas deben emplearse, qué nivel de error resulta aceptable o cómo deben resolverse situaciones en las que dos principios entran en conflicto.

Una segunda dificultad reside en la limitada capacidad de los enfoques exclusivamente conceptuales para cuantificar las compensaciones entre principios éticos en conflicto. En determinadas aplicaciones sanitarias, aumentar la sensibilidad de un modelo puede incrementar los falsos positivos, mientras que elevar su especificidad puede producir el efecto contrario. De manera semejante, maximizar la eficiencia en la distribución de determinados recursos puede entrar en tensión con objetivos de equidad. Los marcos tradicionales no siempre proporcionan metodologías suficientes para medir y comparar estas compensaciones de manera sistemática (McDougall, 2019).

Una tercera limitación corresponde al predominio de perspectivas retrospectivas. Con frecuencia, los problemas éticos adquieren visibilidad después de que una tecnología ha sido implementada y sus efectos adversos han comenzado a manifestarse. Esta lógica reactiva resulta especialmente problemática en sistemas algorítmicos capaces de operar a gran escala. Por ello, resulta necesario desarrollar instrumentos que permitan identificar riesgos antes de la implementación, simular escenarios potenciales y estimar la magnitud y distribución de las consecuencias previsibles (Morley et al., 2020).

A estas dificultades se añade la naturaleza dinámica de algunos sistemas de inteligencia artificial. Los cambios en las poblaciones, en las prácticas clínicas, en la disponibilidad de datos o en las condiciones institucionales pueden modificar el rendimiento de un modelo a lo largo del tiempo. En consecuencia, una evaluación ética realizada exclusivamente antes de la implementación no garantiza que el sistema continúe siendo adecuado posteriormente. La supervisión ética debe concebirse como un proceso continuo que acompañe el ciclo de vida de la tecnología y permita detectar cambios relevantes en su comportamiento y en sus efectos sobre diferentes grupos.

Estas limitaciones han contribuido a generar una brecha de implementación ética, en la que los desarrolladores de sistemas de inteligencia artificial pueden reconocer la importancia de principios como justicia, transparencia, autonomía o no maleficencia, pero carecer de herramientas prácticas suficientes para traducirlos en decisiones concretas de diseño y evaluación algorítmica (Jobin et

al., 2019). La existencia de declaraciones éticas, códigos de conducta y principios generales representa un avance importante, aunque su utilidad disminuye cuando no existen mecanismos que permitan verificar empíricamente su cumplimiento.

En el ámbito sanitario, esta brecha adquiere especial trascendencia porque los efectos de una decisión algorítmica pueden repercutir directamente sobre la salud, el bienestar y las oportunidades de atención de las personas. Por ello, la evaluación ética necesita indicadores capaces de determinar no solamente si un sistema funciona, sino también para quién funciona, en qué condiciones, con qué nivel de incertidumbre y qué grupos soportan una mayor proporción de sus posibles errores. La innovación tecnológica aplicada al cuidado refuerza precisamente la necesidad de que la adopción de nuevas herramientas permanezca subordinada a objetivos relacionados con la calidad de la atención y el bienestar de los pacientes (Albán-Sabando, 2024).

En respuesta a estas limitaciones emerge la bioética cuantitativa como una perspectiva que incorpora métodos matemáticos, estadísticos y computacionales al análisis de problemas éticos relacionados con la atención sanitaria (Battin, 2023). Su importancia reside en la posibilidad de complementar la deliberación normativa con procedimientos capaces de representar escenarios complejos, estimar consecuencias y comparar diferentes alternativas mediante información cuantificable. Esta aproximación permite trasladar determinados problemas éticos desde formulaciones exclusivamente conceptuales hacia modelos que hagan explícitas las variables, supuestos y consecuencias involucradas.

La bioética cuantitativa no implica reducir la ética a números ni asumir que todos los valores humanos pueden expresarse mediante una métrica. Su contribución consiste en utilizar herramientas cuantitativas para hacer visibles dimensiones que podrían permanecer ocultas en un análisis exclusivamente cualitativo. Por ejemplo, puede permitir determinar si una determinada tasa de error se distribuye de manera desigual entre poblaciones, estimar cuántas personas podrían resultar afectadas por una decisión automatizada o comparar los efectos previsibles de diferentes criterios de priorización.

Desde esta perspectiva, la cuantificación puede convertirse en un instrumento para fortalecer la deliberación ética. Los resultados matemáticos no sustituyen la decisión normativa sobre qué valores deben protegerse, pero pueden proporcionar evidencia sobre las consecuencias de elegir una alternativa determinada. De esta manera, la deliberación bioética puede combinar principios normativos con análisis empírico y modelado prospectivo, generando procesos de evaluación más transparentes y reproducibles.

Esta perspectiva permite también desarrollar escenarios de simulación antes de que una tecnología sea implementada en condiciones reales. La posibilidad de estimar cómo diferentes grupos poblacionales podrían verse afectados constituye una herramienta particularmente relevante para la prevención de daños. Un sistema de inteligencia artificial destinado a apoyar decisiones clínicas podría evaluarse no solamente mediante indicadores generales de precisión, sino también mediante escenarios que permitan identificar desigualdades potenciales, niveles de incertidumbre y consecuencias derivadas de diferentes umbrales de decisión.

La evaluación prospectiva adquiere así una función preventiva. En lugar de esperar a que los daños aparezcan para iniciar una discusión ética, el análisis puede incorporarse desde las primeras etapas del diseño tecnológico. Esto permite que desarrolladores, profesionales sanitarios, instituciones y especialistas en bioética participen en la identificación anticipada de riesgos y en la definición de mecanismos de mitigación.

La participación de los profesionales sanitarios resulta indispensable en este proceso. La innovación tecnológica aplicada a la enfermería demuestra que las herramientas digitales adquieren sentido cuando responden a necesidades concretas del cuidado y se integran adecuadamente en la práctica profesional (Albán-Sabando, 2024). De manera semejante, los sistemas de inteligencia artificial utilizados en medicina deben diseñarse y evaluarse considerando el conocimiento de quienes interactúan directamente con los pacientes y comprenden las particularidades del contexto clínico.

La evaluación ética tampoco puede limitarse al momento inicial de adquisición o implementación de una tecnología. Debe acompañar todo su ciclo de vida, desde la selección de datos y el desarrollo del modelo hasta su validación, implementación, supervisión, actualización y eventual retirada. Cada una de estas etapas puede introducir riesgos diferentes y modificar la relación entre beneficios y posibles daños. Por esta razón, un marco de evaluación de impacto ético debe incorporar mecanismos de seguimiento que permitan detectar desviaciones y reevaluar periódicamente la pertinencia del sistema.

Otro componente fundamental corresponde a la responsabilidad. La utilización de inteligencia artificial no elimina la responsabilidad de los actores humanos e institucionales involucrados en la atención sanitaria. Por el contrario, puede distribuirla entre desarrolladores, proveedores tecnológicos, profesionales, administradores e instituciones. Un marco ético adecuado debe identificar con claridad quién supervisa el funcionamiento del sistema, quién responde ante errores previsibles, quién tiene capacidad para suspender su utilización y qué procedimientos

existen para que pacientes y profesionales cuestionen una decisión apoyada algorítmicamente.

También resulta necesario preservar la autonomía del paciente. La incorporación de sistemas inteligentes puede modificar la forma en que se presenta la información clínica y aumentar la influencia de las recomendaciones automatizadas sobre las decisiones profesionales. En consecuencia, la utilización de estas tecnologías debe evitar que la autoridad atribuida al algoritmo reduzca las posibilidades de deliberación o convierta una recomendación probabilística en una decisión aparentemente incuestionable. La supervisión humana debe garantizar que las circunstancias particulares del paciente continúen formando parte sustancial del proceso decisorio.

La integración entre bioética tradicional y herramientas cuantitativas ofrece, por tanto, una vía para abordar la complejidad creciente de la medicina contemporánea. Los principios de autonomía, beneficencia, no maleficencia y justicia continúan proporcionando fundamentos normativos esenciales, mientras que los métodos cuantitativos permiten examinar cómo esos principios se expresan en sistemas capaces de operar sobre grandes volúmenes de información y poblaciones heterogéneas. La relación entre ambos enfoques debe entenderse como complementaria y no como sustitutiva.

En este marco, la evaluación del impacto ético puede funcionar como un puente entre los principios bioéticos y las decisiones técnicas. Su propósito consiste en transformar valores generales en dimensiones observables que puedan analizarse durante el diseño, validación e implementación de los sistemas. Esto implica examinar aspectos como la distribución de errores, la representatividad de los datos, la transparencia de las recomendaciones, la posibilidad de supervisión humana, la protección de la autonomía, la responsabilidad institucional y los efectos potenciales sobre poblaciones vulnerables.

