Universidad Andina Simón Bolívar

Sede Ecuador

Área de Gestión

Maestría en Gerencia Integrada de la Calidad e Innovación

Impacto de la inteligencia artificial en el sector salud

Un análisis bibliográfico de gestión de riesgos con enfoque en el Sistema de Gestión ISO/IEC 42001:2023

Christian Raúl Valdivieso Meza

Tutor: Ricardo Romero

Quito, 2025

Cláusula de sesión de derecho de publicación

Yo, Christian Raul Valdivieso Meza, autor del trabajo intitulado “Impacto de la inteligencia artificial en el sector salud: Un análisis bibliográfico de gestión de riesgos con enfoque en el Sistema de Gestión ISO/IEC 42001:2023”, mediante el presente documento dejo constancia de que la obra es de mi exclusiva autoría y producción, que la he elaborado para cumplir con uno de los requisitos previos para la obtención del título de Magíster en Gerencia Integrada en la Calidad e Innovación en la Universidad Andina Simón Bolívar, Sede Ecuador.

1. Cedo a la Universidad Andina Simón Bolívar, Sede Ecuador, los derechos exclusivos de reproducción, comunicación pública, distribución y divulgación, durante 36 meses a partir de mi graduación, pudiendo por lo tanto la Universidad, utilizar y usar esta obra por cualquier medio conocido o por conocer, siempre y cuando no se lo haga para obtener beneficio económico. Esta autorización incluye la reproducción total o parcial en los formatos virtual, electrónico, digital, óptico, como usos en red local y en internet.
2. Declaro que, en caso de presentarse cualquier reclamación de parte de terceros respecto de los derechos de autor/a de la obra antes referida, yo asumiré toda responsabilidad frente a terceros y a la Universidad.
3. En esta fecha entrego a la Secretaría General, el ejemplar respectivo y sus anexos en formato impreso y digital o electrónico.

01 de septiembre de 2025

Firma: _____

Resumen

Desde una perspectiva analítica, la intención de esta investigación es explorar y describir el impacto de la inteligencia artificial (IA) en el sector salud. El trabajo se focaliza en la identificación de desafíos éticos que surgen a lo largo del ciclo de vida de algunas aplicaciones de IA, desde su diseño y desarrollo hasta su implementación.

Para lograrlo, se establece un marco teórico que familiarice al lector con los conceptos fundamentales de la IA y sus aplicaciones en el ámbito de la salud. Además, se contextualiza el desarrollo de esta tecnología a nivel global y, de manera específica, en América Latina, haciendo un énfasis particular en Ecuador. Se constata, a su vez, que los aspectos legales y normativos son una instancia fundamental de mediación en este proceso.

Se destaca la importancia de la norma ISO/IEC 42001:2023 y otras normativas ISO de soporte como herramientas clave para la gestión ética y responsable de los sistemas de IA por parte de las organizaciones, reconociendo también la relevancia de marcos normativos complementarios, como los elaborados por el NIST, para fortalecer la seguridad y confiabilidad de la IA en entornos clínicos.

Posteriormente, esta investigación explora conceptos éticos como la responsabilidad y la confiabilidad, apoyándose en un marco propuesto por la OMS. Se utiliza esta estructura para validar los desafíos éticos identificados en áreas de la salud como el diagnóstico automático, el diagnóstico por imagen, la salud mental y el desarrollo de medicamentos.

En la última parte de este análisis, se presentan recomendaciones elaboradas con base en la norma ISO/IEC 42001:2023 para mitigar dichos desafíos. Además, se detallan mecanismos que promueven una implementación segura, equitativa y confiable de la IA en el ámbito sanitario.

Esta investigación no solo busca visibilizar la importancia de que las organizaciones que desarrollan o utilizan la IA actúen en conformidad con valores éticos, sino que también resalta la necesidad de que el país cuente con instituciones regulatorias robustas. Estas instituciones deben ser inclusivas para generar confianza, permitir un desarrollo ético de la IA en el sector de la salud y convertirse en catalizadores en el aprovechamiento de esta tecnología para el desarrollo del país.

Palabras clave: aprendizaje de máquina, aprendizaje profundo, ciclo de vida de la IA, redes neuronales, sesgo, explicabilidad.

A mi amada hija Nathy, por ser el motor de mi vida e inspirarme para mejorar cada día.

A mi amada Sofy, por su paciencia y apoyo en este reto.

A mi madre, por brindarme siempre su amor y ayuda incondicional.

A mis pequeños, por acompañarme pacientemente en el desarrollo de este trabajo.

A mi nube, quien con su alegría hacia más llevadero el esfuerzo y con su partida dejó un gran vacío.

Tabla de contenidos

Introducción

En la coyuntura actual, muchos de los problemas sociales se derivan de la falta o distorsión de los valores éticos en las esferas de la colectividad. Por lo tanto, las acciones u omisiones de las organizaciones pueden acarrear repercusiones profundas y de largo alcance. En este sentido, las tecnologías emergentes plantean desafíos éticos significativos que si no son abordados adecuadamente podrían dar lugar a mayor desigualdad y descontento. En consecuencia, la ética adquiere un valor fundamental tanto para las empresas como para las personas que desarrollan y utilizan estas tecnologías (Grover y Dogra 2024, 3).

Entre las tecnologías emergentes, las herramientas de IA están marcando un cambio radical en varias áreas del saber humano, principalmente en sectores como la salud, en donde la implementación de esta tecnología representa un factor de innovación tremendamente significativo en la atención médica debido a su potencial de revolucionar desde la gestión hospitalaria hasta la forma de diagnóstico y tratamiento de enfermedades.

Sin embargo, todo este potencial de la IA trae consigo un bagaje de retos y consideraciones éticas que deben ser abordadas meticulosamente como la privacidad y seguridad de los datos de los pacientes, el sesgo algorítmico que puede afectar la equidad en la atención, la necesidad de garantizar la transparencia de los sistemas, etc. (Persons y MCGinnis 2019, 7).

En este dilema se proponen en el contexto mundial varias normativas, entre ellas el estándar internacional ISO/IEC 42001:2023 que establece un marco para la gestión de sistemas de inteligencia artificial (IA) en las organizaciones, proporcionando directrices que garantizan la transparencia, la responsabilidad y la ética en el uso de tecnologías de IA en sus etapas de desarrollo, implementación o uso.

Aunque el aspecto legal de la IA en diversos países aun es incipiente, se prevé que cada uno adoptará enfoques diferentes en sus regulaciones locales en el futuro. Por ello, resulta relevante abordar esta cuestión desde la óptica de un estándar de reconocimiento global, como lo es el ISO/IEC 42001: 2023. Esta norma, al ser la primera certificable en su tipo, no solo facilita la estandarización, sino que también promueve la confianza en las soluciones de inteligencia artificial a nivel global.

La motivación de esta investigación surge a partir del crecimiento exponencial en la utilización de tecnologías de inteligencia artificial en el ámbito de la salud en el Ecuador. Fenómeno que el autor observa de manera frecuente en su desempeño profesional como ingeniero biomédico. Es notable que muchas de estas herramientas se utilizan vía online sin ningún tipo de regulación, y otras tantas se adquieren principalmente por sus características innovadoras y como estrategia de marketing institucional más que por una evaluación real de las necesidades clínicas. Además, en numerosos casos, su implementación se realiza sin un sistema de gestión adecuado que garantice un uso responsable y seguro. A esto se suma que, frecuentemente, las bases de datos étnicas que utilizan ciertos sistemas médicos no corresponden a la región. Esto plantea la necesidad de investigar y proponer lineamientos que aseguren la integración efectiva y ética de estas tecnologías en el ámbito de la salud.

A partir de lo expuesto, surge la pregunta: ¿Que desafíos éticos derivan en la implementación de ciertas herramientas de IA en el área de la salud como: el diagnóstico automático, diagnóstico por imagen, salud mental y desarrollo de medicamentos y como gestionarlos en el marco de la normativa ISO/IEC 42001:2023?

Para responderla se han establecido los siguientes objetivos específicos:

Definir un marco de referencia que integre aspectos teóricos, normativos y legales relacionados con la Inteligencia Artificial. Esto incluirá la conceptualización de las principales técnicas de IA, el análisis de las normativas pertinentes para su gestión e implementación, y la comprensión del marco legal que las regula.

Analizar y describir ciertas herramientas de IA en las siguientes áreas: diagnóstico automático, diagnóstico por imagen, salud mental y desarrollo de medicamentos e identificar los desafíos éticos y limitaciones que derivan de su implementación.

Establecer una relación entre los desafíos éticos y limitaciones descritos en la implementación de sistemas de IA con un conjunto de recomendaciones fundamentadas en los criterios de la normativa ISO/IEC 42001:2023, así como de otras normativas relacionadas, como ISO 23053, ISO 23894 e ISO/IEC 27001.

Para alcanzar estos objetivos, se realiza una investigación exploratoria que consiste en una revisión bibliográfica de enfoque cualitativo a través de bases de datos especializadas y buscadores académicos con el fin de recopilar y analizar información relevante proveniente de artículos científicos, publicaciones biomédicas, publicaciones de organismos

internacionales y documentos normativos con el fin de obtener una visión holística sobre el estado del arte de la IA en las áreas de salud propuestas. Cada objetivo específico se desarrolla dentro de un capítulo, así:

El primer capítulo plantea los aspectos teóricos de la IA desde su historia, clasificación y áreas de aplicación, hasta el impacto de la IA tanto en la gestión de las organizaciones como en el sector salud. En este capítulo también se exploran los avances de la IA a nivel mundial, principalmente en el Ecuador y en la región latinoamericana. Además, se propone una guía de acciones para que el país incremente su participación en el desarrollo y uso adecuado de esta tecnología. En el marco legal se abordan las propuestas de ley de inteligencia artificial en el país y la ley de protección de datos personales Por otro lado, en el marco normativo se describe la norma ISO/IEC 42001:2023 para gestión de la IA, la norma ISO/IEC 27001:2022 de seguridad de la información y otras normativas de soporte. Por último, se aborda la ética en la IA mediante un marco de principios en el área de la salud definidos por la OMS y con él se describen las características que otorgan confiabilidad a estos sistemas.

El segundo capítulo describe las principales aplicaciones de la IA en las áreas propuestas, identifica sus desafíos éticos particulares y describe un marco de gestión general para estos desafíos.

El tercer capítulo relaciona los desafíos éticos particulares identificados, con los principios éticos sugeridos por la OMS para posteriormente proponer recomendaciones basadas en la norma ISO/IEC 42001:2023 y otras normativas de soporte para su gestión en el ciclo de vida de las aplicaciones de IA que comprenden el diseño, desarrollo, implementación y operación.

Por otra parte, entre las conclusiones planteadas en esta investigación se destaca la corresponsabilidad que existe entre desarrolladores, usuarios y entes reguladores para la gestión eficiente de la inteligencia artificial. Se enfatiza la importancia de asumir la gestión de los desafíos éticos en todo el ciclo de vida de la IA y mantener canales de comunicación entre las partes involucradas con el fin de garantizar la seguridad, ética y precisión en estas herramientas.

Por último, en el Anexo 2 se busca ofrecer una visión contextual de la implementación de la norma ISO/IEC 42001:2023 en una organización cuyo objetivo es reducir el sesgo en

sus aplicaciones de IA. Para ello, se propone en primer lugar un análisis FODA que describe el contexto organizacional, seguido por la identificación de las partes interesadas y sus requisitos relevantes. Posteriormente, se definen los roles, responsabilidades y autoridades dentro de la organización. Finalmente, se incluye una matriz de riesgos que vincula a las partes interesadas con ciertas medidas de mitigación.

Capítulo primero
Aspectos teóricos, legales y normativos de la inteligencia artificial

Para comprender en forma integral el impacto de esta innovación, es fundamental establecer un marco teórico que permita comprender la historia, evolución y clasificación de la IA, ilustrando su uso mediante ejemplos concretos especialmente en el área de la salud. Esto permitirá contextualizar desde el principio la aplicación de estas tecnologías en la práctica clínica y complementar la discusión posterior sobre sus implicaciones éticas.

1. Aspectos teóricos

1.1. Definición de inteligencia artificial

La inteligencia artificial (IA) es una rama de las ciencias de la computación y de la ingeniería cuyas aplicaciones cumplen tareas propias de la inteligencia humana por medio de recursos computacionales relacionados en estructuras con diversos enfoques y técnicas (Bekdaş, Nigdeli, y Yücel 2020, 3–12), otros autores definen a la IA como un sistema capaz de aprender de datos externos y usar este aprendizaje para realizar tareas específicas mediante una adaptación flexible (Garzón Espitia y Rodríguez Padilla 2022, 4).

