CLIAS

CENTRO DE INTELIGENCIA ARTIFICIAL Y SALUD PARA AMÉRICA LATINA Y EL CARIBE

Marcos de evaluación prácticos para la INTELIGENCIA ARTIFICIAL RESPONSABLE APLICADA A LA SALUD: UNA REVISIÓN DE ALCANCE

DOCUMENTO TÉCNICO 8

Mayo 2025

CIPS IMPLEMENTACIÓN E INNOVACIÓN EN POLÍTICAS DE SALUD

IECS INSTITUTO DE EFECTIVIDAD CLÍNICA Y SANITARIA

CONTENIDO

EQUIPO DE TRABAJO

ANALÍA PASTRANA: Medica Pediatra especialista en Neurología Infantil. Magister en Ciencia de Datos. Miembro de la subcomisión de Tecnologías de Información y Comunicación de la Sociedad Argentina de Pediatría.

MARTIN SABAN: Médico pediatra (Universidad de Buenos Aires). Miembro de la subcomisión de Tecnologías de la Información y Comunicación de la Sociedad Argentina de Pediatría. Candidato a Master en Efectividad Clínica y Sanitaria. Becario CIIPS-IECS.

DENISE ZAVALA: Licenciada en Psicología (UBA). Diplomatura en Gestión de políticas sanitarias en el territorio (UNGS). Candidata a Master en Efectividad Clínica (UBA). Investigadora del CIIPS-IECS.

CINTIA CEJAS: Lic. en Ciencias Políticas (UCA) y Magister en Ciencias Sociales y de la Salud (FLACSO-CEDES). Especialista en gestión de proyectos de salud. Coordinadora del Centro de Implementación e Innovación en Políticas de Salud (CIIPS-IECS) y del Centro de Inteligencia Artificial en Salud para Latinoamérica y el Caribe (CLIAS).

PRESENTACIÓN

El presente documento, elaborado por el Centro de Implementación e Innovación en Políticas de Salud (CIIPS) del Instituto de Efectividad Clínica y Sanitaria (IECS), se enmarca en una Serie de Documentos Técnicos sobre Inteligencia Artificial y Salud, que tienen por objetivo aportar al conocimiento de la región, abordando distintos ejes y perspectivas relevantes en el análisis de esta temática.

Destinados a equipos de salud, formuladores de programas y políticas sanitarias y decisores en todos los niveles y población en general, con especial interés en la transformación digital del sector salud y su vinculación a la salud sexual, reproductiva y materna (SSRM), esta serie de documentos sobre IA se complementan con las actividades llevadas a cabo por el CLIAS (Centro de Inteligencia Artificial en Salud para Latinoamérica y el Caribe) que se desarrolla en el CIIPS, con el apoyo del International Development Research Centre (IDRC) y del Foreign, Commonwealth & Development Office (FCDO). Para más información sobre el CLIAS, visitar http://clias.iecs.org.ar

La evaluación del impacto de las herramientas de inteligencia artificial (IA) aplicadas a la salud es un proceso complejo que requiere tiempo significativo debido a la necesidad de múltiples instancias claramente diferenciadas:

El **diseño y desarrollo inicial** implica evaluar la viabilidad técnica, ética y operativa del proyecto, asegurando que la IA pueda y deba desarrollarse. La **evaluación técnica y clínica en contextos controlados** consiste en validar rigurosamente que la IA cumple eficazmente con el propósito específico para el que fue diseñada, garantizando precisión y seguridad antes de su implementación real; mientras que la **evaluación en entornos reales** permite confirmar que la IA aporta valor tangible al sistema de salud, comparando su desempeño y resultados frente a otras soluciones existentes, para justificar su adopción y generalización. Esta fase es especialmente prolongada debido a la necesidad de generar evidencia robusta en contextos prácticos y dinámicos.

El presente documento **analiza los marcos de referencia identificados en la literatura científica para evaluar el diseño y desarrollo inicial de herramientas y sistemas de IA en el ámbito de la salud.** En particular, se pone énfasis en aquellos marcos centrados en la IA responsable, que además incorporan herramientas prácticas para operacionalizar los principios bioéticos a lo largo de todo el ciclo de vida de estas tecnologías.

Se espera que este documento proporcione a equipos y personas interesadas en incorporar IA a herramientas aplicables y viables para evaluar su desarrollo y diseño de forma responsable, entendiendo que este es un paso inicial y fundamental para la evaluación del impacto en la salud. Además, aspira a servir como punto de partida para el desarrollo de un nuevo marco de referencia que integre todas las dimensiones de análisis encontradas.

RESUMEN EJECUTIVO

Si bien existen marcos de evaluación para la inteligencia artificial (IA), la mayoría se enfoca en principios teóricos y carecen de operacionalizaciones concretas para implementar una IA responsable. Este documento busca cubrir esta necesidad al ofrecer un análisis exhaustivo de los marcos de evaluación prácticos aplicados a la salud. Esta revisión identifica y examina **cuatro marcos** clave que brindan herramientas operativas para trasladar los principios bioéticos a cada etapa del ciclo de vida de las tecnologías de IA en salud, destacando la importancia de integrar consideraciones éticas y de aplicabilidad práctica.

Los marcos analizados son:

Evaluación del impacto ético de la UNESCO(1)

Ética, Evidencia y Equidad de la Asociación Médica Estadounidense (AMA)(2,3)

Marco del Biomedical Data Science Lab (BDSLab) para IA confiable(4)

Marco de ética operacional de Solanki y colaboradores(5)

Estos marcos fueron evaluados por sus herramientas prácticas que desarrollaron, como listas de verificación y preguntas orientadoras, destacándose por su integración de principios éticos y su abordaje de la perspectiva de género. Sus **fortalezas** incluyen metodologías estructuradas para evaluar impactos éticos, criterios claros para la efectividad en la práctica clínica y recursos técnicos detallados, como matrices de evaluación y ejemplos de código abierto, que facilitan la implementación de una IA responsable en salud.

Una de las principales **limitaciones** identificadas es que ninguno de los marcos analizados incorpora una dimensión específica dedicada exclusivamente al género, lo que evidencia una brecha crítica en la capacidad de abordar la equidad de género en las aplicaciones de IA en salud. Aunque existen herramientas prácticas, estas **requieren adaptaciones para ser más inclusivas y transformadoras, especialmente en lo que respecta a la equidad de género y la escalabilidad en contextos con recursos limitados**.

A partir de estos hallazgos, se recomienda:

- Estructurar el desarrollo de futuros marcos de evaluación comenzando desde principios éticos generales y avanzando hacia herramientas específicas y aplicables.

- Es imprescindible en los equipos de trabajo definir roles y responsabilidades claras, fomentar la colaboración interdisciplinaria e inclusiva para desarrollar sistemas de IA culturalmente sensibles y alineados con valores sociales clave. La perspectiva de género debe integrarse como una dimensión específica y transformadora, incorporando herramientas concretas que permitan identificar y corregir sesgos.

- Diseñar marcos de evaluación que sean escalables y adaptables a diferentes contextos, incluyendo mecanismos de monitoreo continuo y auditorías periódicas para mitigar riesgos emergentes.