El desafío fundamental consiste en garantizar que el desarrollo tecnológico permanezca subordinado a las finalidades humanas de la medicina. La capacidad de procesar grandes cantidades de datos, identificar patrones complejos y automatizar determinados procesos puede mejorar sustancialmente la atención sanitaria, pero estos beneficios no constituyen por sí mismos una justificación ética suficiente. La tecnología adquiere legitimidad cuando sus capacidades se orientan hacia la protección de la salud, la dignidad, la equidad y la calidad del cuidado. Esta concepción resulta consistente con la necesidad de integrar las innovaciones tecnológicas en la práctica sanitaria sin perder de vista el carácter humano y relacional de la atención (Albán-Sabando, 2024).

Por consiguiente, la bioética aplicada a la inteligencia artificial requiere avanzar desde la formulación general de principios hacia mecanismos capaces de anticipar, medir

y supervisar sus efectos en contextos clínicos reales. La combinación de deliberación normativa, análisis cuantitativo, evaluación prospectiva y supervisión continua puede contribuir a reducir la distancia entre los principios éticos declarados y las decisiones incorporadas efectivamente en los sistemas tecnológicos.

El objetivo de este artículo es proponer un marco de evaluación de impacto ético diseñado para sistemas de inteligencia artificial utilizados en medicina, que proporcione una guía estructurada para identificar, analizar, cuantificar y mitigar los posibles efectos éticos asociados con la integración de estas tecnologías en la práctica clínica. El marco pretende contribuir a una evaluación prospectiva y continua que articule los principios fundamentales de la bioética con herramientas cuantitativas aplicables al diseño, implementación y supervisión de sistemas algorítmicos, prestando especial atención a la equidad, la transparencia, la autonomía, la responsabilidad, la seguridad y la protección de las poblaciones potencialmente vulnerables.

MATERIALES Y MÉTODOS

Para la ejecución de este estudio se realizó una revisión crítica de la literatura científica publicada entre 2018 y 2025 en las bases de datos PubMed, Scopus, Philosopher's Index, Institute of Electrical and Electronics Engineers Xplore (IEEE Xplore) y Google Scholar. Para la búsqueda sistemática se utilizaron cadenas de búsqueda que combinaron términos de Medical Subject Headings (MeSH) y palabras clave libres: artificial intelligence, machine learning, deep learning, bioethics, medical ethics, ethical dilemmas, decision making, clinical decision support, computational simulation, agent-based modeling y quantitative methods.

Se aplicaron los siguientes filtros: artículos originales y revisiones publicados entre enero de 2018 y diciembre de 2025, en idioma inglés o español y con texto completo disponible. La selección de los estudios se realizó en tres fases: 1) cribado por títulos y resúmenes; 2) lectura completa de los textos potencialmente relevantes; y 3) extracción de datos mediante un formulario estandarizado que incluía autor y año, dominio de aplicación de la inteligencia artificial (IA), tipo de simulación empleada, principios éticos abordados y métricas cuantitativas utilizadas.

El análisis se estructuró a partir de categorías temáticas emergentes relacionadas con la asignación de recursos, los sesgos algorítmicos, la privacidad de los datos, el diagnóstico asistido y los sistemas autónomos.

RESULTADOS Y DISCUSIÓN

La transición de la ética principalista a la ética computacional implica un cambio paradigmático en el ámbito de la bioética, que avanza desde enfoques predominantemente

cualitativos hacia aproximaciones que incorporan herramientas cuantitativas. Este nuevo enfoque no solo amplía las posibilidades del análisis ético, sino que también integra perspectivas procedentes de la teoría de decisiones, la economía del bienestar y la ciencia de los sistemas complejos (Mittelstadt, 2019). La incorporación de estas herramientas y marcos de referencia permite que la bioética responda a un contexto contemporáneo en el que la toma de decisiones éticas puede beneficiarse de métodos analíticos capaces de representar escenarios complejos, comparar alternativas y estimar las posibles consecuencias de las tecnologías emergentes sobre la salud y el bienestar humano.

Este enfoque reconoce que numerosos dilemas éticos contemporáneos implican procesos de optimización sometidos a múltiples restricciones y a valores que pueden entrar en competencia. Por ejemplo, el diseño de un algoritmo de triaje para unidades de cuidados intensivos puede perseguir simultáneamente la maximización de vidas salvadas, la garantía de equidad en el acceso, el respeto de las preferencias de los pacientes y la utilización eficiente de los recursos disponibles. Estos objetivos pueden entrar en conflicto, por lo que resulta necesario disponer de métodos formales que permitan evaluar las compensaciones involucradas y hacer explícitos los criterios utilizados para priorizar determinados resultados.

La teoría de decisiones multicriterio constituye una de las metodologías que pueden aplicarse en bioética para proporcionar un marco matemático de apoyo a la toma de decisiones cuando existen múltiples objetivos en conflicto. En contextos éticos, permite analizar cuantitativamente el desempeño de diferentes cursos de acción a través de diversas dimensiones relevantes (Doupmpou et al., 2019). Por ejemplo, en la evaluación de algoritmos de diagnóstico es posible considerar simultáneamente la sensibilidad, la especificidad, el coste, la equidad distributiva y la transparencia. De esta manera, la decisión no depende exclusivamente de un indicador de rendimiento, sino de la valoración conjunta de diferentes criterios clínicos, económicos y éticos.

El modelado basado en agentes permite simular sistemas complejos mediante la creación de agentes autónomos que interactúan de acuerdo con reglas previamente definidas. Esta metodología resulta particularmente útil para estudiar cómo determinadas propiedades y consecuencias éticamente relevantes pueden emerger en el nivel colectivo a partir de interacciones producidas en el nivel individual (Razavi et al., 2021). Por ejemplo, es posible simular cómo diferentes políticas de asignación de órganos afectan la equidad en los trasplantes entre distintos grupos demográficos, identificando patrones distributivos que podrían no resultar evidentes mediante un análisis exclusivamente individual.

El análisis de sensibilidad ética permite evaluar cómo se modifican las conclusiones cuando varían determinados supuestos fundamentales o parámetros incorporados en un modelo. Esta aproximación resulta especialmente relevante debido a que numerosos parámetros vinculados con decisiones éticas, como el valor atribuido a determinados resultados sanitarios o la ponderación relativa de la equidad frente a la eficiencia, pueden estar sujetos a desacuerdos legítimos entre individuos, profesionales e instituciones (Saltelli, 2021). El análisis de sensibilidad permite determinar hasta qué punto una recomendación depende de esos supuestos y contribuye a identificar aquellos parámetros cuya modificación puede alterar sustancialmente los resultados obtenidos.

Las técnicas de optimización sometidas a restricciones matemáticas también pueden utilizarse para identificar soluciones que maximicen determinados objetivos éticamente relevantes sin vulnerar condiciones previamente establecidas. Por ejemplo, el problema de asignación de recursos sanitarios durante una pandemia puede formularse mediante un modelo de optimización orientado a maximizar los años de vida ajustados por calidad, garantizando simultáneamente niveles mínimos de acceso para todos los grupos poblacionales (Singh & Rebennack, 2025). Este tipo de aproximación permite incorporar explícitamente restricciones relacionadas con la equidad y evitar que la maximización de la eficiencia constituya el único criterio de decisión.

En conjunto, estas metodologías permiten avanzar desde una formulación exclusivamente declarativa de los principios bioéticos hacia procedimientos capaces de analizar de manera sistemática las consecuencias de diferentes decisiones. La cuantificación no sustituye la deliberación ética ni determina por sí misma qué valores deben prevalecer, sino que proporciona información adicional para hacer explícitos los supuestos, comparar alternativas, identificar posibles desigualdades y anticipar consecuencias. De este modo, la ética computacional puede entenderse como una extensión metodológica de la deliberación bioética tradicional, especialmente pertinente ante sistemas de inteligencia artificial capaces de intervenir en decisiones clínicas complejas y producir efectos sobre grandes poblaciones.

Las simulaciones computacionales se han convertido en una herramienta valiosa para abordar dilemas éticos emergentes, incluidos aquellos relacionados con la asignación de recursos sanitarios escasos. Mediante modelos capaces de representar diferentes escenarios, poblaciones, restricciones y criterios de decisión, es posible analizar las implicaciones éticas derivadas de distintas estrategias de distribución de recursos limitados. Esta aproximación permite evaluar de manera prospectiva cómo determinadas decisiones pueden repercutir sobre

principios fundamentales como la justicia, la equidad, la beneficencia y el bienestar colectivo.

De este modo, las simulaciones no solo permiten representar los desafíos que surgen en contextos de escasez, sino que también facilitan la exploración y comparación de posibles soluciones desde perspectivas éticas y sociales. Su principal ventaja reside en la posibilidad de anticipar consecuencias antes de implementar una determinada política en condiciones reales. En lugar de evaluar únicamente los resultados agregados de una decisión, estos modelos permiten analizar cómo los beneficios, riesgos y posibles daños se distribuyen entre diferentes grupos poblacionales. Esta capacidad resulta particularmente relevante cuando la búsqueda de eficiencia puede entrar en conflicto con objetivos relacionados con la equidad y la protección de poblaciones vulnerables.