Los sistemas de IA generan contenido, predicciones, decisiones o recomendaciones hacia objetivos establecidos previamente por humanos, es decir, requieren diseño y supervisión en cada etapa y operan bajo un marco de referencia establecido por humanos (ISO/IEC 2022c, 41). De acuerdo a la norma ISO/IEC 22989:2022(2022c, 41) el grado de vigilancia humana en los sistemas de IA está presente al menos durante su entrenamiento y validación para asegurarse de los posibles impactos que genere en su ciclo de vida. En consecuencia, según Álvarez-Maldonado et al. (2024a, 2) las herramientas de IA están creando nuevas condiciones sociales de convivencia entre humanos y agentes no humanos que interactúan en la toma de decisiones para el logro de objetivos, creando un nuevo agente social artificial que afecta las normativas institucionales y organizacionales hacia un cambio cultural inexorable.

Es necesario acotar que según la norma ISO/IEC 23053:2022 (2022b, 15, 28) la utilización de términos como conocimiento y aprendizaje en el área de IA no pretende atribuir a estos sistemas de características humanas ya que se trata de herramientas funcionales y no seres pensantes.

1.2. Historia y evolución de la inteligencia artificial

La era de la información, también llamada era digital o informática ha marcado el desarrollo vertiginoso de tecnologías informáticas y de comunicación, según Gordon Moore la cantidad de transistores en un chip se duplicaría aproximadamente cada 18 meses (Cheang Wong 2005, 4), esta tendencia ha sido clave en la miniaturización de procesadores y el consecuente desarrollo de la inteligencia artificial.

La inteligencia artificial no es un concepto nuevo, Allan Turing propone en la década de los 50 en su estudio *Computing machinery and intelligence* una prueba para determinar si las máquinas eran capaces de pensar (González 2007, 18). En esta prueba denominada como "Test de Turing", participan un juez, un ser humano y una máquina. Se considera que la máquina demuestra inteligencia si logra confundir al juez haciéndole creer que es un ser humano (Marketing Departamento 2021). Sin embargo, en la práctica no se implementaban estos algoritmos debido a las ingentes cantidades de recursos computacionales que demandan.

El proceso de conexión de neuronas en el cerebro humano desarrollado por Donald Hebb con la teoría hebbiana en 1949 abrió las puertas para la investigación de la IA (Trillo 2023, 14). Sin embargo, el término Inteligencia Artificial se empieza a utilizar desde el proyecto de John McCarthy en 1956 en donde se propuso que "cada aspecto del aprendizaje o cualquier característica de la IA puede describirse con tal precisión que una máquina puede simularlo", este proyecto no llegó a un consenso pero generó interés significativo en el tema (Moor 2006, 6).

La programación tradicional con la que en un principio se enfocó el desarrollo de aplicaciones para emular el pensamiento humano permitía procesar grandes cantidades de información y generar respuestas pero su estructura no poseía características de aprendizaje autónomo ni evolución (Trillo 2023, 5), es en los años 80 que empiezan a surgir las redes neuronales y con ellas el desarrollo de sistemas de IA (Trillo 2023, 6).

Sin embargo, incluso a inicios de este siglo las máquinas no lograban confundir al juez en la prueba de Turing. Es en el año 2011 que surge un punto de inflexión con el supercomputador de Watson diseñado por IBM que muestra capacidad de aprender mientras trabaja, reúne información e interactúa con el lenguaje humano, este hito fue una de las claves en el desarrollo del aprendizaje profundo (Marketing Departamento 2021).

Desde la década pasada se han alcanzado importantes hitos en la evolución de la IA , un ejemplo de esto es el desarrollo del programa AlphaGo que fue capaz de vencer al campeón mundial del juego de mesa Go (Trillo 2023, 7) o la predicción de las estructuras 3D de casi cualquier proteína del cuerpo humano por medio de la herramienta Alphafold ,este avance acelera la investigación de medicinas y vacunas para enfermedades como el Alzheimer (McDonald 2024, 5), ambas aplicaciones fueron desarrolladas por Google Deepmind.

En 2018 se vende en una subasta la imagen llamada “Edond de Belamy” por $432500 generada por IA a partir de un entrenamiento con antiguos retratos del siglo 18 y 19, este evento abre un debate ético sobre el rol de los humanos en actividades creativas frente a la IA (McDonald 2024, 5) y sobre desafíos legales en relación a los derechos de autor y la originalidad de estas obras (Benítez Vargas, Rico Duarte, y Niño Hernández 2023, 7).

El último avance en inteligencia artificial se centra en la democratización de su uso, considerando que anteriormente estaba limitado a supercomputadoras como la IBM Deep Blue. Ahora, sofisticadas herramientas de IA con interfaces amigables, como ChatGPT de OpenAI, Google Bard, IBM Watsonx.ai y Microsoft Copilot, son accesibles al usuario desde cualquier computador de rendimiento promedio (McDonald 2024, 7). Sin embargo, el acceso a la tecnología pone de manifiesto aún más la necesidad de asegurar su uso responsable y ético (Jauregui Moran 2024, 3) con el desarrollo de marcos regulatorios claros mediante la colaboración de académicos, gobiernos y empresas (Jauregui Moran 2024, 8).

1.3. Clasificación de la Inteligencia Artificial

La inteligencia artificial es un campo vasto y en constante evolución, lo que impide la existencia de una clasificación universalmente aceptada. El hecho de que las máquinas simulen procesos de pensamiento ha dado lugar a la clasificación por capacidad de la IA en fuerte o general y débil o estrecha. La IA General es aquella capaz de entender, realizar tareas

y aprender de una forma similar al proceso cognitivo del ser humano, esta categoría actualmente es teórica y representa un objetivo ambicioso en el campo de investigación; mientras que la IA estrecha abarca todas las técnicas de IA que se usan en la actualidad, la ISO/IEC 22989 (2022c, 17) la describe como sistemas capaces de resolver tareas definidas e identificar problemas específicos, estos sistemas parecen inteligentes pero aún no poseen habilidades de razonamiento más allá de su entrenamiento o programación particular.

Por otro lado, según su funcionalidad, la IA se puede clasificar en: sistemas expertos, IA predictiva e IA generativa (McDonald 2024, 22).

Los sistemas expertos imitan el proceso de razonamiento humano para resolver problemas específicos. Estos pueden ser utilizados para mejorar habilidades en la resolución de problemas o para actuar como asistentes en una variedad de tareas (Badaro, Ibañez, y Agüero 2013, 3), ya que se enfocan en un dominio específico e imitan la capacidad de decisión de un humano experto en ámbitos como: diagnóstico médico, asesoría legal, etc. (Manmeet Kaur 2024, 2). Aunque, según Badaro (2013, 3), estos sistemas pueden desempeñarse mejor que cualquier ser humano en un dominio acotado. El concepto de sistemas expertos nace en los años 60 y se lo materializa con el sistema Dendral, creado por la Universidad de Stanford con el objetivo de inferir en la identificación de estructuras moleculares en compuestos orgánicos desconocidos, tarea que tradicionalmente requería un conocimiento profundo de químicos especializados (Badaro, Ibañez, y Agüero 2013, 5). En otro ámbito, según Mohamad (2004, 2), la división de sistemas de negocio de IBM ha apoyado sus decisiones estratégicas con un sistema experto llamado Business Insight que ayuda a integrar datos cualitativos y cuantitativos en apoyo de la gestión organizacional.

La IA predictiva se basa en modelos estadísticos que estiman probabilidades para identificar patrones, tendencias y relaciones a partir de un entrenamiento con grandes cantidades de datos históricos. Estas herramientas se combinan con otras formas de análisis para formar modelos más poderosos que resuelven problemas complejos, las aplicaciones de este tipo de IA en la medicina son variadas, Kumar y Ramesh (2024, 3–8) detallan algunas de ellas: detección temprana de enfermedades, análisis de imágenes médicas, medicina genómina, etc. Por otro lado, en las organizaciones de acuerdo a Trillo (2023, 22) el análisis predictivo tiene una variedad de aplicaciones como: mejora de tiempos de entrega, predicción de la demanda, personalización del marketing, predicción de fallas de equipos, etc.

Por último, está la IA generativa, esta utiliza modelos de redes neuronales profundas para generar diferentes tipos de información como textos, imágenes, música, etc., también categoriza el contenido para crear contextos, significados o aprender patrones.

La tecnología que antecedió a la IA Generativa aparece en el 2014 con el nombre de redes antagónicas generativas "GAN" de la mano de Ian Goodfellow, la cual era ya capaz de generar imágenes muy similares a fotografías reales (Bussines & Decision 2023, 4), este modelo se basa en dos redes neuronales profundas, la una genera continuamente imágenes no supervisadas y la otra determina su validez (Bussines & Decision 2023, 9). El nombre de IA Generativa aparece en el 2017 en un documento de Google llamado *Atention is all you need* en el cual surgen los indicios de una arquitectura de IA llamada *Transformers* (McDonald 2024, 24), aquí se describe el entrenamiento de modelos de aprendizaje profundo auto supervisado. El *Transformer* funciona con un codificador y un decodificador, se diferencia de sus antecesores en que el modelo analiza diferentes secuencias de la información en forma simultánea (Bussines & Decision 2023, 9). Posteriormente se desarrolló la capacidad para hacer funcionar esta tecnología a través de una plataforma de cálculo controlada por un procesador gráfico utilizando un corpus[1] de entrenamiento (Bussines & Decision 2023, 5). En las secciones siguientes se explorarán diversas aplicaciones de la IA generativa.

1.4. Áreas de aplicación de los sistemas de Inteligencia Artificial

La IA ha avanzado significativamente en la simulación de procesos cognitivos como el razonamiento, la percepción y el aprendizaje, permitiendo su implementación en una variedad de áreas. A medida que se exploran en esta sección ciertas capacidades cognitivas que los sistemas de IA buscan replicar, se detallan algunas aplicaciones y cómo estas contribuyen a resolver problemas complejos y a mejorar la eficiencia de varias disciplinas:

Percepción y reconocimiento de patrones: Esta tarea consiste en el reconocimiento de señales, a veces con ayuda de sensores. La señal obtenida se clasifica, se codifica y se procesa por medio de diferentes modelos que se utilizan en áreas como reconocimiento de caracteres ópticos, sistemas de visión industrial de robots, sistemas de identificación de objetivos

[1] Conjunto de datos utilizados para entrenar modelos de aprendizaje automático, especialmente en el campo del procesamiento del lenguaje natural (NLP) y el análisis de datos. Este corpus puede consistir en texto, audio, imágenes u otro tipo de datos, dependiendo de la tarea específica que se esté abordando.

militares o sistemas de diagnóstico y reconocimiento de imágenes médicas (Flasiński 2016, 224). Por otro lado, en el ámbito empresarial el procesamiento y reconocimiento de patrones permite a las organizaciones analizar tendencias en áreas como: el comportamiento del mercado, satisfacción del cliente, gestión de recursos humanos, eficiencia de procesos, etc. Con el objetivo de tomar decisiones informadas, mejorar la eficiencia y responder de forma más ágil a las necesidades del mercado (Solano Hernández y Soriano Ávila 2024, 12).

Representación de conocimiento: De acuerdo a Jordi Duran (2013, 5) se basa en la codificación del conocimiento humano para ser usado por los algoritmos computacionales de manera organizada y coherente. Por lo tanto, la representación del conocimiento es crucial para que los sistemas de IA puedan tomar decisiones basadas en el mundo real. Los modelos de representación de conocimiento se clasifican según la forma en que lo adquieren, esta representación puede ser explícita, es decir que permita una interpretación clara por parte de los humanos o implícita en donde el conocimiento es encapsulado o codificado. Algunas técnicas como el reconocimiento de patrones y el aprendizaje automático permiten a ciertos modelos de IA adquirir conocimiento sin intervención humana (Flasiński 2016, 224–25). Las herramientas de análisis de datos permiten a las organizaciones en diversas áreas aprovechar el conocimiento explícito e implícito que poseen, gestionando de manera más efectiva la información sobre sus mercados, procesos y clientes o pacientes. Esto les facilita tomar decisiones más informadas y acertadas.