01. INTRODUCCIÓN

La inteligencia artificial (IA) ha revolucionado la manera en que interactuamos con la tecnología, permitiendo avances sin precedentes en diversos sectores. Desde el reconocimiento de patrones complejos hasta la automatización de tareas, la IA ofrece soluciones que optimizan procesos y mejoran la toma de decisiones en tiempo real.(6) No obstante, a medida que esta tecnología avanza, se hace evidente la necesidad de un enfoque práctico, ético y reflexivo que garantice que la IA no perpetúe desigualdades, sino que promueva equidad y justicia social.(7) Este enfoque ha dado lugar al concepto de **IA responsable**, el cual va más allá del rendimiento técnico, abarcando los principios de rendición de cuentas, responsabilidad y transparencia.[1](8)

El concepto de IA responsable reconoce que, si bien la IA puede generar importantes beneficios, existen riesgos si no se desarrollan marcos adecuados para su implementación.(9) Estos riesgos incluyen la posibilidad de amplificar sesgos existentes en los datos o reproducir estructuras de poder desiguales. Por ello, una IA responsable no solo se preocupa por la eficiencia, sino por la alineación con valores éticos, asegurando que las soluciones tecnológicas respeten los derechos humanos.(10)

En el campo de la salud, la IA ofrece soluciones innovadoras para mejorar el diagnóstico, el tratamiento y la gestión de enfermedades. Los sistemas de IA pueden procesar enormes volúmenes de datos de pacientes, ayudando a los profesionales de la salud a tomar decisiones más rápidas y precisas.(11) Además, la IA puede personalizar los tratamientos y predecir la evolución de enfermedades basándose en análisis predictivos avanzados.(12) La integración de la IA en la práctica clínica no está exenta de desafíos por lo que es necesario tener instrumentos prácticos que aborden los aspectos técnicos, éticos y los posibles impactos sociales de su uso. Sin un enfoque holístico, los sistemas de IA podrían perpetuar desigualdades ya presentes en los sistemas de salud, especialmente aquellas relacionadas con el acceso a los servicios de atención médica en poblaciones vulnerables.

[1] La rendición de cuentas se refiere al requisito de que el sistema sea capaz de explicar y justificar sus decisiones a los usuarios y otros actores relevantes. Para garantizar la rendición de cuentas, las decisiones deben derivarse de los mecanismos de toma de decisiones utilizados y explicarse por ellos. También requiere que los valores morales y las normas sociales que informan el propósito del sistema, así como sus interpretaciones operativas, se hayan obtenido de manera abierta involucrando a todas las partes interesadas.
La responsabilidad se refiere al papel de las propias personas en su relación con los sistemas de IA. La responsabilidad no se trata solo de establecer reglas para gobernar las máquinas inteligentes; se trata de todo el sistema sociotécnico en el que opera el sistema, y que abarca a las personas, las máquinas y las instituciones.
La transparencia indica la capacidad de describir, inspeccionar y reproducir los mecanismos a través de los cuales los sistemas de IA toman decisiones y aprenden a adaptarse a su entorno, y la procedencia y dinámica de los datos que utiliza y crea el sistema. La transparencia implica ser explícito y abierto en cuanto a las opciones y decisiones relativas a las fuentes de datos, los procesos de desarrollo y las partes interesadas. Las partes interesadas también deberían participar en las decisiones sobre todos los modelos que utilizan datos humanos o afectan a seres humanos o pueden tener otro impacto moralmente significativo.

Es en este contexto que la **perspectiva de género** y la **interseccionalidad** se vuelven esenciales para el desarrollo de una IA responsable en salud. Los algoritmos de IA, si no se diseñan cuidadosamente, pueden reproducir sesgos de género que afectan negativamente a poblaciones vulnerables. Estos sesgos no solo impactan en la precisión de los diagnósticos, sino también en el acceso equitativo a los tratamientos. Además, las dimensiones interseccionales como la etnia, el nivel socioeconómico y la ubicación geográfica, también deben considerarse para evitar que la IA refuerce barreras estructurales en la atención sanitaria.

Dado que la IA puede generar consecuencias significativas en el bienestar humano y social, es crucial contar con instrumentos de evaluación antes de su implementación. Para orientar dicha evaluación, se pueden utilizar **marcos (*frameworks*)**.

A través de los marcos, se puede hacer una evaluación integral de los sistemas de IA, midiendo no sólo su rendimiento técnico, sino también su impacto social, ético y clínico. Los marcos entonces, servirán a modo de guías para garantizar que los sistemas de IA se alineen con los principios IA responsable, mencionados anteriormente.

Algunos marcos incluyen herramientas concretas y operativas, como listas de verificación que permiten evaluar el cumplimiento de criterios específicos, guías estructuradas con preguntas orientadoras para analizar aspectos éticos y técnicos, o plantillas predefinidas para documentar y monitorear el progreso en cada etapa del ciclo de vida de la IA. La presencia de estas herramientas indica que el marco no solo propone principios teóricos, sino que también ofrece mecanismos claros y accionables para aplicarlos en contextos reales, ayudando a garantizar su utilidad en la práctica.

Por su parte, la evaluación podría ser llevada a cabo por desarrolladores, reguladores, comités éticos y profesionales de la salud, quienes pueden contribuir a que la IA cumpla con estándares de seguridad, equidad y transparencia antes de su aplicación en entornos clínicos.

La presente revisión tiene por objetivo, entonces, **identificar, describir y analizar marcos de evaluación para la IA responsable que aporten herramientas prácticas y que además, incluyan una perspectiva de género en su desarrollo**. Finalmente, se proponen recomendaciones para la evaluación de la IA responsable.

02. METODOLOGÍA

Para identificar los marcos disponibles se realizó una revisión de alcance que siguió el marco metodológico de cinco etapas de Arksey y O'Malley (13), en el que se evaluó al alcance de la bibliografía disponible sobre un tema concreto.

La **PRIMERA ETAPA** comenzó con la identificación de una/s pregunta/s de investigación, que en este estudio fueron las siguientes:

- ¿Qué marcos de evaluación para la IA responsable existen y qué dimensiones de análisis presentan?
- ¿De qué manera los marcos operacionalizan sus dimensiones? ¿Utilizan indicadores, listas de verificación u otras herramientas prácticas que permitan su medición o implementación?
- ¿Los marcos incluyen una dimensión de género?
- ¿El género está incorporado de forma implícita o explícita dentro de alguna de las dimensiones propuestas por los marcos?

La **SEGUNDA ETAPA** se refirió a la identificación de estudios relevantes para responder las preguntas anteriormente planteadas. Se realizó una búsqueda en PubMed, IEEE, Scopus, Scielo y Google (ver estrategia de búsqueda en el Anexo A). Además, se identificaron estudios a partir de listas de referencias y de consultas a expertos. En la Figura 1 se presenta el diagrama de flujo en el proceso de búsqueda y relevamiento de la literatura.