La pandemia de la enfermedad por coronavirus de 2019 evidenció de manera especialmente clara la necesidad de disponer de marcos éticos para orientar la asignación de recursos sanitarios escasos. La elevada demanda asistencial, combinada con la disponibilidad limitada de camas hospitalarias, ventiladores, medicamentos, vacunas y personal sanitario, generó escenarios en los que las instituciones tuvieron que establecer prioridades. En este contexto, las simulaciones computacionales adquirieron especial relevancia como herramientas para comparar estrategias y estimar anticipadamente sus posibles consecuencias.

Una de sus aplicaciones corresponde a los algoritmos de triaje en cuidados intensivos. Diferentes estudios han empleado modelos de simulación para evaluar estrategias de priorización aplicables a las unidades de cuidados intensivos durante períodos de elevada demanda. White y Lo (2020) compararon 12 estrategias de asignación de ventiladores, considerando no solamente la mortalidad total, sino también las posibles disparidades en mortalidad entre grupos raciales y socioeconómicos. Los resultados mostraron que los algoritmos orientados exclusivamente a maximizar la probabilidad de supervivencia podían exacerbar desigualdades preexistentes, mientras que la incorporación de criterios relacionados con la equidad podía contribuir a disminuir esas disparidades con una reducción relativamente limitada de la eficiencia global.

Este tipo de resultados evidencia una cuestión fundamental para la bioética cuantitativa: una estrategia capaz de producir el mejor resultado agregado no necesariamente constituye la alternativa más justa. La maximización del número de supervivientes puede favorecer determinados perfiles clínicos o poblacionales cuando las condiciones de salud previas se encuentran asociadas con desigualdades estructurales. Por ello, la simulación permite visualizar las consecuencias distributivas de los criterios de priorización y determinar hasta qué punto diferentes

definiciones de eficiencia pueden entrar en tensión con el principio de justicia.

Otra aplicación importante corresponde a la distribución de vacunas y tratamientos. Los modelos de simulación permiten evaluar estrategias que consideran simultáneamente múltiples dimensiones éticas. Giordano et al. (2021) desarrollaron un modelo basado en agentes mediante el cual simulaban 15 estrategias diferentes de priorización vacunal, evaluando no solamente sus efectos sobre la transmisión y la mortalidad, sino también sus consecuencias para grupos vulnerables y la equidad intergeneracional. Los resultados indicaron que las estrategias basadas exclusivamente en la edad, aunque podían resultar eficaces para reducir la mortalidad, podían mantener desigualdades existentes en comunidades marginadas.

Estos hallazgos ponen de manifiesto la importancia de incorporar variables sociales y distributivas en los modelos utilizados para fundamentar decisiones sanitarias. La edad, el riesgo clínico o la probabilidad de supervivencia constituyen variables relevantes, pero pueden resultar insuficientes cuando se analizan de manera aislada. Las condiciones socioeconómicas, la exposición diferencial al riesgo, el acceso desigual a los servicios sanitarios y la presencia de enfermedades asociadas con determinantes sociales también pueden influir en los resultados. En consecuencia, las simulaciones permiten evaluar escenarios alternativos en los que estos factores sean considerados explícitamente.

Las simulaciones aplicadas a la equidad y a la mitigación de sesgos algorítmicos constituyen otro ámbito en el que estos métodos resultan particularmente valiosos. Los sistemas predictivos utilizados en medicina dependen considerablemente de la calidad y representatividad de los datos empleados durante su desarrollo. Cuando determinados grupos poblacionales están infrarrepresentados en los datos de entrenamiento, el rendimiento de los algoritmos puede disminuir precisamente en las poblaciones que presentan una menor representación estadística.

En este sentido, los métodos de simulación pueden contribuir a detectar sesgos en modelos predictivos al permitir evaluar sistemáticamente el comportamiento de los algoritmos en diferentes subpoblaciones, incluidas aquellas que pueden encontrarse infrarrepresentadas en los datos utilizados para su entrenamiento (Chen et al., 2021). Mediante la generación de escenarios controlados es posible modificar determinadas características poblacionales y observar cómo varían las predicciones, los errores y las recomendaciones del sistema. De esta forma, la evaluación puede trascender los indicadores globales de precisión para analizar la distribución de los errores entre grupos específicos.

Esta perspectiva resulta particularmente relevante porque un modelo puede presentar un elevado rendimiento

general y, simultáneamente, producir tasas de error significativamente diferentes entre grupos poblacionales. Una evaluación basada exclusivamente en el promedio podría ocultar estas diferencias. En cambio, la simulación permite identificar qué grupos presentan mayores tasas de falsos positivos, falsos negativos u otras formas de error clínicamente relevantes, proporcionando evidencia para valorar la equidad del sistema antes de su implementación generalizada.

Las simulaciones también posibilitan la evaluación comparativa de diferentes estrategias destinadas a mitigar los sesgos algorítmicos. En lugar de aplicar una única intervención correctiva, pueden modelarse distintas alternativas y estimarse sus efectos sobre la precisión, la equidad y otros criterios relevantes. Esto permite determinar si una estrategia destinada a disminuir una determinada desigualdad genera consecuencias adversas en otras dimensiones del sistema y facilita la identificación de soluciones capaces de alcanzar un equilibrio más adecuado entre múltiples objetivos.

En esta línea, Zhang (2024) examina las complejidades asociadas con la asignación de recursos durante emergencias sanitarias y presta especial atención a las disparidades raciales que afectan a las comunidades negras. El análisis cuestiona las limitaciones de los enfoques centrados exclusivamente en salvar el mayor número posible de vidas y examina alternativas destinadas a incorporar explícitamente consideraciones de equidad. Esta perspectiva resulta relevante porque determinadas condiciones clínicas utilizadas para establecer prioridades pueden encontrarse vinculadas con desigualdades sociales y sanitarias preexistentes, por lo que su utilización aparentemente neutral puede reproducir o intensificar dichas desigualdades.

Zhang (2024) también analiza el Sistema de Doble Principio propuesto por Bruce y Tallman (2021), cuyo propósito consiste en alcanzar un mayor equilibrio entre equidad y utilidad mediante ajustes en las puntuaciones de triaje asociados con enfermedades que presentan disparidades raciales. Este enfoque intenta evitar que las desigualdades sanitarias históricas se traduzcan automáticamente en menores posibilidades de acceso a recursos durante situaciones de emergencia. Sin embargo, su implementación plantea desafíos relacionados con la subjetividad de los criterios, la complejidad operativa y la propia definición de equidad.

Ante estas limitaciones, Zhang (2024) plantea alternativas dirigidas a fortalecer la objetividad y capacidad adaptativa de los mecanismos de asignación. Entre ellas se encuentran el establecimiento de criterios de selección sustentados en análisis de datos, la actualización de la información en tiempo real, la adaptación de los modelos a las características regionales y la incorporación de

mecanismos de lotería combinados con ajustes en las puntuaciones de triaje. Estas propuestas muestran que la equidad no necesariamente debe concebirse como un criterio opuesto a la eficiencia, sino como una dimensión que puede integrarse explícitamente en el diseño de los sistemas de priorización.

En conjunto, estas aplicaciones muestran que las simulaciones computacionales pueden desempeñar una doble función en el análisis bioético. Por una parte, permiten anticipar las consecuencias clínicas y poblacionales de diferentes decisiones antes de su implementación. Por otra, hacen visibles las implicaciones distributivas que pueden permanecer ocultas cuando únicamente se consideran indicadores agregados de eficiencia. Su utilización permite comparar escenarios, identificar poblaciones potencialmente perjudicadas, evaluar estrategias de mitigación y analizar la estabilidad de las decisiones frente a diferentes supuestos.

No obstante, los resultados obtenidos mediante simulaciones deben interpretarse como instrumentos de apoyo a la deliberación y no como sustitutos del razonamiento ético. Todo modelo depende de supuestos, variables, parámetros y criterios previamente seleccionados, por lo que las decisiones sobre qué elementos incorporar también poseen una dimensión normativa. La utilidad bioética de las simulaciones reside precisamente en hacer explícitos estos supuestos y permitir analizar de manera transparente cómo diferentes elecciones metodológicas pueden modificar las consecuencias previstas. De este modo, la simulación computacional puede contribuir a una toma de decisiones sanitarias más informada, prospectiva y sensible a los principios de justicia, equidad y protección de las poblaciones vulnerables.