Resolución de problemas: Este concepto busca desarrollar métodos generales de IA que se apliquen a la resolución de una variedad de problemas. Sin embargo, su consecución ha sido esquiva. Los sistemas basados en estrategias de búsqueda solventan de cierta manera el problema pero en cooperación con el diseño humano en su primera fase que se basa en la construcción del modelo abstracto del problema, en donde se definen los estados para la posterior determinación del espacio de estados[2] (Flasiński 2016, 226).

Razonamiento: La IA funciona mejor en las áreas que implican inferir conclusiones en base a ciertas reglas generales. Estos modelos se aplican en negocios, medicina, industria, comunicación, transporte, etc. En este contexto, los modelos de IA que implican

[2] El término *state space* (espacio de estados) se refiere al conjunto completo de todas las configuraciones o estados posibles que un sistema puede asumir durante la solución de un problema particular. Este concepto es fundamental en la representación y resolución de problemas en IA, especialmente en áreas como la planificación, la búsqueda y los algoritmos de solución de problemas.

razonamiento deductivo simulan capacidades humanas mejor que los que implican razonamiento cognitivo. Esto se debe al desarrollo previo de los modelos matemáticos y lógicos en el periodo que precedió el nacimiento de la IA (Flasiński 2016, 227). En el área de la salud, el razonamiento automatizado se utiliza para analizar datos clínicos o interpretar resultados de pruebas y correlacionar síntomas con diagnósticos potenciales con el fin de recomendar tratamientos específicos apoyados en evidencia y reducir el riesgo de errores humanos.

Toma de decisiones: Esta ha sido una de las primeras aplicaciones de la IA, para lograrlo se necesita conocimiento explícito o reglas provenientes de humanos expertos que orienten a los modelos de IA a reconocer futuros escenarios. El objetivo de estos sistemas es facilitar la toma de decisiones en situaciones que involucran múltiples criterios y variables (Zapata Cortés 2020, 3,4). De acuerdo a Flasiński (2016, 227-28), los sistemas de decisiones con patrones de reconocimiento estadístico no proponen una decisión determinística, pero sugieren varias posibilidades. Por su parte, Zapata Cortés (2020, 3) identifica tres tipos de sistemas de soporte de decisiones. Los pasivos que solo organizan información, los activos que sugieren soluciones y los cooperativos que combinan ambos enfoques en el apoyo de decisiones. Una aplicación de estos sistemas en las organizaciones es la predicción de la demanda mediante el análisis de datos históricos, en donde no solo se incluyen datos estructurados sobre ventas sino también datos no estructurados como comentarios en redes sociales y encuestas; esta información permite modelar diferentes escenarios de mercado y prever como la variación de factores como cambios de precio, lanzamiento de nuevos productos, campañas de marketing o incluso variables macroeconómicas podrían incidir.

Planeación: Según Guzmán Luna (2003, 2) la planeación consiste en construir algoritmos que permitan obtener automáticamente planes adecuados para alcanzar un objetivo, la simulación de este proceso es compleja porque requiere elementos de predicción de consecuencias para las distintas acciones y la habilidad para reestructurarse en tiempo real de acuerdo con los distintos escenarios. Un ejemplo es la determinación de secuencia de movimientos de un robot para alcanzar un objetivo, la cual incluye cálculos geométricos, posición, orientación, velocidades, aceleraciones y la presión necesaria a ejercer en el objeto para evitar su deformación, el robot concatena las acciones básicas hasta alcanzar el resultado deseado (Jiménez Schlegl y Torras Genís 2020, 20).

Procesamiento de lenguaje natural (PLN): Takale y su equipo (2024, 1-2) lo definen como la rama de la IA enfocada en permitir que las computadoras comprendan y generen lenguaje humano a nivel semántico, esto incluye: entendimiento o contextualización de texto, traducción de un lenguaje natural a otro, sistemas humano-máquina de comunicación verbal, etc. El punto de partida en esta área es la teoría Chomsky de generación gramatical[3].

A finales del siglo XX se desarrollaron aproximaciones estadísticas para el análisis del lenguaje en donde se utilizó el corpus a manera de entrenamiento para determinar características estadísticas del lenguaje en particular, tales como: contextos en los que las palabras ocurren, estadísticas de uso de palabras, etc. (Flasiński 2016, 229). El análisis semántico enfrenta diversas dificultades, entre ellas los aspectos no verbales del lenguaje, como la entonación y el estrés. En este contexto, la capacidad de comprender el significado correcto de las palabras está estrechamente vinculada a la inteligencia social, un componente que resulta sumamente complicado de replicar en un sistema de inteligencia artificial (Flasiński 2016, 229,230). En el área del cuidado de la salud se utilizan estas técnicas para la extracción y uso de información de registros médicos no estructurados, con el fin de analizarlos y detectar patrones que permitan identificar e incluso anticipar enfermedades (Takale 2024, 2-3). Estos sistemas permiten mejorar el tiempo y esfuerzo necesarios para el manejo de registros e ingreso de información y optimizar la toma de decisiones clínicas. Por otro lado, los asistentes virtuales impulsados por PLN permiten facilitar el acceso a la salud y fomentar la gestión y adherencia del paciente a programas de tratamiento (Takale 2024, 4–6). Además, las organizaciones utilizan el PNL en diversas áreas como: servicio al cliente y soporte técnico, marketing, publicidad, etc.

Aprendizaje: Los modelos de aprendizaje en IA se pueden dividir en generalizadores de experiencia y transformadores de representación del dominio del problema. En los primeros, el set de aprendizaje representa la experiencia que modifica el comportamiento del sistema, por ejemplo, en el aprendizaje no supervisado el modelo se entrena con datos no etiquetados y busca identificar patrones subyacentes sin ninguna referencia explícita, es especialmente útil cuando se busca descubrir nuevas clasificaciones, pero su interpretación puede ser más compleja (Flasiński 2016, 230-231). Por otro lado, la representación del

[3] Se refiere a la capacidad innata de los seres humanos de generar y comprender un número ilimitado de oraciones en base a las reglas subyacentes de una lengua.

dominio del problema es un modelo que debe construir una ontología, o sea la representación estructurada del contexto en el que opera, esto implica la recopilación de datos de su entorno que le permita definir conceptos relevantes que serán estructurados para establecer relaciones semánticas de su interconexión. Su implementación práctica es aún un desafío para los sistemas de IA (Flasiński 2016, 231).

Manipulación y locomoción: La simulación de habilidades de inteligencia cinética es uno de los problemas más complejos en IA porque las habilidades de manipulación y locomoción de robots dependen de la interrelación de sistemas de reconocimiento de patrones, de resolución de problemas o de planeación con sistemas de movimiento o actuadores. Es pertinente mencionar que estos sistemas no siempre buscan imitar las habilidades humanas, sino también replicar las características animales que muestran claras ventajas sobre el ser humano. Además, algunas características de los robots han sobrepasado las habilidades humanas, especialmente en aplicaciones que requieren alta precisión o resistencia (Flasiński 2016, 232-233). Según Pablo Jiménez (2020, 7), el robot proporciona cuerpo a la IA ya que le permite percibir el mundo, moverse y actuar, en consecuencia, ya no es una máquina más, sino que se convierte en una máquina capaz de recordar, aprender, planificar, razonar y tomar decisiones más o menos complejas de acuerdo a su entorno.

Dentro de este campo, en la industria e investigación son muy utilizados los brazos robóticos para realizar manipulación de objetos en procesos industriales repetitivos o en la manipulación de objetos peligrosos. En la medicina, realizan operaciones de cirugía, asisten en tareas rutinarias como la distribución de medicamentos, pueden prestar apoyo a la rehabilitación o prestar asistencia física en el movimiento de pacientes como por ejemplo en las prótesis robóticas, en donde estos sistemas prestan apoyo a ciertos pacientes con pérdida de capacidades físicas originadas por enfermedades adquiridas, accidentes o malformación genética.

Inteligencia Social: Sufyan y su equipo (2024, 1) definen la inteligencia social como la capacidad de entender las emociones, sentimientos y necesidades de las personas. En este contexto, la percepción de habilidades humanas como la expresión facial o connotaciones del lenguaje son aspectos muy difíciles de adquirir y procesar para los modelos de IA. Para lidiar con este problema se han propuesto patrones avanzados de reconocimiento basados en reglas que integran visión y sonido. La investigación en esta área es de vital importancia para

la robótica y sus aplicaciones en medicina, seguridad social, etc. Una investigación realizada por Sufyan (2024, 2–8) sobre sistemas de IA que utilizan inteligencia social en el contexto de la interacción humana y psicoterapia recalca que ciertos modelos LLM[4] como ChatGPT-4 o Bing han adquirido capacidad de comprensión similar a la humana en su respuesta a sentimientos y emociones. El estudio evaluó tanto a modelos de IA como a psicólogos y mostró que estos modelos son capaces de tomar decisiones en base a la valoración de distintos escenarios que requieren comprensión social, aunque la investigación también advierte que la simulación de habilidades sociales no reemplaza las relaciones terapéuticas genuinas y la profunda comprensión de las estrategias de psicoterapia.

Por último, la simulación de la creatividad humana a través de sistemas de IA generativa es un tema polémico y representa un desafío significativo para la IA, puesto que los procesos creativos se consideran intrínsecamente humanos, especialmente en lo que respecta a la creatividad transformacional, que exige un enfoque disruptivo.

Actualmente, los sistemas de inteligencia artificial generativa se emplean en una variedad de aplicaciones, como la generación de música, arte visual. (Flasiński 2016, 233-234), también en el área de la salud en aplicaciones como optimización de fármacos moleculares, educación médica y odontología (Sai et al. 2024, 1–4). La habilidad generativa confiere a estas herramientas de un gran potencial transformador (Bussines & Decision 2023, 4). Pedro Pernías Peco (2024, 2–12) destaca de los modelos generacionales en el ámbito empresarial, la automatización y personalización de contenido que permite la creación de textos, imágenes o videos en base a la identificación de tendencias y preferencias del cliente, esto permite el desarrollo de innovadoras campañas de marketing adaptadas a necesidades específicas de uno o varios segmentos del público objetivo.

1.4.1. Impacto de la IA en la gestión de las organizaciones

La IA está redefiniendo el ámbito empresarial hasta el punto de convertirse en catalizador de cambio en la forma en que las organizaciones operan y compiten en el mercado (Garcia, Carrillo, y Cuellar 2023, 1) y a su vez está constituyendo un nuevo activo intangible que genera valor a largo plazo.

[4]LLM por sus siglas en inglés large language models. Son modelos entrenados con vastos volúmenes de texto.

El impacto de la IA se puede apreciar en casi todas las áreas tecnológicas como fábricas inteligentes, internet de las cosas, comunicación entre máquinas y operaciones productivas sin intervención humana por medio de la generación y análisis de datos , flujos de trabajo, calidad de resultados y actividades de mantenimiento (Álvarez-Maldonado et al. 2024b, 11). En consecuencia, también sirve de apoyo a la gestión clínica en su labor de garantizar la óptima atención a los pacientes teniendo en cuenta un adecuado control de recursos, lo que contribuye a la satisfacción del paciente y a mejores resultados clínicos.

García (2018, 3) argumenta que la implementación de IA en las organizaciones se orienta en general a la reducción de errores, automatización de procesos y optimización de operaciones en toda la cadena de valor, permitiendo a las empresas fundamentar una cultura basada en datos que incide en la mejora de índices de innovación e impulsa la toma de decisiones informadas.

La implementación de herramientas de IA facilita entre otras cosas el pronóstico de entornos dinámicos y altamente competitivos utilizando para esto técnicas de *big data*, cuyos resultados contribuyen a racionalizar las decisiones otrora sujetas a recursos cognitivos limitados, esto favorece a la capacidad de supervivencia del emprendimiento de negocios contemporáneos (Álvarez-Maldonado et al. 2024b, 13). En este sentido, las capacidades de extracción y análisis de grandes volúmenes de datos que poseen ciertas herramientas de IA facilitan a las empresas la toma de decisiones fundamentadas en evidencia (Rubín 2024, 13).