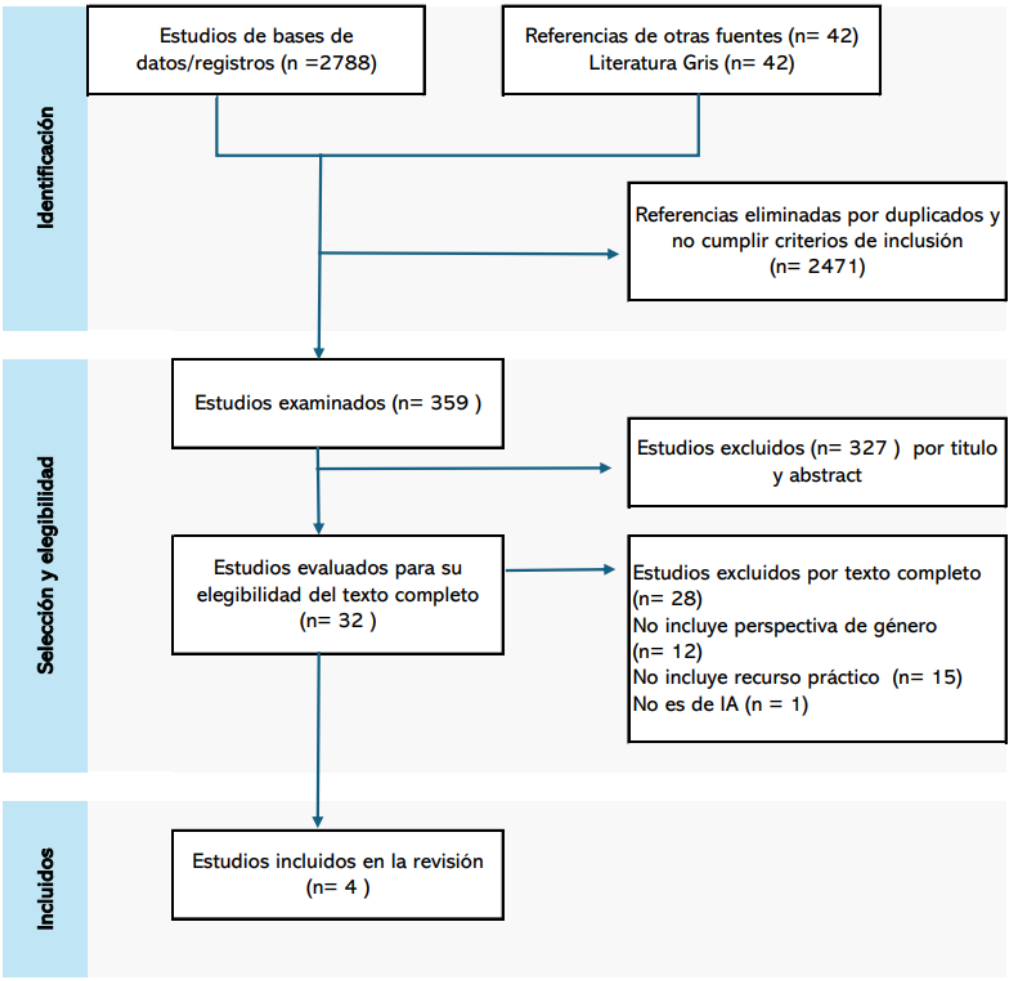


Figura 1. Diagrama de flujo PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) del proceso de revisión.

La **TERCERA ETAPA** comprende la selección de los estudios apropiados y la eliminación de aquellos que no responden a las preguntas de investigación propuestas. Los criterios de inclusión y de exclusión definidos para esta etapa se especifican en la Tabla 1.

INCLUSIÓN	EXCLUSIÓN
Artículos que incluyan marcos y que aborden simultáneamente la salud, el género y la inteligencia artificial	Artículos sobre marcos teóricos que no presenten una operacionalización de sus dimensiones
Artículos publicados a partir del 2019	Artículos publicados antes del 2019
Artículos publicados en español o ingles	
La ubicación del estudio puede ser global	

Tabla 1. *Criterios de inclusión y de exclusión.*

Para la selección de los estudios, en un primer momento se identificaron aquellos artículos que incluyeran en sus títulos o resúmenes las siguientes palabras: “marco”, “género”, “salud” e “inteligencia artificial”, sus traducciones o similares. En un segundo momento, se analizaron de manera completa los textos anteriormente seleccionados. Dicho análisis fue realizado por dos revisores de forma independiente y los desacuerdos se resolvieron por consenso.

En la **CUARTA ETAPA** se registraron en un Excel los datos de todos los estudios seleccionados y se confeccionó una tabla que incluye información sobre cada estudio, como el título, autores, año de publicación, tipo de recurso práctico, abordaje del género, dimensiones del marco, objetivos y destinatarios.

La **QUINTA Y ÚLTIMA ETAPA** consistió en resumir y reportar los resultados. Para el análisis y el reporte de los resultados se separó la literatura relevada en dos categorías: el “diseño conceptual” y el “componente técnico”. La dimensión de género está analizada tanto en el diseño conceptual como en el componente técnico.

Aquellos marcos ubicados dentro de la categoría de “**diseño conceptual**” buscan comprender los aspectos relacionados con el diseño del proyecto de IA, su propósito, el público objetivo al que está dirigido, etcétera. Este enfoque responde las preguntas relacionadas con el "por qué" y el "para qué" de la herramienta, lo que proporciona una visión global que considera tanto el contexto de implementación como las implicancias sociales y éticas asociadas a su adopción o uso.

Los marcos categorizados dentro del “**componente técnico**” proponen una serie de métodos para analizar la calidad de los conjuntos de datos utilizados, la precisión de los algoritmos y la adherencia a estándares técnicos con el propósito de que los objetivos y decisiones establecidos en la etapa anterior se hagan explícitos. Este enfoque aborda el "qué" de la herramienta, la robustez y fiabilidad técnica de la misma.

03. RESULTADOS

CARACTERÍSTICAS DE LA LITERATURA RELEVADA

Durante el proceso de revisión sistemática, se identificaron inicialmente **2,788 registros** a través de bases de datos específicas como PubMed, IEEE, Scopus, Scielo y Google, a los que se sumaron **42 registros adicionales** obtenidos de referencias y consultas a expertos, alcanzando un total de **2,830 registros**.

Tras la eliminación de duplicados, quedaron **359 registros** para el cribado. En esta etapa, se excluyeron **327 registros** tras la revisión de títulos y resúmenes, dejando **32 textos completos** para la evaluación de elegibilidad.

En la fase de elegibilidad, se excluyeron **28 textos completos** por no cumplir con los criterios definidos. Específicamente, **12 textos** fueron descartados por no incluir una perspectiva de género, **15** por carecer de recursos prácticos, y **1** por no estar relacionado con inteligencia artificial.

Finalmente, se incluyeron **4 estudios** en el análisis final, cumpliendo con todos los criterios establecidos para esta revisión. En la **Tabla 1** se presenta un resumen individual de cada uno de ellos.

DESCRIPCIÓN DE LOS MARCOS SELECCIONADOS

Para facilitar la descripción y conceptualización de los marcos seleccionados, estos se dividieron en dos categorías: aquellos enfocados en el **diseño conceptual** y los orientados al **componente técnico**.

MARCOS CON FOCO EN EL DISEÑO CONCEPTUAL

Los marcos centrados en el diseño conceptual establecen principios y lineamientos éticos fundamentales para la incorporación de la IA en salud. Su objetivo es asegurar que la tecnología esté alineada con valores como equidad, transparencia y derechos humanos desde las etapas iniciales de su desarrollo. Estos marcos proporcionan herramientas para evaluar los impactos sociales y bioéticos de la IA antes de su implementación, facilitando su aceptación y uso responsable en entornos sanitarios. Bajo esta categoría se detectaron dos marcos de evaluación: el marco de la UNESCO(1) y el de Asociación Médica

Estadounidense (AMA por sus siglas en inglés)[2].(2,3) A continuación, se desarrolla cada uno de ellos.

ORGANIZACIÓN DE LAS NACIONES UNIDAS PARA LA EDUCACIÓN, LA CIENCIA Y LA CULTURA (UNESCO)

En 2021, la UNESCO publicó la "*Recomendación sobre la ética de la IA*", un marco para garantizar que los desarrollos en inteligencia artificial estén alineados con la promoción y protección de los derechos humanos y la dignidad humana, la sostenibilidad ambiental, la equidad, la inclusión y la igualdad de género. En la **Tabla 2** se resumen los principios de la UNESCO y una breve descripción de cada uno de ellos. Posteriormente, y con el objetivo de lograr una implementación efectiva de los principios éticos establecidos en dicho documento, la UNESCO desarrolló un instrumento para la evaluación del impacto ético de la IA (*Ethical Impact Assessment*).(1)

PRINCIPIO ÉTICO UNESCO	DESCRIPCIÓN
Seguridad y protección	Evitar daños y abordar riesgos de seguridad y vulnerabilidades en sistemas de IA.
Sostenibilidad	Promover la justicia social y la inclusión, asegurando que los beneficios de la IA sean accesibles para todos.
Privacidad y protección de datos	Evaluar y minimizar el impacto ambiental directo e indirecto de los sistemas de IA.
Supervisión y determinación humanas	Proteger la privacidad a lo largo del ciclo de vida de la IA con marcos adecuados de protección de datos.

Tabla 2. *Principios éticos de la UNESCO y su descripción.*

Este instrumento se diseñó especialmente para ayudar a equipos e individuos de gobiernos a llevar a cabo una evaluación de impacto ético en cualquier sector, más allá del de salud. Si bien el objetivo es tener un marco para evaluar los proyectos de IA aplicada a la salud, este instrumento podría utilizarse en otros ámbitos.