Las simulaciones aplicadas a la equidad y a la mitigación de sesgos algorítmicos resultan especialmente valiosas para anticipar posibles consecuencias discriminatorias antes de que un sistema sea incorporado a la práctica clínica. En este contexto, los métodos de simulación permiten evaluar el comportamiento de los algoritmos en subpoblaciones que pueden encontrarse infrarrepresentadas en los datos de entrenamiento, lo que facilita la identificación de diferencias en precisión, sensibilidad, especificidad y distribución de errores. Esta capacidad adquiere especial importancia en medicina, donde un rendimiento global elevado puede ocultar resultados significativamente inferiores en determinados grupos poblacionales.

La detección de sesgos en modelos predictivos constituye una de las principales aplicaciones de estas metodologías. Chen et al. (2021) desarrollaron un marco metodológico capaz de generar datos sintéticos con diferentes tipos y niveles de sesgo, lo que posibilita evaluar la robustez de las técnicas de mitigación de sesgos antes de su implementación clínica. La utilización de datos sintéticos

permite modificar sistemáticamente determinadas características poblacionales y analizar cómo responde el modelo ante diferentes escenarios de representación, distribución y desigualdad. De esta manera, es posible identificar anticipadamente condiciones bajo las cuales un algoritmo podría producir resultados inequitativos.

Asimismo, las simulaciones resultan útiles para comparar diferentes estrategias orientadas a mitigar los sesgos algorítmicos. Esta comparación permite determinar si las intervenciones destinadas a mejorar la equidad afectan otras dimensiones relevantes del rendimiento del sistema. La reducción de una disparidad específica puede, por ejemplo, modificar la precisión global o alterar las tasas de falsos positivos y falsos negativos. Por ello, la evaluación cuantitativa de diferentes estrategias permite analizar las compensaciones existentes entre rendimiento, seguridad y equidad antes de seleccionar una alternativa para su aplicación clínica.

Las simulaciones relacionadas con la privacidad y la gobernanza de los datos constituyen otra aplicación relevante de la bioética cuantitativa. La utilización de grandes volúmenes de información médica plantea dilemas entre el aprovechamiento científico de los datos y la protección de la privacidad de los pacientes. La información clínica puede contribuir al desarrollo de mejores herramientas diagnósticas y predictivas, pero su utilización también puede generar riesgos relacionados con accesos no autorizados, identificación de individuos y utilización secundaria de información sensible.

En este marco, los modelos de simulación permiten cuantificar el equilibrio entre utilidad analítica y riesgo de reidentificación mediante la evaluación de diferentes técnicas de anonimización. El Emam et al. (2020) simuló ataques de reidentificación en conjuntos de datos anonimizados mediante diferentes métodos, proporcionando métricas cuantitativas para valorar el equilibrio entre la protección de la identidad y la conservación de la utilidad de la información. Este tipo de análisis permite determinar hasta qué punto un procedimiento de anonimización disminuye el riesgo de identificación sin deteriorar excesivamente el valor científico del conjunto de datos.

Esta cuestión evidencia que la máxima protección de la privacidad y la máxima utilidad de los datos no siempre pueden alcanzarse simultáneamente. Una transformación excesiva de la información puede reducir el riesgo de reidentificación, pero también puede disminuir su utilidad para determinadas investigaciones. En sentido contrario, conservar un elevado nivel de detalle puede incrementar el valor analítico de los datos y, simultáneamente, aumentar los riesgos para la privacidad. Las simulaciones permiten representar cuantitativamente esta tensión y proporcionar evidencia para establecer niveles

de protección compatibles con los objetivos científicos y las obligaciones éticas.

Por otra parte, los sistemas de diagnóstico asistido por inteligencia artificial plantean dilemas éticos específicos que pueden analizarse mediante aproximaciones cuantitativas. Una de las principales cuestiones corresponde al grado de autonomía que debe concederse al sistema y al nivel de supervisión humana necesario para garantizar decisiones seguras. La utilización de inteligencia artificial puede mejorar determinadas capacidades diagnósticas, pero también puede modificar la forma en que los profesionales interpretan la información clínica y toman decisiones.

Los modelos de simulación permiten evaluar diferentes niveles de autonomía diagnóstica y supervisión humana. Cabitza et al. (2023) analizaron distintas configuraciones de colaboración entre seres humanos e inteligencia artificial en el diagnóstico médico, examinando cómo diferentes protocolos de interacción pueden modificar el desempeño diagnóstico. Este tipo de análisis resulta relevante porque la eficacia de un sistema no depende exclusivamente de su precisión individual, sino también de la manera en que sus recomendaciones son integradas en el razonamiento del profesional.

La interacción entre profesionales y sistemas algorítmicos introduce además el riesgo de dependencia excesiva de las recomendaciones automatizadas. Cuando un sistema presenta sus resultados con un elevado nivel aparente de seguridad, el profesional puede atribuirles una autoridad superior a la que realmente poseen. En consecuencia, la evaluación ética debe considerar no solamente si el algoritmo produce predicciones correctas, sino también cómo la presentación de esas predicciones modifica el comportamiento de quienes las utilizan.

Las simulaciones pueden contribuir igualmente al diseño de mecanismos mediante los cuales los sistemas de inteligencia artificial comuniquen su incertidumbre y sus limitaciones. La comunicación de la incertidumbre constituye un componente esencial de la seguridad clínica, debido a que las predicciones algorítmicas son probabilísticas y se encuentran condicionadas por las características de los datos y del modelo. Gray et al. (2025) desarrollaron un modelo destinado a evaluar diferentes interfaces para comunicar la confianza algorítmica, midiendo su impacto sobre la toma de decisiones clínicas y sobre el riesgo de generar una confianza excesiva en la inteligencia artificial.

Este aspecto resulta fundamental para preservar la autonomía profesional. La comunicación adecuada de los niveles de confianza puede ayudar al personal sanitario a interpretar una recomendación como información de apoyo y no como una decisión definitiva. Por consiguiente, la transparencia debe incluir no solamente explicaciones sobre los factores que contribuyeron a una predicción,

sino también información comprensible acerca de su incertidumbre, limitaciones y condiciones de validez.

Los modelos predictivos orientados a identificar pacientes con alto riesgo para una intervención temprana generan igualmente dilemas entre el beneficio poblacional y la potencial estigmatización. La clasificación de una persona como paciente de alto riesgo puede facilitar intervenciones preventivas y mejorar determinados resultados clínicos, pero también puede producir consecuencias relacionadas con etiquetamiento, medicalización innecesaria o utilización de recursos en personas que finalmente no desarrollarán el evento previsto.

En este contexto, los modelos matemáticos permiten optimizar los umbrales utilizados para activar intervenciones preventivas considerando tanto el beneficio clínico esperado como las posibles consecuencias derivadas de clasificaciones incorrectas. Obermeyer et al. (2019) analizaron el funcionamiento de algoritmos utilizados para identificar pacientes que requerían atención adicional y demostraron que determinadas variables empleadas como aproximaciones de las necesidades sanitarias podían introducir importantes sesgos raciales. Este hallazgo evidencia que la selección de variables y umbrales no constituye una decisión exclusivamente técnica, debido a que puede modificar la distribución de los beneficios asistenciales entre diferentes grupos poblacionales.

Las simulaciones también posibilitan evaluar cómo los sistemas predictivos pueden afectar las disparidades existentes en salud. Este análisis del impacto sobre la equidad permite determinar si una herramienta destinada a mejorar la eficiencia asistencial beneficia de manera semejante a diferentes poblaciones o reproduce desigualdades previamente existentes. Pierson et al. (2021) analizaron mediante modelado las consecuencias de utilizar sistemas predictivos en diferentes poblaciones y mostraron la importancia de considerar explícitamente la distribución de los beneficios derivados de las decisiones algorítmicas. En ausencia de ajustes orientados a la equidad, una herramienta predictiva puede reproducir patrones presentes en los datos históricos y trasladarlos a futuras decisiones clínicas.

La creciente autonomía de los sistemas quirúrgicos y robóticos introduce otra dimensión relevante del debate bioético. A medida que estos dispositivos adquieren mayores capacidades para ejecutar acciones, interpretar información y participar en determinadas decisiones durante los procedimientos médicos, surgen interrogantes relacionados con la responsabilidad, el control humano, la seguridad y los límites aceptables de la autonomía tecnológica. La cuestión fundamental no reside únicamente en determinar qué tareas puede realizar técnicamente un sistema, sino cuáles deben permanecer bajo control directo de profesionales sanitarios.

La delegación de determinadas funciones quirúrgicas a sistemas con mayores niveles de autonomía puede incrementar la precisión de algunas intervenciones, estandarizar procedimientos y reducir determinadas formas de error humano. Sin embargo, estos beneficios deben analizarse conjuntamente con los riesgos derivados de fallos técnicos, interpretaciones incorrectas de situaciones imprevistas o respuestas inadecuadas ante condiciones que no fueron suficientemente representadas durante el desarrollo y validación del sistema.