La IA optimiza la automatización de procesos repetitivos basados en reglas, lo cual permite enfocar los recursos de la organización hacia actividades estratégicas, creativas y de innovación (Rubín 2024, 13). Por otro lado, la IA ha demostrado su eficiencia en sistemas de información para la gestión empresarial ERP y CRM que brindan soporte a las actividades operativas, mejorando en forma significativa sus prestaciones y otorgándoles funciones avanzadas como por ejemplo el reabastecimiento automático para mantener niveles de inventario adecuados y la gestión de cadena de suministro optimizada que involucra la coordinación con proveedores y optimización de rutas y tiempos de entrega, además estas herramientas pueden integrarse con tecnologías IoT para brindar funciones de análisis y monitoreo en tiempo real (Solano Hernández y Soriano Ávila 2024, 8-9). Entre las aplicaciones relevantes que utilizan IA en la gestión de inventario y cadena de suministros

Solano Hernández detalla las siguientes: Llamasoft, Kinaxis RapidResponse, ToolsGroup, Slimstock, Oracle Autonomous Database, IBM Watson supply Chain.

Por otro lado, ciertas herramientas de marketing basadas en IA como los algoritmos de Amazon y Netflix o *chatbots* con asistentes virtuales como Zendesk e Intercom contribuyen a la personalización de la experiencia del cliente por medio de análisis de preferencias individuales, lo cual incide en el valor que la empresa entrega al usuario desde la recomendación de productos hasta la atención que le brinda(Rubín 2024, 13).

En cuanto a seguridad, algunos sistemas de IA como Fortinet, Sophos Intercept o Palo Alto Networks Cortex XDR pueden identificar de forma efectiva actividades sospechosas que representan amenazas cibernéticas, lo cual es fundamental en la protección de activos digitales e información confidencial de la organización (Rubín 2024, 14).

Sin embargo, la implementación de estas herramientas requiere no solo personal capacitado sino el entorno sociotécnico que facilite la integración (Álvarez-Maldonado et al. 2024b, 14). En este sentido, las barreras cognitivas, la resistencia al cambio y la falta de familiaridad que surgen en la adaptación de esta tecnología pueden convertirse en obstáculos insuperables si no son abordados con cambios estructurales, adecuado liderazgo y rutinas digitales apropiadas. Es decir, la mera adopción de la IA no garantiza el éxito de la organización (Garcia, Carrillo, y Cuellar 2023, 2).

También es importante señalar el potencial desbalance en la fuerza laboral que la inteligencia artificial podría generar en las empresas, lo cual podría llevar a un aumento del desempleo (Álvarez-Maldonado et al. 2024b, 14), ante esto la importancia de crear sinergia en el complemento de las capacidades humanas con el potencial de las herramientas de IA con el fin de generar sistemas híbridos que potencien las fortalezas humanas con la velocidad, replicabilidad, predictibilidad y escalabilidad propias de las máquinas inteligentes (Álvarez-Maldonado et al. 2024b, 13).

1.4.2. Impacto de la IA en el sector salud

Las tecnologías de IA han acelerado la investigación en el área de la salud, desde el diagnóstico de enfermedades hasta la predicción de resultados de pacientes, debido entre otras cosas a su potencial para el manejo de grandes volúmenes de datos como registros electrónicos, biobancos e imágenes médicas. En este contexto, la efectividad de los modelos

de aprendizaje profundo en esta área radica en su capacidad para identificar y clasificar patrones complejos de información que integran registros, síntomas, y resultados de pruebas de los pacientes para generar diagnósticos más precisos. (Lim et al. 2023, 228:25-32). En general, las herramientas de IA en el sector de la salud brindan oportunidades significativas para mejorar resultados tanto para los profesionales como para los pacientes en áreas como automatización, síntesis de información, generación de recomendaciones y visualizaciones para toma de decisiones (Matheny et al. 2019, 18).

La IA tiene diversas aplicaciones en el ámbito de la salud, entre ellas el control de enfermedades crónicas cardiovasculares, diabetes o depresión, en estas la IA puede facilitar tareas como la administración de medicamentos, el monitoreo del paciente o la intervención personalizada por medio de agentes conversacionales que mediante técnicas de procesamiento de lenguaje natural brindan apoyo emocional al paciente o asesoramiento sobre su salud (Matheny et al. 2019, 76-77).

Además, mediante el uso de chatbots, robots y tecnologías de monitoreo, la IA puede contribuir en el tratamiento de personas con falencias cognitivas (Matheny et al. 2019, 78), brindar respuestas instantáneas, agendamiento de citas y consejos médicos preliminares a través de comunicación personalizada dirigida hacia las necesidades individuales del paciente y adaptándose a su idioma. Estos sistemas son capaces de reducir las ambigüedades en la comunicación y extraen valiosos hallazgos de la interacción del paciente que ayudan enormemente a la toma de decisiones médicas (Nathaniel Kumar Sarella y Therissa Mangam 2024, 2)

En el ámbito clínico, las herramientas de IA pueden analizar imágenes médicas para asistir en el diagnóstico temprano de patologías y mejorar la gestión de información optimizando la atención (Matheny et al. 2019, 80). Esto se debe a su capacidad de identificar patrones que se escapan de la percepción humana (Lim et al. 2023, 228:25). Las imágenes de TAC, MRI y rayos X han sido utilizadas por varios años para el diagnóstico en varias etapas de ciertos tipos de cáncer. Sin embargo, esta enfermedad muchas veces no es detectable por la pericia humana en etapas tempranas, mientras que varias herramientas de IA como Aview LCS+ de Coreline Soft o LungQ Clinical Suitte de Tirona son capaces de reconocerlo en las mismas imágenes generando mejores resultados de alta precisión que sobrepasan con creces la percepción humana (Rayan, Zafar, y Tsagkaris 2021, 9).

Ahmad y Alquarshi (2024, 2) mencionan que los sistemas de diagnóstico asistido por computadora CAD y potenciados con IA muestran beneficios al minimizar los errores humanos, también aumentan la eficiencia del análisis al aprovechan el aprendizaje adaptativo de nuevos datos y técnicas de imagen mediante aprendizaje continuo. Además que se apalancan del aprendizaje por transferencia que consiste en aprovechar el aprendizaje de otro modelo previamente entrenado para evitar el entrenamiento desde cero, ahorrando así recursos computacionales, tiempo de desarrollo y recursos económicos.

En este contexto, la detección de anomalías por medio de imágenes médicas, junto al análisis de registros médicos electrónicos (EHR) y datos genéticos pueden llevar a la IA a identificar patrones que sugieran intervenciones quirúrgicas (Zarghami 2024, 2) al integrar datos relevantes que ayudan a identificar factores de predisposición como el riesgo del paciente, anatomía, evolución de la enfermedad y costos asociados. Además de ofrecer al cirujano mapas anatómicos que lo guían durante la operación con el objetivo de disminuir riesgos (Matheny et al. 2019, 80). Es así que los sistemas de análisis intra operatorio pueden brindar información en tiempo real sobre el paciente (Lim et al. 2023, 228:25), apoyando al cirujano en el análisis de videos e imágenes quirúrgicas en tiempo real a fin de identificar estructuras críticas y facilitar incisiones precisas.

En este ámbito, la cirugía robótica también representa un campo que puede ser potenciado por IA para realizar procedimientos complejos que garanticen precisión (Matheny et al. 2019, 80) y mejora en los tiempos de recuperación del paciente (Lim et al. 2023, 228:25-26).

El monitoreo post operatorio es también un campo que la IA puede optimizar a través del análisis continuo de datos del paciente para garantizar la detección temprana de complicaciones, además puede incidir en la adherencia del paciente al tratamiento mediante comunicación efectiva (Zarghami 2024, 2).

Otra área importante en este sector es la salud pública. En este sentido, los programas poblacionales que identifican áreas de riesgo mediante el análisis de patrones de propagación de enfermedades (Lim et al. 2023, 228:14, 28) ayudan a optimizar la administración de recursos, servicios y estrategias de prevención. La IA puede utilizarse en el modelado espacial y predicción de riesgos por medio del análisis de datos geográficos de Sistemas de Información Geográfica “GIS” para identificar patrones y tendencias que sugieran afectación

a la salud pública (Olawade et al. 2023, 4). Además, facilita la visualización de la posible distribución de epidemias a largo plazo, brindando un soporte estratégico para la toma de decisiones y la implementación de medidas específicas en las zonas identificadas de alto riesgo. De acuerdo a Olawade (2023, 4) el uso de algoritmos de aprendizaje automático permite detectar patrones complejos en relaciones no lineales entre diferentes variables que ayudan a dirigir estrategias de salud personalizadas hacia grupos que presenten predisposición hacia ciertas enfermedades.

Otro aspecto importante de la IA en la salud pública es el control de la información errónea por medio de la colaboración de las plataformas de redes sociales a través de la implementación de medidas para identificar y contener información no verificada capaz de propagar mitos o pánico generalizado (Olawade et al. 2023, 2). En este sentido, la IA mostró su eficiencia en la pandemia del COVID apoyando en el rastreo de patrones de comunicación al evaluar la información compartida y alertando sobre tendencias de desinformación. Sin embargo, en la actualidad el desarrollo exponencial de la IA en aspectos como la generación de video o simulación de voz amerita un avance paralelo de los sistemas de detección ante los riesgos que representa el uso mal intencionado de estas tecnologías.

2. Técnicas de IA (aprendizaje de máquina y aprendizaje profundo)

El aprendizaje de máquina de acuerdo a Krishnan et al. (2021, 3) es una técnica computacional que emplea la estadística para aprender patrones o tendencias de un set de datos y luego tomar decisiones autónomas. Estos algoritmos a diferencia de la programación tradicional, no utilizan instrucciones programadas sino modelos con variables de entrada de donde evalúan patrones aprendidos en el entrenamiento (Krishnan et al. 2021, 4) para generar en su salida predicciones, clasificaciones o regresiones (ISO/IEC 2022b, 31-32) y se los utiliza generalmente para análisis de tendencia y proyecciones (Manmeet Kaur 2024, 2).

De acuerdo a la ISO/IEC 23053 (2022b, 28) los sistemas de aprendizaje de máquina se clasifican en:

Sistemas supervisados: utilizan etiquetas para clasificar los datos en su entrenamiento, como por ejemplo el entrenamiento de modelos de IA con imágenes médicas en distintas categorías patológicas debidamente etiquetas para luego identificar dichas patologías en el mundo real.

Sistemas no supervisados: no utilizan etiquetas, un ejemplo de uso es la detección de anomalías en registros de salud electrónicos (EHR) que por su naturaleza muchas veces no poseen identificadores.

Sistemas semi supervisados: utilizan un entrenamiento híbrido con algunos datos etiquetados y otros sin etiquetas, un ejemplo de estos sistemas es el sistema FocalMix que se utiliza para la detección de lesiones en nódulos pulmonares utilizando grandes cantidades de imágenes médica no etiquetadas junto a una cantidad limitada de imágenes etiquetadas (Wang et al. 2020, 1), los datos no etiquetados son pseudo etiquetados por el sistema para mejorar su entrenamiento.

Sistemas auto supervisados: utilizan datos no etiquetados en su entrenamiento y a partir de estos datos se crean señales de supervisión, como por ejemplo en el análisis de secuencias de ADN. El modelo aprende patrones de secuencias conocidas para luego identificar segmentos faltantes en posiciones específicas (Aravind Ayyagiri, Anshika Aggarwal, y Shalu Jain 2024, 7).

Aprendizaje reforzado: consiste en asignar recompensas o penalizaciones a los resultados obtenidos en el aprendizaje a modo de prueba y error. De esta forma, el sistema se acerca progresivamente a su objetivo en base a retroalimentación, como por ejemplo la rehabilitación robótica personalizada, en donde el sistema aprende por refuerzo hasta optimizar la terapia adecuada para cada paciente, o el entrenamiento de asistentes robóticos con prueba y error para perfeccionarse en sutura de heridas (Esteva et al. 2019, 1)

La norma ISO/IEC 23053:2022 (2022b, 11) identifica en el aprendizaje de máquina cuatro partes principales: La tarea o definición del problema, el modelo, los datos y las herramientas o técnicas

La tarea es en donde se define el problema a ser resuelto mediante la aplicación de un modelo, las tareas pueden ser: regresión, clasificación, agrupamiento, detección de anomalías, reducción de dimensionalidad, etc.

La regresión se utiliza para predecir valores continuos y su objetivo es determinar una función que relaciona una o más variables independientes de entrada con una variable de salida (ISO/IEC 2022b, 11-12), un ejemplo de aplicación en el área de la salud es el estudio de la relación de enfermedades cardiovasculares con factores como antecedentes familiares, nivel de colesterol, edad, etc., o la relación ente consumo de tabaco y cáncer de pulmón .