El instrumento se divide en dos partes. La primera parte consiste en una serie de preguntas en relación con la:

[2] La AMA es la asociación más grande a nivel nacional en Estados Unidos. Agrupa a más de 190 sociedades médicas estatales y especializadas, así como a otras partes interesadas clave.

1. **Descripción del proyecto**: descripción, objetivos, problema que busca resolver, tipos de usuarios que interactuarían con la IA, entre otros aspectos generales que hacen al diseño del proyecto.

2. **Tamizaje de la proporcionalidad**: refiere a chequear que la IA que se desarrolle no sea utilizada para actividades éticamente cuestionables, evaluar si los beneficios esperados justifican los posibles riesgos asociados con la tecnología. Esto es útil para manejar posibles conflictos entre los valores éticos, como la tensión entre la transparencia y la protección de la privacidad.

3. **Gobernanza del proyecto**: con esto se pretende evaluar la diversidad (de disciplinas y de actores) del equipo que diseña la herramienta de IA y determinar claramente roles y responsabilidad de cada uno para minimizar riesgos y evitar confusiones.

4. **Gobernanza multi-stakeholder**: de acuerdo con el tipo de proyecto, pueden involucrarse actores de diversos niveles. Las preguntas de este ítem indagan la existencia y el diseño del plan de participación de las partes interesadas.

La segunda parte del instrumento evalúa que el diseño, desarrollo e implementación de la IA sea consistente con los principios éticos de la IA de UNESCO. Para cada principio se listan una serie de preguntas que tienen como objetivo evaluar:

a. Si se han establecido suficientes garantías procesales para garantizar que este sistema cumpla con la Recomendación.

b. Los (potenciales) resultados positivos y los impactos adversos que pueden surgir de la adquisición y el despliegue de la herramienta o sistema de IA, específicos de su contexto de uso.

La **Tabla 3** contiene las preguntas de la segunda etapa del instrumento aplicado al principio de *Equidad/Justicia, No discriminación, Diversidad.*

ASOCIACIÓN MÉDICA ESTADOUNIDENSE (AMA)

Por su parte, la AMA construyó un marco para el desarrollo y uso de IA en el campo de la salud, basándose en su política interna para la IA y en una revisión de la literatura que llevó a cabo para identificar los principales desafíos que plantea la implementación de la IA en la atención sanitaria. Para abordar estos desafíos, la AMA examinó las guías existentes y los modelos regulatorios aplicados a la introducción de tecnologías innovadoras en la práctica clínica.(2,3)

Este marco está dirigido a profesionales de la salud y *stakeholders* involucrados en el desarrollo, implementación y uso de la IA en salud. Su propósito es guiar el desarrollo e implementación responsable de la IA en el ámbito de la salud, asegurando que se utilice

para mejorar la atención médica y la salud de la población, sin comprometer los valores éticos y la equidad. En este sentido, la ética, la evidencia y la equidad son los tres principios esenciales para construir confianza entre los pacientes y los profesionales de la salud.

Los tres principios están inspirados en la visión de la AMA sobre el cuádruple objetivo de la IA:

a. Mejorar la atención del paciente,
b. Mejorar la salud de la población,
c. Mejorar la calidad de vida de los profesionales de la salud y,
d. Reducir los costos

De este modo, y con un enfoque práctico, los autores proponen una lista de verificación con dos categorías principales (planificación y desarrollo, e implementación y monitoreo) y los roles y las responsabilidades necesarias de cada grupo de involucrados. Este *checklist* incluye tareas como la identificación de objetivos clínicos, la supervisión continua del sistema y la corrección de sesgos.

En la **Tabla 4** se encuentra la lista de verificación propuesta por los autores.

En sintonía con los principios de la IA responsable, la propuesta de la AMA posiciona a las partes interesadas como responsables y transparenta el proceso para que todos rindan cuentas por el cumplimiento de estas expectativas ante la adopción de una herramienta.

En este sentido, los equipos podrían utilizar el marco para evaluar si la herramienta **funciona** (es decir, si cumple con las expectativas en materia de ética, evidencia y equidad), si **funciona para sus pacientes** (es decir, si ha demostrado que mejora la atención en una población de pacientes con características similares a las suyas), y si **mejoran los resultados de salud** (es decir, si se ha demostrado que, por ejemplo, la herramienta agrega valor a la relación médico-paciente, permitiendo una atención centrada en el paciente o, los datos sobre el desempeño clínico y la experiencia del paciente demuestran un impacto positivo en las medidas de calidad de vida a través de métodos de investigación cualitativos y cuantitativos, entre otros aspectos) (**Tabla 5**).

MARCOS CON FOCO EN COMPONENTES TÉCNICOS

Los marcos con foco en el componente técnico ofrecen herramientas y metodologías prácticas para evaluar la calidad de los datos, la precisión de los algoritmos y la adherencia a estándares de seguridad y confiabilidad en la IA aplicada a la salud. Estos marcos están diseñados para proporcionar guías concretas a desarrolladores y técnicos, asegurando que la IA cumpla con criterios rigurosos de transparencia, equidad y robustez técnica en su funcionamiento real.

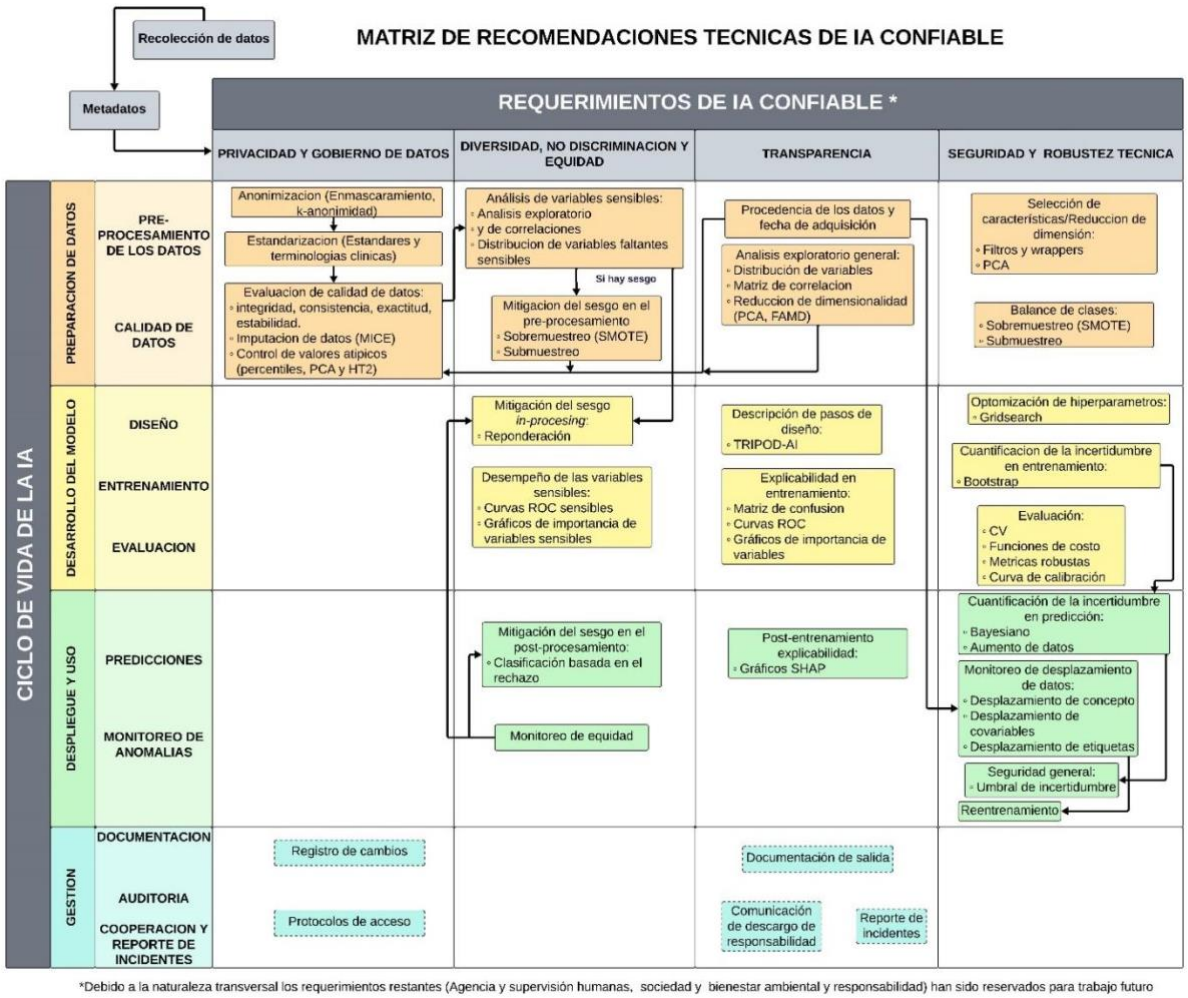
En relación con los marcos cuyo análisis se centra en el **componente técnico,** encontramos dos artículos.(4,5) Uno de ellos, De Manuel Vicente et al.(4) y el otro desarrollado por Solanki, Grundy y Hussain.(5)