La responsabilidad constituye uno de los problemas más complejos en este escenario. Cuando una decisión es resultado de la interacción entre un profesional sanitario, un sistema robótico, un modelo algorítmico, una institución asistencial y un fabricante tecnológico, la atribución de responsabilidad puede resultar considerablemente más compleja que en los procedimientos convencionales. La existencia de múltiples actores obliga a establecer previamente qué decisiones pueden ser delegadas, quién conserva la autoridad final y qué mecanismos deben activarse cuando el comportamiento del sistema se aparta de los resultados esperados.

Por esta razón, resulta fundamental establecer marcos claros de responsabilidad que definan el grado de control que los profesionales sanitarios deben mantener sobre estos sistemas, así como protocolos de supervisión, intervención y suspensión que garanticen una utilización segura y éticamente justificable. La autonomía tecnológica debe concebirse de manera gradual y proporcional al nivel de riesgo asociado con cada función. Cuanto mayor sea la capacidad de un sistema para intervenir directamente sobre decisiones con consecuencias clínicas significativas, mayores deberán ser las exigencias relacionadas con validación, transparencia, supervisión y rendición de cuentas.

El desafío consiste, por tanto, en equilibrar los beneficios derivados de la innovación tecnológica con la necesidad de preservar un enfoque humano en la práctica médica. La eficiencia, precisión y capacidad de procesamiento proporcionadas por los sistemas automatizados no eliminan la importancia del juicio profesional, la comunicación con el paciente y la responsabilidad ética. Por el contrario, la creciente autonomía tecnológica hace necesario definir con mayor precisión los límites de la delegación y los mecanismos mediante los cuales los profesionales pueden mantener un control significativo sobre las decisiones clínicas.

En conjunto, las simulaciones aplicadas a sesgos algorítmicos, privacidad, diagnóstico asistido, predicción de riesgos y sistemas autónomos muestran la amplitud de aplicaciones de los métodos cuantitativos en bioética. Estas herramientas permiten anticipar consecuencias, comparar alternativas y hacer visibles posibles conflictos

entre eficiencia, equidad, privacidad, autonomía y seguridad. Sin embargo, sus resultados deben incorporarse a procesos de deliberación ética en los que los valores y criterios normativos permanezcan explícitos. De esta manera, la simulación computacional puede contribuir a que la innovación tecnológica avance de forma compatible con la protección de los pacientes, la responsabilidad profesional y los principios fundamentales de la atención sanitaria.

El presente estudio ha permitido sistematizar el creciente cuerpo de literatura que emplea métodos cuantitativos y simulaciones computacionales para abordar dilemas éticos en sistemas de inteligencia artificial aplicados a la medicina. A partir de esta sistematización, se ha delineado un marco de Evaluación de Impacto Ético (EIE) estructurado en cuatro fases. En esta sección se discuten las implicaciones de estos hallazgos, se establecen comparaciones con propuestas existentes en la literatura y se reconocen las principales limitaciones del estudio.

Tres hallazgos emergen como especialmente significativos. En primer lugar, la identificación de tres dominios críticos, correspondientes a la asignación de recursos, la mitigación de sesgos algorítmicos y la gobernanza de datos, en los que los enfoques bioéticos tradicionales resultan insuficientes por sí solos, sugiere que la complejidad computacional de los sistemas de inteligencia artificial exige un avance metodológico. Este hallazgo no implica que los principios formulados por Beauchamp y Childress (2019) hayan quedado obsoletos, sino que requieren ser operacionalizados mediante instrumentos capaces de representar cuantitativamente las consecuencias de diferentes decisiones. Como señala Mittelstadt (2019), los principios abstractos ofrecen una orientación limitada para la implementación algorítmica; las simulaciones basadas en agentes y los modelos de optimización multiobjetivo examinados en esta revisión constituyen una posible respuesta a esta brecha entre formulación normativa y aplicación tecnológica.

En segundo lugar, la evidencia revisada muestra de manera consistente que las optimizaciones unidimensionales, como maximizar exclusivamente la supervivencia o la eficiencia, pueden exacerbar desigualdades preexistentes (Obermeyer et al., 2019; White & Lo, 2020). Este patrón permite plantear un principio relevante para el diseño de sistemas de inteligencia artificial aplicados a la medicina: la dimensión ética no debe concebirse como una restricción externa que pueda incorporarse después del desarrollo tecnológico, sino como una propiedad que debe considerarse conjuntamente con los objetivos clínicos durante el diseño, validación e implementación del sistema.

Desde esta perspectiva, la bioética cuantitativa no se limita a evaluar retrospectivamente las consecuencias éticas

una vez que el sistema ha sido implementado, sino que propone incorporar métricas relacionadas con justicia, equidad, autonomía, seguridad y beneficio clínico dentro del propio proceso de diseño algorítmico. Este cambio implica pasar de una ética predominantemente reactiva a una evaluación prospectiva, en la que diferentes configuraciones puedan compararse antes de afectar a pacientes reales. La simulación permite precisamente analizar cómo determinadas decisiones técnicas pueden modificar la distribución de beneficios, riesgos y errores entre diferentes grupos poblacionales.

En tercer lugar, la recurrencia de estudios que emplean análisis de sensibilidad (Razavi et al., 2021; Saltelli, 2021) apunta a una característica relevante de los dilemas éticos asociados con la inteligencia artificial: las conclusiones pueden presentar una elevada sensibilidad ante variaciones en determinados supuestos o parámetros fundamentales. Entre estos se encuentran la ponderación relativa atribuida a la eficiencia y la equidad, los niveles aceptables de riesgo, los umbrales utilizados para activar una intervención y las tasas de error diagnóstico registradas en diferentes subpoblaciones.

Este hallazgo presenta una implicación práctica importante. Los comités de ética hospitalaria, las instituciones sanitarias y los desarrolladores de sistemas de inteligencia artificial deberían considerar, como parte de los procesos de evaluación ética, la elaboración de análisis de sensibilidad que permitan identificar aquellos parámetros cuya modificación produce cambios sustanciales en los resultados. De esta manera, sería posible determinar los umbrales críticos a partir de los cuales una determinada configuración tecnológica puede generar consecuencias incompatibles con los criterios éticos previamente establecidos.

El marco de Evaluación de Impacto Ético propuesto en este estudio no constituye el primer intento de estructurar la evaluación ética de sistemas de inteligencia artificial en el ámbito sanitario. Jobin et al. (2019) analizaron 84 guías internacionales sobre ética de la inteligencia artificial y encontraron que, aunque existe un considerable grado de convergencia en torno a principios como transparencia, justicia y responsabilidad, persisten diferencias importantes respecto de su interpretación y aplicación. Este problema evidencia la distancia existente entre la formulación de principios generales y su traducción en procedimientos concretos de evaluación.

El marco de Evaluación de Impacto Ético pretende contribuir a reducir esta distancia mediante la articulación de cuatro fases complementarias. La primera comprende la identificación de las dimensiones éticas relevantes y su traducción en indicadores susceptibles de evaluación. La segunda incorpora modelos computacionales y simulaciones prospectivas destinados a anticipar posibles

consecuencias. La tercera introduce el análisis de sensibilidad para determinar cómo las variaciones en parámetros y supuestos modifican los resultados. Finalmente, la cuarta fase contempla mecanismos de seguimiento, auditoría y ajuste continuo una vez que el sistema ha sido incorporado al entorno clínico.

En comparación con propuestas internacionales de evaluación del impacto algorítmico y con las Directrices Éticas para una Inteligencia Artificial Fiable elaboradas por el Grupo Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial de la Comisión Europea, el marco de Evaluación de Impacto Ético presenta dos diferencias conceptuales relevantes. La primera corresponde a la incorporación explícita de la simulación prospectiva en su segunda fase como herramienta para anticipar consecuencias que podrían no resultar intuitivamente previsibles. Esta característica pretende superar las limitaciones de los enfoques predominantemente reactivos o estructurados mediante listas de verificación (Morley et al., 2020).

La segunda diferencia corresponde al énfasis otorgado al análisis de sensibilidad ética durante la tercera fase. Este componente parte del reconocimiento de que determinados parámetros relacionados con valores y prioridades pueden estar sujetos a desacuerdos razonables. Por consiguiente, un marco de evaluación robusto no debería asumir necesariamente la existencia de una única ponderación universalmente válida entre principios como eficiencia, equidad, autonomía o seguridad. Por el contrario, debería permitir examinar cómo se modifican los resultados cuando cambian esas ponderaciones y determinar si las recomendaciones permanecen estables ante diferentes posiciones éticas legítimas.

No obstante, el marco de Evaluación de Impacto Ético también presenta similitudes con enfoques previos, particularmente con propuestas orientadas a la evaluación de sesgos antes y después de la implementación de modelos de aprendizaje automático (Chen et al., 2021). El marco propuesto integra este tipo de auditorías en su cuarta fase, pero las sitúa dentro de un proceso continuo de seguimiento y ajuste. Esta característica responde al reconocimiento de que los sistemas de inteligencia artificial utilizados en medicina operan en entornos dinámicos, donde pueden modificarse las características de las poblaciones, los patrones presentes en los datos, las tecnologías disponibles y las normas de la práctica clínica.