La clasificación consiste en determinar la categoría o clase a la que una instancia de datos pertenece de acuerdo a sus atributos específicos o características proporcionadas en la entrada, por ejemplo cuando en el diagnóstico médico un conjunto de síntomas de un paciente pueden ser clasificados hacia una patología en particular (ISO/IEC 2022b, 12).

El *clustering* o agrupamiento busca agrupar un conjunto de datos de acuerdo a elementos que comparten características similares, pero, a diferencia de la clasificación , en esta técnica no se utilizan etiquetas predefinidas (ISO/IEC 2022b, 12). El clustering se pueden utilizar por ejemplo en la agrupación de pacientes con características similares para optimizar el tratamiento y atención médica.

La detección de anomalías intenta identificar instancias de datos que no coinciden con el patrón esperado, esta técnica se utiliza en variedad de aplicaciones como detección de fraude (ISO/IEC 2022b, 12), o la detección temprana de anomalías conductuales mediante sensores pasivos para determinar problemas de salud mental en personas de riesgo (Rogan, Bucci, y Firth 2024, 11).

La reducción de dimensionalidad consiste en reducir el número de variables de entrada sin que afecte la información útil con el fin de optimizar los recursos computacionales y reducir costos (ISO/IEC 2022b, 13). Una técnica de reducción de dimensiones muy utilizada es el análisis de componentes principales PCA, esta técnica se ha empleado para reducir el ruido que producen por ejemplo miles de genes en un estudio en particular como la leucemia, se identifican las variables que producen la mayor varianza en los datos y se eliminan las que no aportan al estudio.

Otro componente del aprendizaje de máquina es el modelo, que consiste en una construcción matemática que genera una predicción basada en datos de entrada, el modelo entrenado puede reconocer propiedades estadísticas relevantes en forma efectiva con el fin de aproximar resultados que se asemejan a los patrones aprendidos. Según la norma ISO/IEC 23053:2022 (2022b, 7) el modelo es el resultado de aplicar un algoritmo a un conjunto de datos de entrenamiento. Si los datos usados son incompletos, el entrenamiento del modelo será insuficiente y sus resultados reflejarán falencias (ISO/IEC 2022b, 14).

Los datos son un componente esencial los sistemas de aprendizaje de máquina y pueden ser: de entrenamiento, son los que otorgan al sistema los patrones y características para el proceso de análisis; los datos de validación que se utilizan para ajustar los hiper

parámetros[5] del modelo con el objetivo de optimizar su desempeño durante el desarrollo; los datos de prueba se utilizan para revisar la capacidad de generalización del modelo y determinar su desempeño con datos futuros. Por último, los datos de producción son los utilizados una vez que el modelo está operando en un entorno real (ISO/IEC 2022b, 14).

En el área de la salud existen sets de datos de acceso público que sirven para entrenar y validar estos sistemas, tales como: el set de datos de cáncer de mama de Wisconsin que contiene 569 instancias, cada una con 32 variables, 357 de las instancias están etiquetadas como benignas y 212 como malignas, o el set de datos de hepatitis que contiene 155 instancias con 20 variables cada una, 32 instancias pertenecen a la etiqueta *muere* y 123 a la etiqueta *vive* (Krishnan et al. 2021, 31).

En cuanto a las herramientas en el aprendizaje de máquina, estas se categorizan en: herramientas de preparación de datos, algoritmos de aprendizaje de máquina, métodos de optimización y métricas de evaluación (ISO/IEC 2022b, 15). En este contexto, la evaluación de los sistemas médicos de IA es un aspecto crítico, ya que en estos se debe garantizar la eficacia y seguridad de diagnósticos y tratamientos, un sistema de diagnóstico basado en IA con insuficiente evaluación tiene altas probabilidades de detección errónea de una condición patológica que el paciente no la tiene, lo que se conoce como falso positivo, o la omisión de una condición patológica que si está presente en el paciente, lo que se conoce como falso negativo (Krishnan et al. 2021, 32).

Luego de explorar los fundamentos del aprendizaje de máquina, es importante destacar una de sus subáreas más avanzadas, se trata del aprendizaje profundo. Según la norma ISO/IEC 23053:2022 (2022b, 5) a diferencia del aprendizaje de máquina, este posee muchas capas ocultas que le permite procesar grandes conjuntos de datos con variados atributos y la creación de complejas representaciones jerárquicas a través del entrenamiento; por consiguiente, tanto el aprendizaje profundo como el aprendizaje de máquina comparten las mismas propiedades, pero el primero es capaz de abordar problemas más sofisticados.

Las capas de los modelos de aprendizaje profundo están organizadas de manera secuencial emulando las conexiones neuronales del cerebro humano (Rayan, Zafar, y Tsagkaris 2021, 6). La capa inicial recibe diferentes tipos de datos de entrada, como imágenes

[5] Son variables que se optimizan en el proceso de aprendizaje del modelo.

médicas y registros de pacientes. A medida que los datos avanzan a través de las capas profundas, se transforman en representaciones abstractas, en donde se identifican características complejas en grandes volúmenes de datos sin necesidad de intervención humana (Esteva et al. 2019, 1). La evolución de estos modelos ha sido exponencial en la última década debido a los avances computacionales y a la disponibilidad de sets de datos masivos. Según Esteva (2019, 1) los datos generados en el ámbito médico estriban los 150 exabytes y crecen 48 % anualmente. Por su parte, Shenavarmasouleh y sus colaboradores (2023, 1) sostienen que la industria de la salud genera anualmente casi la tercera parte de todos los datos del planeta.

Entre las técnicas de aprendizaje profundo con relevancia en la medicina se puede destacar las redes neuronales convolucionales (CNN), son una arquitectura de aprendizaje profundo supervisado, se componen de varias capas totalmente conectadas y utilizan *kernels* o filtros que se deslizan por la imagen para captar sus características y patrones en diferentes niveles de abstracción (Dai, Zhou, y Liu 2023, 8), mediante análisis convolucional realizan actividades de clasificación de imágenes, detección de objetos y segmentación de imágenes (Shenavarmasouleh et al. 2023, 2) . Una de sus aplicaciones principales es la visión por computadora y el procesamiento visual. En el ámbito médico se centran en la interpretación de imágenes y videos para identificar por ejemplo patrones anormales que pudieren representar patologías (Esteva et al. 2019, 1).

Otra técnica de aprendizaje profundo de relevancia en el área de la salud son los *autoencoders*, se trata de un modelo de aprendizaje no supervisado cuyo objetivo es el cifrado eficiente de los datos por parte de un codificador de entrada que transforma la información original en una representación comprimida con las características más relevantes y de menor dimensión, y por otro lado el decodificador que toma esta representación y busca replicar la entrada original (Nawaz, Khan, y Ahmad 2024, 2). Los *autoencoders* se utilizan para reducción de dimensionalidad, eliminación del ruido, detección de anomalías y como modelos generativos. En la medicina se utilizan principalmente los *autoencoders* apilados *stacked autoencoders* que están compuestos por varios niveles de codificadores y decodificadores, estos han mostrado su eficacia en el diagnóstico de enfermedades como el Alzheimer y Parkinson, en el apoyo al análisis de imágenes médica, en la detección de nódulos, detección de células en imágenes de biopsias de médula ósea, etc.

(Shenavarmasouleh et al. 2023, 3). Es pertinente mencionar que el entrenamiento de múltiples *autoencoders* demanda grandes recursos computacionales y tiempo de procesamiento que podría ser crítico en el ámbito médico (Nawaz, Khan, y Ahmad 2024, 12).

Las redes de creencias profundas o *deep belief networks* (DBN) son redes neuronales compuestas por múltiples capas de unidades visibles y ocultas, cada una de estas se forma por un conjunto de RBM[6] que se entrena de forma independiente para luego realizar un ajuste fino a la red completa. Estos modelos se utilizan principalmente para el aprendizaje no supervisado, sus principales aplicaciones son la clasificación y reconocimiento de patrones, la visión por computadora y el procesamiento de lenguaje natural. En el área de la salud, se las utiliza como clasificadores para identificar pacientes esquizofrénicos, para identificar características complejas en imágenes de resonancia magnética funcional[7] (Shenavarmasouleh et al. 2023, 3) e incluso en la detección de cáncer cervical por medio de imágenes de colposcopía (Rini Novitasari et al. 2020, 1). Entre sus desafíos, Novitasari (2020, 2) menciona que el proceso de entrenamiento de las DBN demanda altos recursos computacionales y mayor tiempo que otras técnicas de aprendizaje profundo debido a que requiere de un pre entrenamiento seguido de un ajuste.

Las Redes Neuronales Recurrentes (RNN) son un tipo de red neuronal diseñada para el procesamiento de datos temporales o secuenciales. Su principal característica es la capacidad de mantener una memoria interna, lo que les permite retener información previa y utilizarla para mejorar su rendimiento en contextos temporales, estas redes presentan ciertas deficiencias en el entrenamiento debido a su naturaleza, ante esto aparecen modelos evolucionados como *el Long short-term memory* (LSTM) que retiene información a largo plazo (Shenavarmasouleh et al. 2023, 3-4). Una de sus principales aplicaciones es el procesamiento de lenguaje natural, que permite interpretar patrones en secuencia de palabras. En el área de las salud se utiliza por ejemplo en el análisis de registros médicos, funciona extrayendo información útil de datos no estructurado para optimizar los procesos clínicos; otra aplicación son los asistentes de voz que transcriben y estructuran conversaciones entre

[6] Las Máquinas de Boltzmann Restringidas son herramientas eficientes en la modelación y aprendizaje de datos complejos de alta dimensionalidad (Shenavarmasouleh et al. 2023, 4).

[7] Es una técnica de neuroimagen que muestra la actividad cerebral en función del flujo sanguíneo.

médicos y pacientes, de esta forma colaboran en la administración y permiten que el profesional de la salud se enfoque más en el paciente (Esteva et al. 2019, 2-3).

3. Desarrollo de la IA en el contexto mundial y regional

El desarrollo actual de la IA está fomentando una competencia entre países que según Šinkovec y Miloš se compara con la carrera espacial de los años 60 (2021, 4). Este fenómeno ha llevado a las principales naciones desarrolladas a una intensa lucha por la supremacía en el dominio de esta tecnología. Países como Estados Unidos y China están invirtiendo masivamente en la investigación y desarrollo de IA con el objetivo de liderar el mercado y las aplicaciones futuras, desde la automatización hasta el sector militar. Esta rivalidad no solo se manifiesta en la inversión económica, sino también en el control de recursos clave como el hardware de procesamiento de datos y la retención del talento humano especializado.

Es así que China se ha impuesto la meta de llegar al liderato mundial de la IA en el 2030 mediante la inversión de 150 billones de dólares por parte del gobierno (Jagodič y Šinkovec 2021, 3). Mientras que Estados Unidos dispone de empresas referentes en este ámbito, como por ejemplo NVDIA que mantiene en estos momentos el liderazgo en el desarrollo de hardware GPU que facilita el procesamiento de modelos de aprendizaje profundo debido a sus características de procesamiento masivo y paralelo de datos. De acuerdo con Márquez (2025, 6-7), el gobierno de EE. UU. con el fin de mantener el liderazgo tecnológico en las aplicaciones de IA implementó rigurosos controles dirigidos hacia los clientes que buscan adquirir sus chips de alto rendimiento, enfatizando las restricciones a países como China, Rusia o Corea del Norte, política que por China ha sido calificada como bloqueo tecnológico y guerra económica. Por su parte, China en esta contienda, viene implementando restricciones en la exportación de tierras raras (Márquez 2025, 7), elemento necesario en la fabricación de estos componentes.

Estas medidas restrictivas afectan económicamente a las mismas empresas estadounidenses, las cuales según Márquez (2025, 21), han reducido su valor en 240000 millones de dólares, situación que merma su capacidad de inversión en investigación y desarrollo. Además, EE. UU. se enfrenta con la escasez de profesionales calificados en áreas

STEM[8], debido a que la inversión en su formación se pierde con las atractivas propuestas de trabajo que promueven países como Australia, Canadá y el Reino Unido mediante visas de talento (Márquez 2025, 7).