MARCO DEL BIOMEDICAL DATA SCIENCE LAB (BDSLAB) PARA IA CONFIABLE

Este artículo propone un marco para guiar la creación de sistemas de IA confiables en el ámbito de la salud. El mismo está respaldado por ejemplos de código abierto que implementan los métodos necesarios para cumplir con los requisitos de confianza, validado mediante dos conjuntos de datos de salud de acceso abierto. Además, se propone una lista de verificación para facilitar la implementación de sistemas de IA confiables según los lineamientos de la UE. (Tabla 6)

Asimismo, los autores incluyeron una matriz de IA confiable, que clasifica métodos técnicos para cumplir con los requisitos de confianza: **privacidad y gobernanza de datos; diversidad, no discriminación y equidad; transparencia; y robustez técnica y seguridad**. La matriz organiza estos métodos a lo largo de las distintas etapas del ciclo de vida de la IA.

La **Figura 1** ilustra la matriz elaborada por los autores.



*Debido a la naturaleza transversal los requerimientos restantes (Agencia y supervisión humanas, sociedad y bienestar ambiental y responsabilidad) han sido reservados para trabajo futuro

Figura 1. *Matriz de recomendaciones técnicas de IA confiables. En las columnas están representados los lineamientos de la UE para la IA confiable y en las filas el ciclo de vida de la IA. Los cuadrantes representan los componentes necesarios para cumplir con cada lineamiento. Los métodos técnicos se indican mediante elementos dentro de las casillas. Los cuadrantes descatalogados indican que los componentes no están estrictamente relacionados con los métodos técnicos proporcionados. Las flechas indican posibles dependencias o flujos de trabajo entre los componentes. Modificado de de-Manuel-Vicente et al. (4)*

MARCO DE ÉTICA OPERACIONAL DE SOLANKI Y COLABORADORES

El segundo de los marcos es el desarrollado por Solanki, Grundy y Hussain(5), que tiene como público objetivo no solo a desarrolladores de IA en salud, sino también a los profesionales de la salud que usan estas herramientas.

El objetivo principal de su trabajo es proporcionar una descripción general de los problemas éticos específicos que surgen al integrar valores humanos en la IA aplicada a la atención médica. A partir de eso, se propone un marco operativo para la ética, basado en pautas existentes que ofrecen soluciones viables.

En la **Figura 2** se ilustra el marco propuesto por los autores, en el cual se considera el ciclo de vida de la IA en salud, desglosado en sus diferentes subcomponentes. Para cada uno de estos subcomponentes, se proporciona una lista de soluciones que se basan en principios éticos.

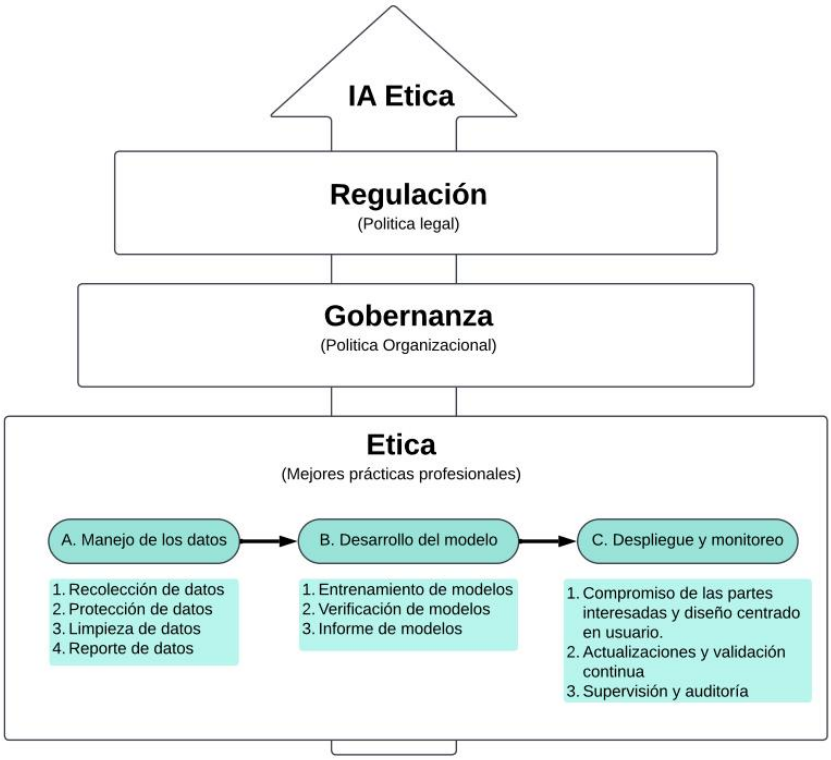


Figura 2. *Marco de la IA ética. Modificado de Solansky et al.(5)*

En la figura, se puede observar que la creación y el uso de IA ética es el resultado de múltiples niveles de influencia:

- **Prácticas profesionales de desarrolladores y usuarios:** Las decisiones y acciones de quienes crean y utilizan sistemas de IA afectan directamente su comportamiento ético. Esto incluye la selección de datos, el diseño de algoritmos y la implementación de medidas para evitar sesgos.

- **Gobernanza organizacional:** Las políticas y estructuras internas de las organizaciones que desarrollan o implementan IA determinan cómo se manejan las cuestiones éticas. Esto abarca desde códigos de conducta hasta comités de ética que supervisan proyectos de IA.

- **Regulación legal:** Las leyes y normativas establecidas por gobiernos y entidades reguladoras proporcionan un marco legal que guía y limita las acciones de individuos y organizaciones en relación con la IA, asegurando que se respeten estándares éticos y derechos humanos.

Estas capas interactúan entre sí, creando un entorno complejo donde la ética en la IA es moldeada por influencias técnicas, organizacionales y legales.

Por otra parte, el desarrollo de sistemas de IA se divide en tres fases secuenciales, cada una de las cuales debe completarse antes de avanzar a la siguiente:

- **Diseño:** En esta etapa se conceptualiza el sistema de IA, definiendo sus objetivos, funcionalidades y especificaciones técnicas. Se consideran aspectos éticos desde el inicio, como la equidad, la transparencia y la privacidad.
- **Desarrollo:** Aquí se lleva a cabo la construcción del sistema, incluyendo la programación, el entrenamiento de modelos con datos y la realización de pruebas iniciales. Es crucial implementar controles éticos, como la mitigación de sesgos y la garantía de seguridad.
- **Despliegue y uso:** Una vez desarrollado, el sistema se implementa en entornos reales. Durante esta fase, se monitorea su desempeño, se recopilan retroalimentaciones y se realizan ajustes necesarios para mantener la alineación con los principios éticos establecidos.

Este enfoque escalonado busca asegurar que, en cada fase del ciclo de vida de la IA, se aborden y refuercen las consideraciones éticas, para promover el desarrollo de sistemas responsables y confiables.