El seguimiento continuo resulta especialmente importante porque un sistema que presenta un comportamiento adecuado durante su validación inicial puede experimentar cambios en su rendimiento cuando las condiciones del entorno se modifican. Por esta razón, la evaluación ética no debería finalizar con la autorización o implementación del sistema, sino continuar durante todo su ciclo de vida. La vigilancia posterior permite identificar desviaciones,

nuevas desigualdades, cambios en la distribución de errores o consecuencias no anticipadas durante las etapas iniciales.

A pesar de estas contribuciones, es necesario reconocer diversas limitaciones que condicionan el alcance y la generalización de los hallazgos obtenidos.

La primera limitación corresponde a la ausencia de validación empírica del marco de Evaluación de Impacto Ético. El marco propuesto ha sido desarrollado a partir de una revisión crítica de la literatura, pero no ha sido sometido a validación mediante estudios de caso, implementaciones piloto en entornos clínicos reales ni consultas formales a paneles de expertos mediante el método Delphi. Por consiguiente, debe entenderse como una propuesta conceptual que requiere procesos posteriores de refinamiento, validación y evaluación empírica antes de considerar su adopción generalizada.

En particular, todavía debe determinarse si las cuatro fases propuestas poseen la misma relevancia para todos los tipos de sistemas de inteligencia artificial utilizados en medicina. Los sistemas destinados al diagnóstico, pronóstico, asignación de recursos, apoyo a decisiones clínicas o robótica quirúrgica presentan características y niveles de riesgo diferentes. Es posible, por tanto, que algunos dominios requieran modificaciones en las fases propuestas, indicadores específicos o mecanismos adicionales de evaluación y supervisión.

La segunda limitación corresponde a la heterogeneidad de los estudios revisados. La literatura sobre bioética cuantitativa constituye un campo emergente y presenta una considerable diversidad metodológica. Los estudios incluidos difieren en los tipos de simulaciones utilizadas, entre ellas los modelos basados en agentes, el modelo continuo y las simulaciones de Monte Carlo. También existen diferencias importantes en las métricas éticas empleadas, debido a que algunos estudios utilizan indicadores desarrollados específicamente para determinados contextos, mientras que otros recurren a medidas previamente establecidas.

Esta heterogeneidad también se observa en los dominios clínicos analizados, las poblaciones estudiadas y los resultados evaluados. Aunque esta diversidad refleja la expansión y el carácter interdisciplinario del campo, dificulta la realización de una síntesis cuantitativa agregada y limita la posibilidad de formular recomendaciones universalmente aplicables. No se realizó un metaanálisis cuantitativo debido a la diversidad de los diseños metodológicos, intervenciones, modelos y desenlaces, por lo que las conclusiones obtenidas poseen principalmente un carácter narrativo y temático.

La tercera limitación corresponde a los posibles sesgos de publicación y búsqueda. Aunque se consultaron cinco

bases de datos mediante una estrategia de búsqueda amplia, es posible que investigaciones relevantes publicadas en revistas con menor visibilidad, literatura gris, informes técnicos, tesis doctorales, documentos regulatorios o publicaciones en idiomas diferentes del inglés y el español no hayan sido identificadas. Esta situación puede limitar la representatividad del conjunto de estudios analizados.

Asimismo, debe considerarse la posibilidad de sesgo de publicación hacia resultados positivos, novedosos o metodológicamente exitosos. Es posible que existan investigaciones en las que las simulaciones no hayan logrado representar adecuadamente determinados dilemas éticos, hayan producido resultados inconsistentes o hayan identificado limitaciones sustanciales para la cuantificación de valores morales. Estos resultados negativos o inconclusos podrían encontrarse infrarrepresentados en la literatura publicada y, en consecuencia, también en la presente revisión.

Estas limitaciones no invalidan la utilidad potencial del marco propuesto, pero señalan la necesidad de interpretarlo como un punto de partida susceptible de evaluación y perfeccionamiento. Las investigaciones futuras deberán validar empíricamente sus cuatro fases en diferentes contextos clínicos, comparar su funcionamiento entre distintos tipos de sistemas de inteligencia artificial e incorporar la participación de profesionales sanitarios, especialistas en bioética, desarrolladores, responsables institucionales y pacientes. Esta validación permitirá determinar hasta qué punto la Evaluación de Impacto Ético puede convertirse en una herramienta reproducible para integrar criterios éticos cuantificables en el diseño, implementación y seguimiento de sistemas de inteligencia artificial aplicados a la medicina.

La cuarta limitación corresponde a la simplificación inherente a la modelización cuantitativa. Los enfoques de bioética cuantitativa, por su propia naturaleza, requieren traducir determinados valores morales, como la autonomía, la dignidad, la justicia o la justicia relacional, en variables, indicadores y parámetros susceptibles de representación matemática. Esta transformación implica necesariamente cierto grado de simplificación y puede excluir matices cualitativos relevantes para la comprensión integral de los problemas éticos. La posibilidad de cuantificar una determinada dimensión no significa necesariamente que todos sus componentes moralmente significativos puedan representarse mediante una medida numérica.

Esta dificultad resulta particularmente evidente en conceptos caracterizados por una elevada complejidad relacional y contextual. La confianza existente en la relación entre profesionales sanitarios y pacientes, por ejemplo, difícilmente puede ser representada en toda su

complejidad mediante una variable ordinal o una probabilidad bayesiana. De manera semejante, la equidad no necesariamente puede reducirse completamente a un único indicador estadístico, como el coeficiente de Gini, ni a la igualación de las tasas de falsos positivos entre diferentes grupos poblacionales. Estos indicadores pueden capturar dimensiones relevantes del fenómeno, pero no necesariamente representan todas sus implicaciones normativas, sociales y contextuales.

Esta limitación pone de manifiesto una tensión epistemológica fundamental para la bioética cuantitativa: la diferencia entre aquello que puede medirse y aquello que posee significado moral. La construcción de un indicador exige seleccionar determinadas dimensiones del fenómeno, establecer reglas de medición y determinar cómo deben interpretarse sus valores. Estas decisiones no son completamente neutrales, debido a que reflejan supuestos acerca de qué aspectos se consideran relevantes y qué concepción de justicia, autonomía o beneficio se pretende representar.

El marco de Evaluación de Impacto Ético no pretende resolver esta tensión epistemológica, sino hacerla explícita dentro del proceso de evaluación. Permanece abierta la cuestión relativa a hasta qué punto los dilemas éticos pueden traducirse en modelos matemáticos sin producir una pérdida significativa de contenido moral. Por esta razón, los resultados cuantitativos deben interpretarse como una fuente adicional de evidencia para la deliberación y no como una representación exhaustiva de los valores involucrados. El marco de Evaluación de Impacto Ético debe entenderse, por tanto, como un complemento y no como un sustituto del razonamiento ético cualitativo, contextual y deliberativo.

La quinta limitación corresponde a la aplicabilidad del marco en contextos caracterizados por recursos limitados. La implementación completa de la Evaluación de Impacto Ético puede requerir capacidades técnicas significativas, entre ellas profesionales especializados en ciencia de datos, modelización computacional y análisis estadístico, además de acceso a infraestructura tecnológica adecuada y disponibilidad de tiempo para ejecutar simulaciones y análisis de sensibilidad. Estas condiciones pueden no encontrarse igualmente disponibles en todos los sistemas sanitarios.

Esta limitación adquiere especial relevancia en países de ingresos bajos y medios, instituciones sanitarias rurales y organizaciones con capacidades tecnológicas reducidas. La ausencia de personal especializado o de infraestructura computacional puede generar una nueva forma de desigualdad en la gobernanza ética de la inteligencia artificial: las instituciones con mayores recursos podrían disponer de mecanismos más sofisticados para evaluar los riesgos de sus sistemas, mientras que aquellas con

menores capacidades podrían adoptar tecnologías similares sin disponer de herramientas equivalentes para examinar sus consecuencias.

El presente estudio no aborda suficientemente cómo adaptar el marco de Evaluación de Impacto Ético a estos contextos ni desarrolla versiones simplificadas que requieran menores capacidades técnicas y computacionales. Esta cuestión deberá abordarse en investigaciones posteriores, debido a que un marco destinado a promover la equidad en los sistemas de inteligencia artificial podría producir paradójicamente nuevas desigualdades si su aplicación solamente fuera viable para instituciones con elevados niveles de recursos económicos, tecnológicos y humanos.