Por su parte, la Unión Europea busca disminuir la dependencia externa y evitar riesgos en la disrupción de su cadena de suministros fortaleciendo su soberanía tecnológica, mediante la creación de un fondo de inversión en tecnologías estratégicas de chips e IA que contempla: microprocesadores avanzados, arquitecturas alternativas, fotónica integrada y chips cuánticos. Estas acciones impulsarían la reindustrialización de Europa en alineación con metas de sostenibilidad, neutralidad climática y economía circular (Márquez 2025, 31–35).

En esta coyuntura mundial, Filgueira (2023, 9–23) señala que América Latina enfrenta una fuerte dependencia tecnológica debido a su especialización en la economía primaria, lo cual limita su independencia económica y el desarrollo de infraestructura local robusta que apoye los procesos de innovación relacionados con la IA. Además, la falta de control en su infraestructura de información la expone a riesgos de *colonialismo de datos,* que implica el uso de su información pública por grandes corporaciones extranjeras para fines gubernamentales o comerciales.

Con el objetivo de medir el desarrollo de la IA en América Latina, El CEPAL en conjunto con el CENIA[9] realizaron un diagnóstico integral del estado de desarrollo, adopción y gobernanza de la IA en esta parte del mundo, enfocándose especialmente en sectores industriales, educativos y científico. Sus resultados se reflejan en el Índice Latinoamericano de IA, por sus siglas ILIA. Este índice se construye sobre varias dimensiones y subdimensiones ponderadas de la siguiente forma: factores habilitantes con una ponderación del 40% que incluye infraestructura, datos y talento humano. Por otro lado, investigación, desarrollo y adopción, con una ponderación del 35%, y por último gobernanza que se pondera con el 25% e incluye visión e institucionalidad, cooperación internacional y regulaciones (2024, 164).

El informe sugiere que existe una marcada disparidad entre los países de la región en lo referente a la adopción y desarrollo de tecnologías de IA. En este sentido, países como

[8] Acrónimo en inglés de Ciencia, tecnología, ingeniería y matemáticas
[9] Es el Centro Nacional de Inteligencia Artificial de Chile.

Brasil y Uruguay muestran el mejor desempeño en la región en cuanto a la adopción, infraestructura y regulación de esta tecnología, mientras que otros países como El Salvador y Honduras muestran una participación considerablemente menor (CENIA 2024, 29). Es interesante mencionar que el PIB per cápita no incide directamente en el desempeño de los países de la región en la adopción y desarrollo de la IA. Sino más bien factores estructurales como la composición industrial, inversión en investigación y desarrollo y capacidad para atraer industrias de alto valor tecnológico (CENIA 2024, 97).

El documento menciona que la brecha entre Latinoamérica y los países del Norte Global se ha mantenido en los últimos ocho años. Por otro lado, acota que la alfabetización ha incrementado en algunos países gracias a las tecnologías de IA. Sin embargo, desde el 2019 existe en la región una tendencia a la fuga de talento. Por lo tanto, un desafío importante en la región es retener sus especialistas. Es importante mencionar que la caracterización económica de cada país y sus políticas públicas son factores importantes que impactan directamente en la adopción de IA. Además, el número de publicaciones multidisciplinarias asociadas con IA alcanzó el 80% en la región, lo que evidencia una mejora en la utilización de estas herramientas en apoyo al desarrollo científico. Y, por último, el documento resalta que ya existen en la región 38 iniciativas legales en materia de IA, que contemplan marcos regulatorios más amplios que permitirían, entre otras cosas, sancionar el uso indebido de IA generativa, tal es el caso de estafas telefónicas en Chile y violación a la privacidad en México (CENIA 2024, 17).

Respecto a la calificación, el informe del CENIA agrupa a los países según el grado de madurez en las dimensionas antes mencionadas y los divide en tres categorías (2024, 14):

Pioneros: abarcan el tercer tercil con los valores más altos, este lugar ocupa aquellos países líderes en infraestructura tecnológica, desarrollo especializado, productividad científica y capacidad de innovación.

Adoptantes: comprende el segundo tercil con los puntajes intermedios y corresponde a los países que utilizan la IA de manera incipiente en sectores productivos, administraciones públicas, servicios e investigación. Estos países muestran disposición para invertir en políticas de fomento de la IA.

Exploradores: Corresponde al primer tercil con los valores más bajos y comprende los países que apenas están desarrollando capacidades básicas en esta área, carecen de una comunidad de investigación y apenas empiezan a impulsar políticas públicas preliminares.

Según la figura 1, Ecuador se sitúa en el segundo tercil. Es decir, el de los países adoptantes, con una puntuación de 35.59. Se encuentra muy cerca del límite con el primer tercil, el de los países exploradores, junto a Panamá, que tiene 37.48 puntos y Jamaica, que logra 32.38 puntos. Es notable que sus vecinos, Perú y Colombia, obtengan calificaciones significativamente más altas, con 45.52 y 52.64 puntos respectivamente. El caso de Colombia, de hecho, la acerca al grupo de los países pioneros.

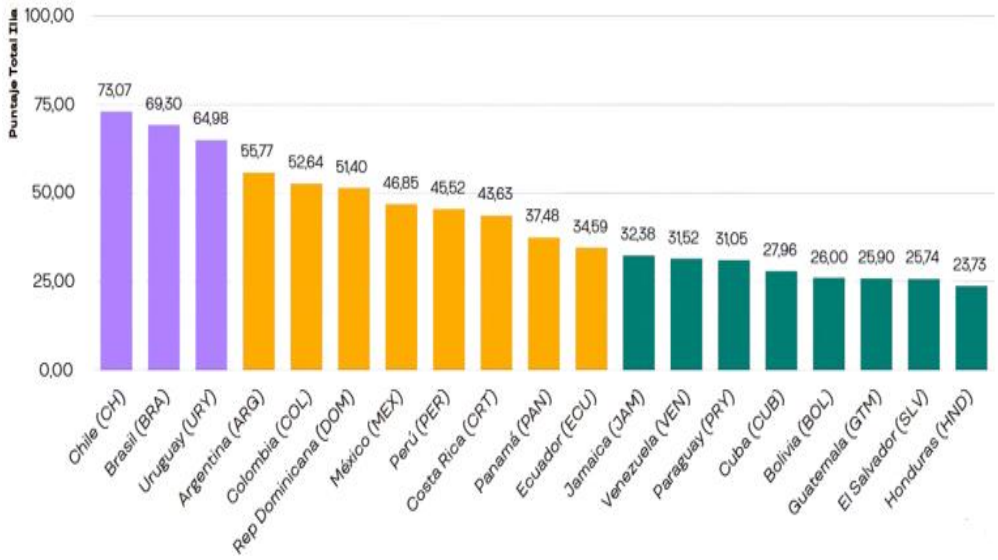

Figura1. Puntaje total ILIA de los países latinoamericanos. Fuente: Índice latinoamericano de inteligencia artificial (CENIA, 2024). Elaboración: CENIA

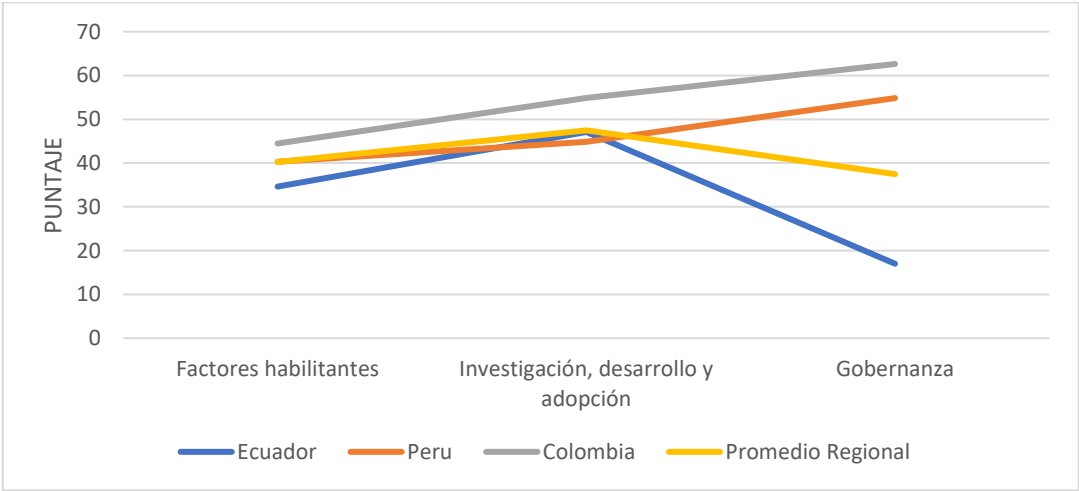

Figura 2. Comparación entre Ecuador, Perú y Colombia en las dimensiones principales del ILIA. Fuente: Índice latinoamericano de inteligencia artificial (CENIA, 2024). Elaboración propia

La figura 2 indica que, tanto en la dimensión de factores habilitantes y en la dimensión de investigación, desarrollo y adopción, el Ecuador muestra un desempeño similar al promedio regional y a sus países vecinos. Sin embargo, cuando hablamos de gobernanza, el puntaje de Ecuador cae apenas a 17 puntos, mientras que el promedio regional es de 37, sin mencionar el desempeño mucho más alto al promedio en esta dimensión por parte de los países vecinos. Según el proyecto *Diagnóstico sobre la IA en el Ecuador* diseñado por el MINTEL (2021, 6), la gobernanza es la capacidad del estado para establecer redes simbióticas entre instituciones públicas y privadas destinadas a resolver problemas sociales o crear oportunidades.

Para entender el bajo desempeño del Ecuador en esta dimensión, es pertinente desentrañar sus componentes, los cuales son:

Visión e institucionalidad: comprende las estrategias y políticas de IA desde su contenido hasta los mecanismos de seguimiento y alcances. En este componente el Ecuador obtiene una calificación de cero puntos (CENIA 2024, 114).

Vinculación internacional: se construye a través de la adhesión a acuerdos internacionales más allá de América Latina y el Caribe, además se considera la relevancia de participar en procesos de definición de normas ISO. En este componente el Ecuador obtiene un promedio de 50 puntos que resulta de cero puntos por la participación en estándares ISO y 100 puntos por la participación con organismos internacionales (CENIA 2024, 117).

Regulación: considera la revisión de los proyectos de ley vigentes y también los que están en discusión, además de un análisis del contexto regulatorio evidenciado tras este proceso. También se considera la ética y sustentabilidad que toma en cuenta el ambiente, sociedad y derechos humanos. En este ámbito, el Ecuador obtiene 23.33 puntos, calificación por debajo del promedio regional que corresponde a 45.28 puntos (CENIA 2024, 118). Por otra parte, la regulación se compone del subindicador mitigación de riesgos que identifica la presencia de mecanismos de gestión de riesgos relacionados a la IA. En este componente, el Ecuador obtiene una calificación de cero puntos (CENIA 2024, 119). El subindicador de ciberseguridad forma también parte de la regulación y refleja el nivel de madurez de los países en la protección de datos públicos. En este aspecto el Ecuador obtiene 27.23 puntos,

puntuación por debajo de los 49.85 puntos que representa el promedio regional (CENIA 2024, 123).

En este contexto, la AILA[10] (2025, 8–9) también realiza una evaluación de la participación del gobierno ecuatoriano en el desarrollo de la IA, tomando en cuenta tres aspectos: gobierno como habilitador del entorno de la IA, gobierno como usuario de la IA y gobierno como garante del uso ético de la IA. El informe presenta los siguientes hallazgos:

En la evaluación del gobierno como habilitador del entorno de la IA, el Ecuador obtiene un puntaje de 2.9 sobre 5 puntos, lo que refleja algún compromiso político para impulsar la adopción y desarrollo de la IA en el país (PNUD 2025, 21). Sin embargo, la ALIA acota que existen limitaciones estructurales en infraestructura tecnológica como la baja adopción de tecnologías centradas en la nube, retraso en la red 5G y falta de especialización en los centros de datos. Además, la gestión de datos públicos es insuficiente debido a que existen brechas significativas en su adopción y uso, puesto que estos presentan baja calidad y estandarización insuficiente. Por otro lado, se identifica también deficiencia en seguridad y portabilidad de datos. Otro de los aspectos clave es el déficit crítico del país en talento especializado, debido a la baja oferta de profesionales y baja retención del talento. Por último, la ausencia de una estrategia nacional de IA y falta de financiamiento que incluya incentivos adecuados para implementar un ecosistema de innovación, junto con la existencia de instituciones estatales que no brindan confianza al sector privado y restringen la inversión (PNUD 2025, 8).