Para operacionalizar el marco de trabajo ético de IA identificaron pautas que brindan recomendaciones viables sobre necesidades éticas específicas en estas tres etapas (diseño, desarrollo, despliegue y uso), para prevenir, mitigar y abordar problemas éticos. En la Tabla 7 del Anexo se encuentra el checklist elaborado por los autores para abordar

las necesidades específicas de la IA para la atención sanitaria y los principios éticos involucrados.

REVISIÓN DE LA PERSPECTIVA DE GÉNERO

El **Marco de la UNESCO** integra el género dentro de la dimensión de Equidad, No Discriminación y Diversidad.(2) Este marco reconoce el impacto diferenciado que las tecnologías de IA pueden tener en las personas, siendo el género una variable más que deberá ser analizada por subgrupos, asegurando resultados y accesos equitativos. Con esta lógica, se definen una serie de preguntas para identificar y mitigar potenciales sesgos en los datos y promover la representación de género en los equipos para la toma de decisiones relacionadas con el desarrollo y uso de la IA.

En el **Marco de la Asociación Médica Estadounidense (AMA)** la perspectiva de género está incorporada dentro de la visión de la AMA sobre el cuádruple objetivo de la IA.(3) En este sentido, reconoce que la IA aborda las necesidades de salud de la población si no reproduce las desigualdades arraigadas en injusticias históricas y contemporáneas y la discriminación que afectan a las comunidades negras; indígenas; mujeres; comunidad LGBTIQ+; comunidades con discapacidades; comunidades con bajos ingresos; comunidades rurales; y otras comunidades usualmente excluidas por la industria de la salud. Asimismo, de no reconocer las diferencias, los sistemas de IA en salud pueden perpetuar o exacerbar desigualdades si los datos de entrenamiento contienen sesgos inherentes. Esto incluye sesgos que podrían influir en el diagnóstico, tratamiento y pronóstico en salud.

De manera similar al marco de la UNESCO, en el **Marco del Biomedical Data Science Lab (BDSLab)** también se hace mención a aspectos vinculados al género dentro de la dimensión de **Diversidad, Equidad y No Discriminación**.(4) El marco está basado en los lineamientos de la Unión Europea para la IA confiable por lo que esta dimensión hace referencia a que para lograr este principio se debe tener en cuenta a todos los afectados y garantizar su participación en todo el proceso. Asimismo, apunta a garantizar la igualdad de acceso mediante procesos de diseño inclusivos, con igualdad de trato independientemente de la edad, género, capacidades o características de las personas.(14) Por último, el **Marco de Solanki, Grundy y Hussain** incluye esta perspectiva en su dimensión de **Justicia y Equidad**.(5) Los autores destacan la importancia de garantizar que los sistemas de IA en salud no perpetúen sesgos de género ni discriminen a las personas por esta razón. Asimismo, se subraya que el desarrollo ético de la IA requiere la revisión crítica de los datos utilizados para entrenar los modelos, con el fin de identificar y corregir posibles sesgos, así como la incorporación de equipos diversos en el proceso de diseño y desarrollo.

04. DISCUSIÓN

A partir del análisis de los marcos seleccionados, se pueden identificar algunas fortalezas y limitaciones para la evaluación del proceso de desarrollo de herramientas de IA de forma responsable.

La literatura relevada es diversa, desde sus objetivos y sus destinatarios hasta las herramientas prácticas que proponen. Esta diversidad en el alcance de cada marco genera complementariedad para evaluar integralmente los sistemas de IA.

FORTALEZAS Y LIMITACIONES DE LOS MARCOS

En el marco de la UNESCO(1) por ejemplo, su herramienta práctica permite una evaluación exhaustiva que abarca desde el diseño y desarrollo de la IA hasta la evaluación de su implementación y los impactos que podrían surgir. De este modo, el marco no solo busca que los diversos actores involucrados reflexionen sobre la efectividad de la tecnología para responder a las necesidades de la población, sino que también los ubica en un lugar de responsabilidad para abordar los impactos positivos y negativos que puedan surgir. Este enfoque está alineado a los principios de la IA responsable y está estrechamente relacionado con el concepto de *Investigación e Innovación Responsable*, que subraya la necesidad de considerar los efectos potenciales de la IA en el ambiente y en la sociedad.(7) De manera similar, el marco de la AMA asigna roles y responsabilidades claras a los distintos actores y plantea preguntas clave para la evaluación de la IA.(3) Invita a los profesionales de la salud a evaluar si la IA está cumpliendo con su propósito, si mejora los resultados en salud y si responde adecuadamente a las necesidades de los pacientes.

La alineación de los marcos con los principios de rendición de cuentas, responsabilidad y transparencia en el desarrollo de IA responsable es crucial. Sin embargo, se reconoce que operacionalizar estos valores y principios éticos representa un desafío ya que suelen ser conceptos abstractos difíciles de incorporar en la práctica de los ingenieros de software. Los marcos de BDSLab(4) y Solanki, Grundy y Hussain(5) proporcionan métodos técnicos específicos y ejemplos de código que permiten a los desarrolladores implementar principios éticos a lo largo del ciclo de vida de la IA. Estos métodos incluyen técnicas de anonimización de datos, control de calidad, mitigación de sesgos y garantía de transparencia y explicabilidad de los modelos. Además, integran principios fundamentales de derechos humanos, como la dignidad, la autonomía y la igualdad, en el diseño de sistemas de IA. Esto representa una fortaleza al abordar con acciones concretas la brecha de operacionalización de la ética en el desarrollo técnico de la IA.

ENFOQUES SEGÚN EL TIPO DE MARCO

Los marcos con foco en el diseño conceptual son útiles para tomadores de decisiones, líderes de proyectos y otros actores estratégicos involucrados en la implementación de IA en salud. Estos marcos suelen abordar dimensiones como la equidad y la diversidad (UNESCO)(1), la gobernanza y la transparencia (AMA)(2,3), y la evaluación del impacto social (UNESCO, AMA)(1–3). Estos marcos son útiles y ofrecen instrumentos para diseñar, hacer seguimiento y evaluar los proyectos de IA en salud. El de UNESCO, por ejemplo, propone una serie de preguntas para orientar el propósito del proyecto, su público objetivo, su alineación con necesidades de las comunidades y su impacto.(1) Este marco, aunque no es específico para el sector salud, ofrece una guía de preguntas que permite analizar si la IA es segura y aporta valor a la atención antes de su implementación, así como identificar y mitigar tanto impactos positivos como negativos. El marco de AMA, por su parte, incluye una pequeña lista para preguntarse si el sistema de IA implementado mejora los resultados de salud de los pacientes.(2,3) Asimismo, proporciona una lista de verificación para planificar, desarrollar, implementar y monitorear el proyecto. Este checklist no solo permite organizar el trabajo de manera eficiente, sino que establece con claridad los roles y responsabilidades de cada actor involucrado, asegurando un esquema de gobernanza sólido.

Por otro lado, los marcos con foco en el componente técnico están dirigidos principalmente a desarrolladores, científicos de datos y profesionales involucrados en la implementación y mantenimiento de soluciones de IA en salud. Estos marcos se centran en la calidad de los datos (BDSLab)(4), la robustez de los algoritmos (Solanki, Grundy y Hussain)(5) y el cumplimiento de estándares técnicos y de seguridad (BDSLab, UE)(4). Para ello, proporcionan herramientas como matrices de evaluación técnica, ejemplos de código abierto (BDSLab)(4), metodologías para la mitigación de sesgos (Solanki, Grundy y Hussain)(5) y controles de calidad (BDSLab, UE) para asegurar que los sistemas de IA sean confiables y efectivos en su funcionamiento clínico.

PERSPECTIVA DE GÉNERO

En relación a la perspectiva de género, se observa que la literatura relevada hace referencia a la importancia de evaluar los datos utilizados para evitar la perpetuación de desigualdades y la reproducción de sesgos. Si se utiliza la Escala GRAS (Gender Responsive Assessment Scale)(15) para el análisis de los marcos, se podría decir que todos incluyen un enfoque sensible al género ya que reconocen las normas, roles y relaciones de género, aunque en la mayoría de los casos no se desarrollan acciones específicas para abordar al género.