Una posible línea de desarrollo consiste en establecer diferentes niveles de evaluación en función del riesgo y la complejidad de los sistemas. Las aplicaciones con consecuencias clínicas limitadas podrían someterse a procedimientos simplificados, mientras que aquellas capaces de influir directamente sobre diagnósticos, tratamientos, asignación de recursos o intervenciones de alto riesgo requerirían evaluaciones cuantitativas más exhaustivas. Esta estructura permitiría conservar los componentes esenciales del marco y, al mismo tiempo, adaptar las exigencias metodológicas a las capacidades institucionales y al nivel de riesgo de cada tecnología.

A pesar de estas limitaciones, los hallazgos obtenidos permiten identificar diferentes líneas prioritarias para futuras investigaciones. En primer lugar, se requieren estudios de implementación que apliquen el marco de Evaluación de Impacto Ético a casos clínicos concretos. Entre ellos podrían incluirse la evaluación prospectiva de un algoritmo de triaje en una unidad de cuidados intensivos, un sistema destinado a predecir readmisiones hospitalarias, una herramienta de apoyo al diagnóstico o un sistema algorítmico utilizado para priorizar determinadas intervenciones sanitarias.

Estos estudios deberían documentar no solamente los resultados favorables obtenidos mediante la aplicación del marco, sino también las dificultades prácticas encontradas durante su implementación. Resultaría especialmente relevante conocer qué métricas pueden operacionalizarse adecuadamente, cuáles generan mayores desacuerdos entre los actores participantes, qué recursos son necesarios para ejecutar las simulaciones y hasta qué punto los resultados de los análisis cuantitativos modifican efectivamente las decisiones adoptadas por profesionales, instituciones y comités de ética.

En segundo lugar, resulta necesario desarrollar bibliotecas de métricas éticas estandarizadas y herramientas informáticas de código abierto que faciliten la aplicación de la segunda fase, correspondiente a la modelización, y de la tercera fase, relacionada con el análisis de sensibilidad.

La disponibilidad de recursos accesibles podría reducir las barreras técnicas para investigadores, profesionales sanitarios y comités de ética que no cuentan con formación matemática o computacional avanzada.

No obstante, la estandarización de las métricas éticas deberá realizarse con precaución. La existencia de indicadores comunes facilitaría la comparación entre sistemas y estudios, pero no debería conducir a asumir que conceptos como justicia, autonomía o equidad poseen una única operacionalización válida en todos los contextos. Las herramientas estandarizadas deberían permitir adaptar las ponderaciones y criterios a las características clínicas, sociales y culturales de cada entorno, manteniendo explícitos los supuestos normativos utilizados.

En tercer lugar, resulta necesario profundizar en la validez ecológica de las simulaciones, entendida como el grado en que los resultados producidos por un modelo computacional representan adecuadamente las situaciones que pueden observarse en la práctica clínica real. Un modelo puede presentar coherencia matemática y producir resultados internamente consistentes sin reproducir completamente la complejidad de los entornos sanitarios. Las interacciones humanas, los cambios organizativos, las diferencias culturales, las limitaciones institucionales y los comportamientos imprevistos pueden modificar considerablemente los efectos reales de una tecnología.

La evaluación de esta correspondencia entre simulación y práctica clínica requiere investigaciones metodológicas específicas que comparen las predicciones obtenidas mediante los modelos con los resultados observados después de la implementación. Esta comparación permitiría identificar qué componentes de las simulaciones poseen mayor capacidad predictiva, qué supuestos requieren modificaciones y qué dimensiones de la realidad clínica resultan difíciles de representar mediante modelos cuantitativos.

Otra línea de investigación debería orientarse al desarrollo de mecanismos participativos para definir las métricas y ponderaciones utilizadas en la Evaluación de Impacto Ético. La selección de criterios relacionados con equidad, autonomía, seguridad o beneficio no debería depender exclusivamente de especialistas en modelización o desarrollo tecnológico. Profesionales sanitarios, especialistas en bioética, pacientes, responsables institucionales y representantes de las poblaciones potencialmente afectadas pueden aportar perspectivas diferentes sobre cuáles consecuencias deben considerarse relevantes y qué niveles de riesgo resultan aceptables.

Esta participación permitiría evitar que las decisiones normativas permanezcan ocultas dentro de parámetros aparentemente técnicos. Un modelo matemático puede producir resultados precisos respecto de los objetivos que se le han asignado, pero no puede determinar por sí

mismo cuáles deberían ser esos objetivos. La legitimidad ética del proceso depende, por tanto, tanto de la calidad técnica de la modelización como de la transparencia y pluralidad mediante las cuales se definen los valores incorporados al modelo.

Finalmente, la bioética cuantitativa no debe desarrollarse de manera aislada de la bioética cualitativa, clínica y filosófica. La creciente utilización de métodos matemáticos y computacionales ofrece nuevas posibilidades para anticipar consecuencias, comparar alternativas y hacer visibles determinadas desigualdades, pero no elimina la necesidad de interpretar esas evidencias dentro de contextos humanos concretos. Los resultados cuantitativos requieren deliberación para determinar su significado moral, del mismo modo que la reflexión ética puede beneficiarse de evidencia cuantitativa que permita estimar las consecuencias previsibles de diferentes alternativas.

En este sentido, una de las perspectivas más prometedoras corresponde al desarrollo de modelos híbridos en los que los análisis cuantitativos informen la deliberación ética cualitativa y, simultáneamente, la reflexión filosófica y clínica determine qué variables, restricciones y resultados deben incorporarse a los modelos cuantitativos. Esta relación bidireccional evita tanto la reducción de los problemas éticos a cálculos matemáticos como la formulación de principios abstractos desvinculados de las consecuencias observables de su aplicación.

El marco de Evaluación de Impacto Ético propuesto pretende constituir un instrumento para favorecer esta integración. Su finalidad no consiste en sustituir la reflexión moral mediante algoritmos, sino en proporcionar una estructura mediante la cual la evidencia cuantitativa, la simulación computacional, el análisis de sensibilidad y la deliberación ética puedan interactuar durante el diseño, implementación y seguimiento de los sistemas de inteligencia artificial aplicados a la medicina. Desde esta perspectiva, el principal aporte de la bioética cuantitativa no reside en convertir los valores humanos en números, sino en utilizar las herramientas cuantitativas disponibles para comprender mejor las consecuencias de las decisiones tecnológicas sin perder de vista la complejidad moral, relacional y humana que caracteriza a la práctica sanitaria.

CONCLUSIONES

El presente estudio identifica la bioética cuantitativa como un campo emergente con un considerable potencial para abordar los complejos dilemas éticos derivados de la incorporación de la inteligencia artificial en la medicina. Los hallazgos muestran que las simulaciones computacionales y los modelos matemáticos pueden proporcionar herramientas relevantes para analizar y cuantificar las compensaciones entre principios éticos en conflicto, evaluar las consecuencias poblacionales de políticas y

algoritmos, identificar efectos no intuitivos y propiedades emergentes en sistemas complejos, así como desarrollar evaluaciones éticas prospectivas antes de la implementación de nuevas tecnologías.

La incorporación de métodos cuantitativos permite ampliar las posibilidades de la deliberación bioética tradicional al proporcionar evidencia sobre las consecuencias previsibles de diferentes alternativas. Esta perspectiva resulta especialmente relevante en sistemas de inteligencia artificial capaces de procesar grandes volúmenes de información, intervenir en decisiones clínicas y producir efectos simultáneos sobre poblaciones heterogéneas. En estos contextos, la evaluación ética no puede limitarse exclusivamente a determinar si una tecnología cumple formalmente determinados principios, sino que debe examinar cómo esos principios se materializan en sus resultados, cómo se distribuyen los beneficios y riesgos y qué poblaciones pueden resultar particularmente afectadas.

El marco de Evaluación de Impacto Ético propuesto en este estudio constituye una herramienta operativa orientada a favorecer la incorporación sistemática de la bioética cuantitativa en entornos clínicos y de investigación. A diferencia de los enfoques centrados exclusivamente en la formulación de principios generales, el marco plantea un ciclo de evaluación compuesto por cuatro fases interrelacionadas. La primera fase comprende la identificación participativa de los valores y principios potencialmente en conflicto y su traducción en dimensiones susceptibles de evaluación. La segunda incorpora la simulación prospectiva de escenarios y el análisis cuantitativo de las compensaciones entre diferentes objetivos clínicos y éticos. La tercera introduce el análisis de sensibilidad para identificar parámetros y umbrales críticos capaces de modificar sustancialmente los resultados. Finalmente, la cuarta fase establece un proceso de seguimiento continuo mediante auditorías éticas periódicas y mecanismos de ajuste durante el ciclo de vida del sistema.