El gobierno como usuario de la IA obtiene 2.5 sobre 5 puntos (PNUD 2025, 21), esta calificación evidencia que existe cierto avance en áreas clave, pero persisten desafíos que incluyen: la falta de una estrategia nacional formal para el uso de IA en la gestión pública, la incipiente transformación digital que se ha centrado únicamente en automatizar ciertos procesos, dejando de lado herramientas avanzadas de IA para la toma de decisiones. Por último, la escasez de talento digital avanzado en el personal de TICs del ecosistema público, cuyo trabajo actual en lugar de centrarse en analítica avanzada se canaliza hacia mantenimiento básico (PNUD 2025, 9).

Otro aspecto muy importante son los factores que impiden que el gobierno se convierta en un garante efectivo del uso ético de la IA. La calificación que recibe el Ecuador

[10] AILA por sus siglas: Artificial Intelligence Landscape Assessment.

en esta área es apenas 1.9 sobre 5 puntos (PNUD 2025, 21), debido a la carencia de un marco normativo sobre la ética en esta tecnología. No existe una ley de IA en el país ni reglamentos éticos oficiales y tampoco una entidad especializada en supervisión. Además, en el país no se practica la transparencia algorítmica puesto que no existe respaldo legal que garantice el derecho a la explicabilidad de las decisiones de la IA. Tampoco se ha desarrollado un sistema formal de gestión de riesgo o algún tipo de financiamiento para su investigación y mitigación. Por último, la inclusión y equidad en el desarrollo de la IA no están garantizados porque no existen lineamientos públicos que aseguren alguna representatividad en los datos de entrenamiento de los sistemas de IA. Por ejemplo, es pertinente mencionar que se observa una sobre representación de mujeres e indígenas en sectores de ciencia, tecnología e innovación (PNUD 2025, 9).

Según la Política Pública de Transformación Digital del Ministerio de Telecomunicaciones de Ecuador (2025, 77, 85), la inteligencia artificial es considerada una tecnología emergente con gran potencial para el desarrollo sostenible en el Ecuador. No obstante, el documento no especifica acciones, planes ni cronogramas concretos para su implementación. En este contexto, se propone una guía de acciones orientadas a fortalecer el posicionamiento de Ecuador frente a los países latinoamericanos y, por qué no, también a nivel mundial en el desarrollo y la aplicación de tecnologías de inteligencia artificial:

Desarrollar una estrategia nacional de IA que no solo reconozca a la IA como tecnología relevante, sino que incluya financiación, incentivos fiscales, programas de colaboración transversal con el sector privado, además de objetivos claros y precisos. Esta estrategia podría incluir el apoyo a la formación de clústeres que fomenten un entorno dinámico de colaboración entre gobierno, emprendimientos, academia y grandes empresas.

Crear un marco regulatorio robusto que establezca directrices que promuevan la transparencia, fortalezcan la seguridad de la información y fomenten la equidad sin asfixiar la innovación. Así, este marco regulatorio no solo debería contener regulaciones y sanciones, sino también incentivos para el desarrollo de soluciones éticas de IA en el país.

Proteger y sistematizar la recolección de datos públicos mediante la implementación de infraestructura tecnológica que permita su recolección y protección, tales como: IoT, internet en la nube, redes 5G, big data, etc. En este contexto, según el informe AILA (2024, 29) el portal *datosabiertos.gob.ec* cuenta con más de 1417 conjuntos de datos provenientes

de 97 instituciones públicas. Sin embargo, su impacto se limita debido a la falta de estándares consistentes. Por lo tanto, es de suma importancia garantizar la estandarización y normalización de datos en todas las fuentes del gobierno, para que estos sean accesibles y utilizables, siempre bajo normas estrictas de protección a la privacidad.

Incentivar la formación de talento humano en IA mediante la estrecha colaboración de la academia, abriendo programas de capacitación en IA a nivel tecnológico, pregrado y doctorado, que incluyan las habilidades que necesita el país y el mundo para el desarrollo y uso correcto de estas tecnologías. En este sentido, el gobierno podría ofrecer becas e incentivos que canalicen afluencia de estudiantes hacia carreras orientadas a la IA. Sin embargo, este esfuerzo sería insuficiente si no va acompañado de políticas que incentiven la retención del talento humano como por ejemplo programas de apoyo a los emprendimientos de IA e incentivos fiscales. Según el INEC (2024), tan solo el 12 % de los títulos nacionales corresponden a áreas de la ingeniería, en su mayoría, las titulaciones se orientan a áreas como la educación, periodismo, administración y en menor medida salud. Por otro lado, el porcentaje de titulaciones en tecnologías de la información apenas llega a 3 %. En este sentido, es necesario mencionar que los incentivos para el aprendizaje de este tipo de tecnologías emergentes comienzan por mejorar los programas de formación inicial de los niños y jóvenes para alentar su involucramiento activo desde el principio en áreas STEM.

Promover el uso ético y responsable de la IA no solo con la creación de leyes que de cierta manera garanticen el desarrollo y uso ético de esta tecnología, sino también con el fortalecimiento de instituciones como el MINTEL para que garantice la corresponsabilidad entre desarrolladores y usuarios y así evitar la perpetuidad de desafíos éticos como el sesgo. Pero principalmente es necesario fomentar en el país una cultura ética mediante campañas de concientización, talleres, integración y práctica de la ética como materia curricular desde etapas iniciales de estudio. Según el índice de percepción de la corrupción del año 2024, el Ecuador apenas alcanza 32/100 puntos (FCD 2025). Por lo tanto, es también de vital importancia depurar las instituciones estatales existentes para que generen confianza y gobernabilidad.

Fomentar la innovación es fundamental para el desarrollo de la IA en el país. En este sentido, un indicador de la capacidad de innovación de un país es el número de patentes

registradas. Según la RICYT [11] (2022), en 2019 se publica la última cifra del Ecuador. En ese año el país contaba con 25 patentes por cada millón de habitantes. Esta cifra es inferior a la de sus países vecinos, ya que Perú alcanzaba 39 patentes por millón de habitantes y Colombia ostentaba las 43 patentes por millón. Ante esta realidad, se concluye que Ecuador necesita generar una cultura de innovación y considerarla como prioridad en lugar de solo delegarla a departamentos estatales insuficientemente articulados, fomentar e incentivar la formación de ingenieros y científicos, brindar incentivos fiscales al emprendimiento dinámico y facilitarle el acceso a crédito. Además, es imperativo garantizar la protección de la propiedad intelectual. Según Andrés Oppenheimer en su libro Crear o Morir (2014, 315), la cultura de innovación se caracteriza por un entusiasmo colectivo por la creatividad que enaltezca a los innovadores productivos de la misma manera con que los medios de comunicación se encargan de glorificar a los grandes deportistas. Entre las iniciativas que el gobierno podría implementar para fomentar la innovación en IA, se destacan la organización de encuentros enfocados en la resolución de problemas relevantes, acompañados de incentivos económicos otorgados tanto por el Estado como por el sector privado a los innovadores que presenten las mejores soluciones (Oppenheimer 2014, 317).

Sin la existencia de políticas públicas y regulaciones que brinden un marco institucional sólido, el desarrollo y uso responsable de la IA en Ecuador carece de un horizonte claro. Además, no es suficiente la promulgación de leyes, sino la formación de instituciones robustas que generen un marco de gobernanza claro que provea confianza, estrategia y mecanismos de coordinación interinstitucional. Estos factores son clave para sentar las bases del desarrollo de la IA y permitir el posicionamiento del país entre el grupo de pioneros de esta tecnología a nivel sudamericano.

4. Marco legal

En el 2008 la IA deja de ser parte de investigaciones académicas y se convierte en componente de productos y servicios ofertados al público en general, en consecuencia, el desarrollo de la IA muestra en la actualidad un crecimiento exponencial (Garzón Espitia y Rodríguez Padilla 2022, 16). Según el CEPAL (2024, 30), un estudio de McKinsey &

[11] RICYT son las siglas de Red de Indicadores de Ciencia y Tecnología Iberoamericana e Interamericana

Company indica que a nivel mundial el 72 % de ejecutivos encuestados utilizó al menos una función de IA en el 2024, lo que representa un aumento significativo en comparación con la misma encuesta que arrojó un resultado del 20 % en el 2017. No obstante, en una investigación sobre los marcos regulatorios de la IA desde el 2008 al 2018, Garzón Espitia y Rodriguez Padilla (2022, 16) concluyen que los países con mayor desarrollo de IA han otorgado preponderancia a sus legislaciones en el área de protección de datos y vehículos autónomos, sin otorgarle importancia relevante a las aplicaciones de IA ni a sus implicaciones éticas.

Entre las consecuencias negativas detalladas por Loján Alvarado y Cárdenas Villavicencio (2024, 9) sobre el uso de la IA que requieren regulación, se pueden destacar las siguientes:

- Vulneración de derechos humanos en aplicaciones de vigilancia masiva, represión política o manipulación psicológica.
- Decisiones discriminatorias sobre contratación laboral, préstamos, libertad condicional, etc.
- Abusos de la privacidad debido a la utilización de datos masivos para el entrenamiento de los sistemas de IA.
- La automatización del trabajo mediante herramientas de IA que puede desequilibrar mercados laborales y generar desempleo.
- El uso de tecnologías de IA en armas que podría desligar a los humanos de responsabilidades éticas y hacer más plausible la guerra.

En este contexto, el rápido crecimiento en el uso de sistemas de IA a nivel mundial amerita la adopción de un enfoque responsable que contemple una perspectiva multidimensional en cuestiones éticas, sociales, técnicas y legales (Azabache Santos, Ángeles Piedra, y Mendoza De Los Santos 2024, 8), debido a que la promulgación de regulaciones en los diferentes países, al momento, es insuficiente y, más aún, las propuestas regulatorias internacionales han permanecido en gran medida solo en forma de recomendaciones o debates (Garzón Espitia y Rodríguez Padilla 2022, 17).

4.1. Ley de Inteligencia Artificial

Ante este escenario, en marzo del 2024, la Unión Europea aprobó su ley de inteligencia artificial, de cumplimiento obligatorio para los 27 Estados miembros y sus socios. Esta normativa se convierte en la primera de su tipo a nivel mundial (Marr 2024).

El principio de esta ley se basa en que "La IA debe ser una tecnología centrada en el ser humano y debe servir como herramienta para las personas, con el objetivo último de aumentar el bienestar humano". En este sentido, la ley prohíbe su utilización para influir comportamientos en forma perjudicial, entre las prohibiciones, se nombra la clasificación biométrica con la intención de inferir creencias políticas, religiosas u orientaciones sexuales o para la identificación remota de personas en lugares públicos como en los sistemas de reconocimiento facial (Marr 2024).

En el reglamento, las aplicaciones consideradas peligrosas se etiquetaron como "inaceptables" y se prohíbe su uso salvo para el gobierno, ciertos estudios científicos y fuerzas de seguridad; el resto de las aplicaciones se clasifican según su riesgo en alto, limitado y mínimo. Entre las aplicaciones de alto riesgo están las aplicaciones médicas y los autos autodirigidos. En consecuencia, las organizaciones involucradas en estos segmentos deberán responder a normas más estrictas de calidad y protección de datos. Las aplicaciones de IA de riesgo limitado y mínimo comprenden videojuegos y procesos creativos, en estas los requisitos son menores pero se les exige transparencia y uso ético de propiedad intelectual (Marr 2024).

La ley intenta regular la transparencia de dos formas, decreta la clara demarcación de las imágenes generadas por IA con el objetivo de limitar el engaño y la desinformación, por otro lado, obliga a los desarrolladores de sistemas de alto riesgo en proporcionar amplia información sobre los pormenores del sistema de IA y se establecen medidas de supervisión y responsabilidad humana (Marr 2024).

En América Latina las regulaciones en materia de IA son incipientes y no existen condiciones ni incentivos para una cooperación regional, apenas se observan esfuerzos tímidos entre los estados por medio de *soft laws*[12] promulgadas por la Asamblea General de Naciones Unidas que proveen ciertos lineamientos de carácter general, cuyo contenido ha sido acogido por países latinoamericanos como: Ecuador, Argentina, Chile, Brasil, Costa

[12] Se trata de principios, conjunto de acuerdos o declaraciones de carácter voluntario.