Además, es importante mencionar que el género se aborda como un factor dentro de los principios éticos de Equidad y Diversidad. Esto puede diluir su análisis o abordaje específico al estar considerando otros factores dentro de estos principios éticos, como el nivel

socioeconómico, la zona geográfica de residencia y la etnia. De hecho, al analizar las herramientas prácticas de los marcos en relación con la perspectiva de género, se observa que, en el checklist de la AMA(2,3) (el cual se guía por los valores de equidad, evidencia y ética), los aspectos relacionados con el género no se abordan de manera explícita sino que pueden inferirse a partir de ítems como “Desarrollar un protocolo claro para identificar y corregir posibles sesgos” y “Asegurarse de que se hayan considerado y abordado los problemas éticos identificados tanto en el momento de la adquisición como durante el uso”. De manera similar, en el marco de Solanki, Grundy y Hussain(5), los ítems tales como “Aplicar técnicas de prevención de la discriminación en los datos para abordar la discriminación directa e indirecta” e “Identificar y seleccionar herramientas, modelos o marcos de trabajo de limpieza de datos adecuados a su contexto, dominio o necesidades” son los que refieren al principio de Justicia y Equidad. Esto es una limitación significativa porque no se observa un abordaje explícito de género.

05. RECOMENDACIONES

A partir de las fortalezas y limitaciones identificadas, en este apartado se proponen una serie de recomendaciones. A través de cada recomendación se pretende capitalizar las fortalezas de los marcos, como la claridad en la asignación de roles, la interdisciplinariedad y las herramientas técnicas concretas, integrándose en soluciones aplicables y transformadoras. Además, se incluyen propuestas para enriquecer el diseño de marcos prácticos de evaluación.

La evaluación de la IA en salud debe ser integral, considerando tanto su diseño conceptual como su desempeño técnico y su impacto social. Un análisis riguroso garantiza que estas tecnologías sean seguras, equitativas y éticas.

A partir de los marcos revisados, se proponen tres recomendaciones clave: evaluar la IA de lo general a lo específico, asegurar equipos interdisciplinarios con roles definidos y establecer un monitoreo continuo que incluya la perspectiva de género. Estas estrategias permiten una implementación más responsable y efectiva.

1. EVALUAR EL SISTEMA DE IA DESDE LO GENERAL A LO ESPECÍFICO

Para garantizar una evaluación integral de la IA en salud, se recomienda comenzar con una evaluación conceptual amplia y avanzar hacia análisis más detallados y técnicos. Este enfoque permite identificar el propósito del sistema, su alineación con valores éticos y su potencial impacto social antes de evaluar aspectos más específicos como la robustez técnica y la calidad de los datos.

Ejemplo: El marco de la UNESCO propone una evaluación de impacto ético que comienza con la definición del proyecto y su propósito, seguido por un análisis de proporcionalidad de beneficios y riesgos. Posteriormente, se analiza la gobernanza del sistema y su cumplimiento con principios éticos clave.

Herramientas aplicables: Listas de verificación de impacto ético (UNESCO), guías de evaluación estructurada (AMA), y matrices de gobernanza (UNESCO, AMA).

2. EVALUAR LA COMPOSICIÓN INTERDISCIPLINARIA DEL EQUIPO, LOS ROLES NECESARIOS Y LAS RESPONSABILIDADES

El desarrollo y evaluación de sistemas de IA en salud requiere la participación de equipos interdisciplinarios que incluyan expertos en ética, medicina, informática, sociología y derecho. La definición clara de roles y responsabilidades es clave para garantizar la rendición de cuentas y mitigar riesgos asociados a la implementación de IA.

Ejemplo: El marco de la AMA enfatiza la necesidad de establecer roles específicos para desarrolladores, reguladores y profesionales de la salud, asegurando que cada actor tenga claridad sobre sus responsabilidades en el diseño, implementación y monitoreo de la IA.

Herramientas aplicables: Modelos de gobernanza multi-stakeholder (UNESCO), listas de verificación de roles y responsabilidades (AMA) y guías de conformación de equipos diversos (Solanki, Grundy y Hussain).

3. IMPLEMENTAR UNA EVALUACIÓN CONTINUA PARA GARANTIZAR UN USO ÉTICO, RESPONSABLE Y CON PERSPECTIVA DE GÉNERO

La evaluación de la IA no debe limitarse a su diseño inicial, sino que debe ser un proceso continuo a lo largo de su implementación y uso. Se recomienda establecer mecanismos de monitoreo y auditoría periódicos que permitan detectar y corregir sesgos, errores y efectos adversos que puedan surgir con el tiempo.

Ejemplo: El marco de BDSLab propone la implementación de controles de calidad continuos, incluyendo auditorías de datos y modelos, mientras que el marco de Solanki, Grundy y Hussain recomienda la adopción de metodologías de mitigación de sesgos durante todo el ciclo de vida del sistema.

Herramientas aplicables: Auditorías periódicas de equidad y transparencia (BDSLab), herramientas de mitigación de sesgos (Solanki, Grundy y Hussain) y matrices de evaluación técnica (BDSLab).

DIMENSIONES SUGERIDAS PARA NUEVOS MARCOS DE REFERENCIA

En base a estas recomendaciones y al análisis de los marcos, encontrarán en esta sección dimensiones sugeridas para la generación de un marco que integre las fortalezas de cada marco. Se pueden proponer una condensación entre dimensiones a partir del análisis de los marcos relevados. Como resultado, proponemos estas dimensiones:

1. **GOBERNANZA Y DISEÑO DEL PROYECTO:** esta dimensión establece la base del desarrollo de la IA en salud, asegurando que su propósito, población objetivo y contexto de aplicación sean claros y alineados con principios éticos y valores del sistema de salud. También se enfoca en la estructura de gobernanza del proyecto, la rendición de cuentas y la participación de múltiples actores en el proceso de toma de decisiones.

   - Elementos clave:
     - **Descripción del proyecto:** objetivos, población objetivo, necesidades a abordar, justificación del uso de IA.
     - **Gobernanza del proyecto:** definición de roles, responsabilidades y procesos de toma de decisiones dentro del equipo.
     - **Evaluación de proporcionalidad:** análisis de riesgos y beneficios, evitando aplicaciones innecesarias o desproporcionadas.
     - **Gobernanza multiactor:** inclusión de partes interesadas (pacientes, profesionales, reguladores) para asegurar una implementación equitativa y contextualizada.

2. **DESARROLLO TÉCNICO Y ÉTICO DE LA IA:** ESTA dimensión integra la calidad técnica del desarrollo con la evaluación ética del modelo, asegurando que la IA sea confiable, segura y libre de sesgos discriminatorios.

   - Elementos clave:
     - **Privacidad y gobernanza de datos:** protección de datos personales y cumplimiento normativo en el manejo de información.
     - **Diversidad, equidad y no discriminación:** evaluación y mitigación de sesgos en datos, modelos y resultados.
     - **Transparencia y explicabilidad:** implementación de mecanismos que permitan comprender cómo funciona la IA y garantizar decisiones auditables.
     - **Supervisión y determinación humana:** definición de límites claros sobre el papel de la IA en la toma de decisiones clínicas, asegurando intervención humana cuando sea necesario.
     - **Seguridad y robustez técnica:** implementación de medidas para garantizar la confiabilidad y estabilidad del sistema en diferentes escenarios.

3. **IMPLEMENTACIÓN, EVALUACIÓN Y SEGUIMIENTO:** esta dimensión abarca el proceso de integración de la IA en los sistemas de salud y su monitoreo continuo para asegurar que cumpla con sus objetivos y no genere efectos no deseados.