El principal aporte del marco consiste en proporcionar una estructura que permita representar cuantitativamente determinadas dimensiones de dilemas éticos complejos sin reducir su significado exclusivamente a variables matemáticas. Esta aproximación puede facilitar que los comités de ética hospitalaria, los desarrolladores de sistemas de inteligencia artificial, los profesionales sanitarios y las autoridades competentes anticipen consecuencias potencialmente no intuitivas, comparen diferentes estrategias de mitigación de sesgos, identifiquen desigualdades en la distribución de beneficios y riesgos y establezcan mecanismos de rendición de cuentas sustentados en evidencia cuantitativa.

La implementación sistemática de este marco puede contribuir a reducir la distancia existente entre los principios bioéticos declarados y su aplicación efectiva en el diseño

algorítmico. Su carácter prospectivo permite incorporar consideraciones éticas antes de que los sistemas produzcan consecuencias sobre pacientes y poblaciones, mientras que el seguimiento continuo reconoce que el comportamiento de una tecnología puede modificarse a medida que cambian los datos, las poblaciones, los contextos institucionales y las prácticas clínicas. De esta manera, la evaluación ética deja de constituir una intervención aislada realizada antes o después de la implementación y pasa a formar parte del ciclo de vida completo de la tecnología.

No obstante, la bioética cuantitativa no debe interpretarse como una alternativa destinada a sustituir la deliberación ética cualitativa, clínica o filosófica. Los modelos matemáticos pueden representar determinadas dimensiones de conceptos como equidad, beneficio, riesgo o justicia distributiva, pero no agotan la complejidad moral de valores como dignidad, autonomía, confianza o responsabilidad. Por ello, la utilidad de los métodos cuantitativos depende de su integración con procesos deliberativos capaces de interpretar los resultados, cuestionar los supuestos incorporados a los modelos y considerar las particularidades humanas y contextuales de cada decisión.

La principal perspectiva derivada de este estudio es, por tanto, la necesidad de avanzar hacia una integración entre bioética cuantitativa y deliberación ética. Los métodos computacionales pueden aportar evidencia sobre las consecuencias previsibles de diferentes decisiones, mientras que la reflexión bioética debe determinar qué valores requieren protección, qué riesgos resultan aceptables y qué criterios deben orientar el diseño y utilización de los sistemas. Esta complementariedad permite aprovechar la capacidad analítica de las herramientas cuantitativas sin reducir los problemas morales a simples procedimientos de optimización.

La medicina del siglo XXI requiere aproximaciones éticas capaces de responder a la creciente sofisticación de las tecnologías que intervienen en la atención sanitaria. En este escenario, la bioética cuantitativa representa una evolución metodológica relevante hacia una gobernanza ética más prospectiva, rigurosa, transparente y sustentada en evidencia para los sistemas de inteligencia artificial aplicados a la salud. Su desarrollo futuro dependerá de la capacidad para integrar modelización computacional, evidencia clínica, participación de los actores involucrados y deliberación bioética, garantizando que la innovación tecnológica permanezca subordinada a la protección de la dignidad, la equidad, la autonomía, la seguridad y el bienestar de las personas.

REFERENCIAS

- Albán-Sabando, E. A. (2024). Innovación en la práctica enfermera: nuevas tecnologías para el cuidado del paciente. *Sophia Research Review*, 1(1), 10-14. <https://doi.org/10.64092/52vny797>
- Beauchamp, T. L., & Childress, J. F. (2019). Principles of biomedical ethics (8th ed.). Oxford University Press.
- Bruce, L., & Tallman, R. (2021). Promoting racial equity in COVID-19 resource allocation. *Journal of Medical Ethics*, 47(4), 208-212. <https://doi.org/10.1136/medethics-2020-106794>
- Cabitza, F., Campagner, A., Ronzio, L., Cameli, M., Mandoli, G. E., Pastore, M. C., Sconfienza, L. M., Folgado, D., Barandas, M., & Gamboa, H. (2023). Rams, hounds and white boxes: Investigating human-AI collaboration protocols in medical diagnosis. *Artificial Intelligence in Medicine*, 138, 102506. <https://doi.org/10.1016/j.artmed.2023.102506>
- Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care—Addressing ethical challenges. *The New England Journal of Medicine*, 378(11), 981-983. <https://doi.org/10.1056/NEJMp1714229>
- Chen, I. Y., Pierson, E., Rose, S., Joshi, S., Ferryman, K., & Ghassemi, M. (2021). Ethical machine learning in healthcare. *Annual Review of Biomedical Data Science*, 4, 123-144. <https://doi.org/10.1146/annurev-bio-datasci-092820-114757>
- Doumpos, M., Figueira, J. R., Greco, S., & Zopounidis, C. (Eds.). (2019). *New perspectives in multiple criteria decision making: Innovative applications and case studies*. Springer.
- El Emam, K., Mosquera, L., & Bass, J. (2020). Evaluating identity disclosure risk in fully synthetic health data: Model development and validation. *Journal of Medical Internet Research*, 22(11), e23139. <https://doi.org/10.2196/23139>
- Emanuel, E. J., Persad, G., Upshur, R., Thome, B., Parker, M., Glickman, A., Zhang, C., Boyle, C., Smith, M., & Phillips, J. P. (2020). Fair allocation of scarce medical resources in the time of Covid-19. *The New England Journal of Medicine*, 382(21), 2049-2055. <https://doi.org/10.1056/NEJMs2005114>
- Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A., & Colaneri, M. (2021). Modeling vaccination rollouts, SARS-CoV-2 variants and the requirement for non-pharmaceutical interventions in Italy. *Nature Medicine*, 28(5), 1051-1058. <https://doi.org/10.1038/s41591-021-01334-5>

- Gray, N., Page, H., Buchan, I., & Joyce, D. W. (2025). *Risk and uncertainty communication in deployed AI-based clinical decision support systems: A scoping review* [Preprint]. medRxiv. <https://doi.org/10.1101/2024.12.06.24318489>
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. <https://doi.org/10.1038/s42256-019-0088-2>
- Juca-Maldonado, A. X., García-Vera, J. S., Juca-Maldonado, F., & Carrión-González, J. T. (2025). Competencias tecnológicas de la Generación Z. Una evaluación cuantitativa de las habilidades frente a la percepción de los nativos digitales. *Revista Transdisciplinaria de Estudios Sociales y Tecnológicos*, 5(2), 16–24. <https://doi.org/10.58594/rtest.v5i2.160>
- McDougall, R. J. (2019). Computer knows best? The need for value-flexibility in medical AI. *Journal of Medical Ethics*, 45(3), 156–160. <https://doi.org/10.1136/medethics-2018-105118>
- Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1(11), 501–507. <https://doi.org/10.1038/s42256-019-0114-4>
- Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*, 26(4), 2141–2168. <https://doi.org/10.1007/s11948-019-00165-5>
- Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. <https://doi.org/10.1126/science.aax2342>
- Pierson, E., Cutler, D. M., Leskovec, J., Mullainathan, S., & Obermeyer, Z. (2021). An algorithmic approach to reducing unexplained pain disparities in underserved populations. *Nature Medicine*, 27(1), 136–140. <https://doi.org/10.1038/s41591-020-01192-7>
- Price, W. N. (2019). Medical AI and contextual bias. *Harvard Journal of Law & Technology*, 33(1), 65–116. <https://repository.law.umich.edu/cgi/viewcontent.cgi?article=3242&context=facarticles>
- Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., et al. (2021). The future of sensitivity analysis: An essential discipline for systems modeling and policy support. *Environmental Modelling & Software*, 137, 104954. <https://doi.org/10.1016/j.envsoft.2020.104954>
- Saltelli, A., Jakeman, A., Razavi, S., & Wu, Q. (2021). Sensitivity analysis: A discipline coming of age. *Environmental Modelling & Software*, 146, 105226. <https://doi.org/10.1016/j.envsoft.2021.105226>
- Singh, B., & Rebennack, S. (2025). Release immediately or sequentially? Strategies for allocating scarce therapeutic resources during disease outbreaks. *IJSE Transactions*, 1–16. <https://doi.org/10.1080/24725854.2025.2525918>
- White, D. B., & Lo, B. (2020). A framework for rationing ventilators and critical care beds during the COVID-19 pandemic. *JAMA*, 323(18), 1773–1774. <https://doi.org/10.1001/jama.2020.5046>
- Zhang, C. (2024). Equitable resource allocation in health emergencies: Addressing racial disparities and ethical dilemmas. *Journal of Medical Ethics*, 52(7), 493–496. <https://doi.org/10.1136/jme-2024-109947>

Conflictos de interés:

Los autores declaran no tener conflictos de interés.

Contribución de los autores:

Loraine María Frutos-Morales, María Caridad González-Alonso, Marisela Angelis Rivero-Fernández: Concepción y diseño del estudio, adquisición de datos, análisis e interpretación, redacción del manuscrito, revisión crítica del contenido, análisis estadístico, supervisión general del estudio.

Declaración ética:

El estudio se basó en el análisis de fuentes documentales y datos de acceso público, por lo que no implicó la participación directa de seres humanos. No se manejó información personal identificable.