   - Elementos clave:
     - **Evaluación de impacto en salud:** Identificación de efectos positivos y negativos de la IA en la atención médica y en la experiencia de pacientes y profesionales. (AMA, UNESCO)

- **Supervisión y monitoreo continuo:** Implementación de mecanismos de control para detectar fallas y mejorar el sistema con base en evidencia real. (Solanski, AMA)
- **Gestión de riesgos y corrección de errores:** Procedimientos para manejar errores del sistema y establecer procesos de actualización. (BDSLab, Solanki)
- **Sostenibilidad:** Evaluación del impacto ambiental, financiero y operativo del sistema de IA en el largo plazo. (UNESCO)

4. **RENDICIÓN DE CUENTAS Y RESPONSABILIDAD ÉTICA:** esta dimensión garantiza que todas las partes involucradas cumplan con estándares éticos y que existan mecanismos claros de rendición de cuentas.

   - Elementos clave:
     - **Claridad en roles y responsabilidades:** Checklist de tareas asignadas a cada actor (desarrolladores, profesionales de salud, reguladores). (AMA, UNESCO)
     - **Evaluación de equidad y acceso:** Análisis de cómo la IA impacta en grupos poblacionales vulnerables y estrategias para reducir desigualdades. (AMA, UNESCO)
     - **Concienciación y alfabetización en IA:** Capacitación de profesionales de la salud y pacientes sobre el uso y limitaciones de la IA en salud. (UNESCO)
     - **Mecanismos de rendición de cuentas:** Procesos de auditoría y transparencia que garanticen el cumplimiento de principios éticos. (UNESCO, AMA)

06. CONCLUSIONES

Se detectaron cuatro marcos que incluyen herramientas prácticas, como listas de verificación, preguntas orientadoras y ejemplos de código abierto, que dentro de sus dimensiones consideran una mirada de género. Sin embargo, ninguno de ellos establece una dimensión específica dedicada exclusivamente al análisis y abordaje de género. Esto pone de manifiesto una brecha importante en la forma en que los marcos actuales tratan esta temática, limitando su capacidad para abordar de manera profunda y transformadora las inequidades de género en el ámbito de la inteligencia artificial.

La evaluación y diseño de marcos para sistemas de IA en salud requieren un enfoque integral que combine principios generales y soluciones específicas. Este documento ha destacado la importancia de partir de bases éticas sólidas, como las propuestas por el marco de la UNESCO, para luego incorporar elementos técnicos y prácticos que aseguren un desarrollo responsable y efectivo.

Un aspecto clave identificado es la necesidad de integrar la perspectiva de género como una dimensión central y operativa, asegurando que las tecnologías no solo eviten perpetuar desigualdades, sino que contribuyan activamente a transformarlas. Además, la flexibilidad y escalabilidad de los marcos emergen como características esenciales para su implementación en diversos contextos, garantizando que sean útiles tanto en entornos avanzados como en aquellos con recursos limitados.

Finalmente, el monitoreo continuo y la actualización regular de los marcos son imprescindibles para mantener su relevancia frente a los rápidos avances tecnológicos y los cambios en las expectativas sociales. La colaboración interdisciplinaria y la asignación clara de roles completan este enfoque integral, asegurando que los sistemas de IA en salud sean éticos, inclusivos y efectivos. Estas recomendaciones y las dimensiones sugeridas proporcionan una hoja de ruta para el desarrollo y la implementación de marcos que respondan a las complejidades y demandas actuales en el campo de la salud.

REFERENCIAS

[1] United Nations Educational, Scientific and Cultural Organization (UNESCO). Ethical impact assessment: a tool of the Recommendation on the Ethics of Artificial Intelligence [Internet]. 2023 [cited 2024 Dec 16]. Available from: https://unesdoc.unesco.org/ark:/48223/pf0000386276

[2] Crigger E, Reinbold K, Hanson C, Kao A, Blake K, Irons M. Trustworthy augmented Intelligence in health care. J Med Syst [Internet]. 2022 Jan 12 [cited 2025 Jan 21];46(2):12. Available from: https://link.springer.com/article/10.1007/s10916-021-01790-z

[3] American Medical Association. Advancing health care AI through ethics, evidence and equity [Internet]. American Medical Association. 2022 [cited 2024 Dec 19]. Available from: https://www.ama-assn.org/practice-management/digital/advancing-health-care-ai-through-ethics-evidence-and-equity

[4] de Manuel Vicente C, Fernández Narro D, Blanes Selva V, García Gómez JM, Sáez C. A development framework for trustworthy Artificial Intelligence in health with example code pipelines [Internet]. medRxiv. 2024 [cited 2024 Dec 16]. p. 2024.07.17.24310418. Available from: https://www.medrxiv.org/content/10.1101/2024.07.17.24310418v1.abstract

[5] Solanki P, Grundy J, Hussain W. Operationalising ethics in artificial intelligence for healthcare: a framework for AI developers. AI Ethics [Internet]. 2023 Feb [cited 2024 Dec 16];3(1):223–40. Available from: https://link.springer.com/article/10.1007/s43681-022-00195-z

[6] Solomonides AE, Koski E, Atabaki SM, Weinberg S, McGreevey JD, Kannry JL, et al. Defining AMIA's artificial intelligence principles. J Am Med Inform Assoc [Internet]. 2022 Mar 15;29(4):585–91. Available from: http://dx.doi.org/10.1093/jamia/ocac006

[7] Frigerio I, Rashidian N. Artificial intelligence for equity. Art Int Surg [Internet]. 2023;3(3):160–2. Available from: http://dx.doi.org/10.20517/ais.2023.22

[8] Dignum V. Taking Responsibility. In: Responsible Artificial Intelligence [Internet]. Cham: Springer International Publishing; 2019 [cited 2025 Jan 21]. p. 47–69. Available from: https://link.springer.com/chapter/10.1007/978-3-030-30371-6_4

[9] He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med [Internet]. 2019 Jan 7 [cited 2024 Sep 27];25(1):30–6. Available from: https://www.nature.com/articles/s41591-018-0307-0

[10] OECD, CAF Development Bank of Latin America. Acciones para desarrollar un abordaje responsable, fiable y centrado en el ser humano [Internet]. OECD. 2022 [cited 2025 Jan 21].

Available from: https://www.oecd.org/es/publications/uso-estrategico-y-responsable-de-la-inteligencia-artificial-en-el-sector-publico-de-america-latina-y-el-caribe_5b189cb4-es/full-report/component-7.html

[11] Basu K, Sinha R, Ong A, Basu T. Artificial Intelligence: How is It Changing Medical Sciences and Its Future? Indian journal of dermatology [Internet]. 2020 Sep [cited 2025 Apr 21];65(5). Available from: http://dx.doi.org/10.4103/ijd.IJD_421_20

[12] Martin Saban, Santiago Esteban, Katherine Pérez-Acuña, Adolfo Rubinstein, Cintia Cejas. El impacto de la inteligencia artificial en la atención de la salud. Perspectivas y enfoques para América Latina y el Caribe [Internet]. 2023. Available from: https://clias.iecs.org.ar/wp-content/uploads/2023/07/DT1_CLIAS.fix_.pdf

[13] Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol [Internet]. 2005 Feb;8(1):19–32. Available from: http://dx.doi.org/10.1080/1364557032000119616

[14] European Comission. Ethics guidelines for trustworthy AI [Internet]. Shaping Europe's digital future. 2019 [cited 2025 Jan 21]. Available from: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[15] International Development Research Centre. Transforming gender relations - insights from IDRC research [Internet]. 2019. Available from: https://idl-bnc-idrc.dspacedirect.org/bitstreams/712e6482-7131-4bef-b5bd-cfcdc78e6f1e/download

CLIAS

CENTRO DE INTELIGENCIA
ARTIFICIAL Y SALUD
PARA AMÉRICA LATINA
Y EL CARIBE

CIPS
IMPLEMENTACIÓN E INNOVACIÓN EN POLÍTICAS DE SALUD

IECS
INSTITUTO DE EFECTIVIDAD CLÍNICA Y SANITARIA