Rica y Perú. Estas resoluciones buscan que los Estados se "abstengan o dejen de usar sistemas de IA que no operen en consonancia con el derecho internacional o que supongan riesgos indebidos para el disfrute de los derechos humanos" (Contreras 2024, 8, 17).

Por otro lado, RIPD[13] ha publicado dos documentos relevantes en materia de IA, se trata de las "Recomendaciones generales para el tratamiento de datos en IA" que consiste en un instrumentos tipo soft law que contiene recomendaciones similares al marco europeo RGPD[14] y la *Declaración sobre neurodatos* que fija directrices genéricas para la protección de la privacidad de datos en materia de neurotecnologías (Contreras 2024, 10) .

El Ecuador no es la excepción en cuanto al crecimiento de uso de estas tecnologías, el 36 % de ecuatorianos las utilizan en sus tareas cotidianas (Vivero 2024). Sin embargo, el país en la actualidad no cuenta con una ley de IA, aunque se han propuesto algunos proyectos para legislarla, tales como:

El Proyecto de Ley Orgánica de Regulación y Promoción de la IA en Ecuador, fue presentado en el 20 de junio de 2024 por la asambleísta Silvia Nuñez, este proyecto parte desde la protección de los derechos humanos y constitucionales y busca garantizar la educación y fomentar las capacidades para el uso de estas herramientas, presenta un enfoque inflexible debido a sus regulaciones excesivas que pretenden regular la IA ligada a la libre competencia, el consumo, la protección de datos y la transformación digital, el proyecto demanda de una compleja infraestructura estatal y recursos. Este proyecto cae en ambigüedades de definición en referencias como proporcional, adecuado, razonable y sustancial (Vivero 2024).

El proyecto de Ley para el Fomento y Desarrollo de la IA fue presentado por la asambleísta Karina Subía el 30 de julio del 2024, es un proyecto breve que recoge aspectos específicos, apunta a reglamentos que permitan aprovechar las leyes de emprendimiento e incentivos tributarios que ya existen en Ecuador y concentra sus esfuerzos regulatorios en las áreas de alto riesgo como las decisiones automatizadas en el área de la salud y el reconocimiento facial (Vivero 2024).

Por último, el proyecto de ley Orgánica de Aprovechamiento Digital e Inteligencia para niñas, niños y adolescentes que fue presentado por la asambleísta Pierina Correa el 17

[13] Es la red iberoamericana de protección de datos
[14] Es la ley que regula el tratamiento de datos personales en la Unión Europea.

de septiembre del 2024, el proyecto pretende etiquetar la IA de alto riesgo y en esta concentrar sus esfuerzos regulatorios, excluyendo de estos a los que tienen fines investigativos, propone también bajo un modelo de transparencia la revelación de datos y algoritmos utilizados en las aplicaciones de IA que serían supervisadas por la ANSIA[15], entidad con múltiples competencias. Se contemplan también regulaciones y obligaciones de información para niños y adolescentes (Vivero 2024).

4.2. Ley de Protección de Datos Personales

En la coyuntura actual, la información proveniente de datos personales se ha posicionado como un bien jurídico, cuyo valor para muchas organizaciones supera el de otros bienes tradicionalmente, debido a que estos pueden ser usados a través de entornos informáticos para influir en diversas dimensiones que incluyen aspectos políticos, económicos o sociales (Gutiérrez Proenza 2022, 5).

En EEUU se aprobó en 1974 la ley que materializa la protección de datos personales, mientras que Europa empezó a regular esta práctica desde 1960 (Gutiérrez Proenza 2022, 4).

Pese a la evidente necesidad de regulación en épocas contemporáneas a nivel global por casos como la sanción que impuso la Comisión Federal de Comercio de EEUU a Facebook en 2018 por exponer la privacidad de 50 millones de usuarios en el caso de Cambridge Analytica (Gutiérrez Proenza 2022, 5) o la multa impuesta por la Comisión Federal de Comercio a Google por la recopilación y promoción de datos personales de niños sin el consentimiento de sus padres (Prensa Asociada 2019). En Ecuador el suceso que marcó la necesidad de regulación fue la fuga de información de 20 millones de ecuatorianos incluyendo 6.7 millones de menores de edad investigada por personal de la compañía de seguridad informática israelí *vpnMentor* el 24 de septiembre del 2019 (Gutiérrez Proenza 2022, 6), este hecho estimuló el debate sobre el tema en la Asamblea Nacional del Ecuador y es así que el 10 de mayo del 2021 fue aprobada la ley orgánica de protección de datos personales. Esta ley establece principios, derechos, obligaciones y mecanismos de tutela para garantizar la protección de datos personales y abarca cualquier modalidad de uso posterior (Asamblea Nacional del Ecuador 2021, 4,38).

[15] ANSIA, siglas de Autoridad Nacional de Supervisión de la IA.

La ley se extiende a personas y entidades públicas o privadas que traten con datos personales y sus principios incluyen legalidad y respeto al derecho de los titulares de los datos, quienes deben consentir su uso, salvo en ciertas excepciones como cuando se trata de fuentes accesibles al público, o una relación jurídica libre y legítima entre el responsable del tratamiento y el titular de los datos, también si los datos son proporcionados a autoridades administrativas o judiciales en base a solicitudes amparadas en normativa vigente o por último si se trata de interés público esencial en salud (Asamblea Nacional del Ecuador 2021, 19).

La ley busca asegurar el tratamiento de datos en forma transparente, justa y lícita. En consecuencia, los titulares de los datos tienen derecho de acceso para conocer de qué forma se está manejando su información o solicitar la corrección de sus datos personales en caso de que estos sean inexactos o estén incompletos, también tienen derecho, mediante un habeas data (Heredia Peñaloza y Santacruz Vélez 2023, 7), a cancelar la autorización cuando lo consideren; es necesario mencionar que los titulares podrían oponerse al tratamiento de sus datos si se los utiliza con objetivos publicitarios, además, tienen derecho a recibir sus datos en formato legible y estructurada y finalmente la ley otorga a los titulares el derecho de ser protegidos ante cualquier forma de discriminación derivada del tratamiento de su información personal (Asamblea Nacional del Ecuador 2021, 11-14).

Por otro lado, la ley establece responsables en el tratamiento de estos datos y además menciona que la actividad se debe centrar en los principios de: legitimidad, que establece la necesidad de una base legal que justifique su utilización ; finalidad, establece que los datos personales solo pueden ser utilizados para fines específicos, explícitos y legítimos; proporcionalidad, establece que el tratamiento de los datos debe ser relevante, adecuado y limitado a lo necesario; calidad, establece que se tomen las medidas razonables para asegurar que los datos personales sean exactos y actualizados; seguridad, establece la obligatoriedad de proteger los datos personales contra acceso no autorizado, pérdida, destrucción o daño accidental; y por último, transparencia, que obliga a los responsables del tratamiento de datos a informar de manera clara a los titulares sobre el propósito y la forma en la que se tratará su información (Asamblea Nacional del Ecuador 2021, 9-10).

Entre las deficiencias de la ley de protección de datos personales LOPDP algunos expertos mencionan: la exclusión hacia personas jurídicas o personas fallecidas, salvo en este

último caso si existiere un sujeto titular de derecho sucesorio, más no aplica a personas con interés legítimo sin derecho derivado (Gutiérrez Proenza 2022, 10). Por otro lado, la ley establece la posibilidad de transferencia de datos en base a la presunción de que el titular sea informado sobre la finalidad del uso de sus datos, más no aclara la exigencia de un consentimiento expreso (Gutiérrez Proenza 2022, 11). Este particular es en la actualidad objeto de abuso o mal interpretación por organizaciones, como por ejemplo las entidades bancarias, quienes transfieren sus bases de datos a empresas gestoras de cobros sin que exista consentimiento expreso del titular (Gutiérrez Proenza 2022, 11).

Uno de los casos documentados en Ecuador sobre la transferencia de datos personales es descrito por Heredia y Santa Cruz (2023, 9–13). Este caso involucra a un paciente que fue evaluado en un hospital por varios estudiantes de diversas especialidades, incluyendo dermatología, alergología, urología, hematología, reumatología e imagenología. Los resultados de dicha evaluación concluyeron en el diagnóstico de una patología conocida como el *Síndrome de Sweet*; pasados algunos meses, el paciente es reconocido mediante fotografías por familiares y amigos en una revista del colegio de médicos de difusión electrónica y física que exponía deliberadamente incluso referencias de salud muy personales del paciente e imágenes que mostraban su rostro.

En este contexto, es crucial resaltar la importancia de la confidencialidad en relación con la información médica, especialmente según Heredia y Santa Cruz (2023, 6) considerando las complejidades asociadas a la propiedad de dichos datos, ya que su generación involucra a múltiples actores. Por un lado, está el paciente; por otro, el médico, quien elabora la historia clínica mediante su experiencia y análisis, finalmente, el hospital que también juega un papel al proporcionar una remuneración al médico. No obstante, la ley protege la información del paciente y exige que los profesionales de la salud asuman la responsabilidad de recopilar y mantener estos datos de forma confidencial, sin otorgarles sobre ellos control o derechos de propiedad.

5. Marco normativo

Según Longo y O'Reilly (2023, 1662:190) la falta de armonización entre marcos regulatorios internacionales puede crear obstáculos significativos para el cumplimiento de las obligaciones establecidas y en consecuencia para el comercio mundial. Ante esta realidad,

se han propuesto varios estándares con aceptación a nivel global que pretenden brindar a las organizaciones un enfoque estructurado de gestión de riesgos para el desarrollo o utilización de sistemas de IA, entre ellos se puede nombrar el NIST AI RMF, ALTAI, ISO 42001, etc.

5.1. Norma ISO/IEC 42001:2023

La ISO/IEC 42001:2023 responde a la necesidad de establecer directrices para la gestión responsable de IA , fue elaborada por el Comité Técnico Conjunto ISO/OEC JTC 1, Subcomité SC42 (ISO/IEC 2023, 7), es el primer estándar ISO para AIMS[16], se centra en la gestión de sistemas de IA y ofrece un marco que permite abordar desafíos y riesgos asociados con el uso de estas herramientas. Además, es la primera normativa certificable de IA a nivel mundial y se compone de una estructura armonizada que facilita su integración con otros sistemas de gestión como por ejemplo el ISO/IEC 27001 o ISO 45001 (Thiers, Md PhD y Harned 2024, 3).

Por otro lado, el estándar ISO/IEC 42001:2023 impulsa prácticas sostenibles en el ámbito de la inteligencia artificial con el objetivo de beneficiar a la sociedad en su conjunto. Este estándar se alinea con los ODS, promoviendo la igualdad de género, fomentando la innovación y respaldando el crecimiento económico (Kappel 2024). Al adoptar esta normativa, las organizaciones no solo garantizan un enfoque ético y responsable en el desarrollo de tecnologías de IA, sino que también aseguran un impacto positivo y duradero en la comunidad. De este modo, contribuyen al avance de una sociedad más justa, mientras potencian un entorno de innovación que favorece el desarrollo económico sostenible.

La norma establece requisitos específicos para la gestión de los riesgos provenientes de la IA y procesos para gestionarlos de forma transparente (ISO/IEC 2023, 47). La información generada bajo la estructura de la norma no es pública por defecto , pero separa cierta información relevante que puede ser revisada por sus partes interesadas (Thiers, Md PhD y Harned 2024, 4), esta información debe ser accesible y comprensible (ISO/IEC 2023, 47).

Otro aspecto fundamental de la norma es el cumplimiento legal. Las entidades certificadas deben identificar los requisitos legales aplicables, asegurarse de cumplir con las regulaciones locales y con los requisitos específicos de seguridad, así como definir roles y

[16] AIMS, por su siglas Artificial Intelligence Management System.

responsabilidades en el manejo de la información. Además, deben asumir la responsabilidad en el desarrollo y la utilización de los sistemas de inteligencia artificial (ISO/IEC 2023, 16, 50, 51).

La gestión de riesgos es un aspecto esencial, permite a las organizaciones la identificación, evaluación y mitigación de los riesgos asociados con el uso de los sistemas de IA. El riesgo se define como el efecto de la incertidumbre sobre los objetivos e incluye tanto los efectos positivos como negativos, este se pondera por la combinación de la probabilidad que ocurra un evento y sus posibles consecuencias (ISO/IEC 2023, 12). De acuerdo con Bogucka (2024, 5), las organizaciones deberían en esta revisión no solo centrarse en aspectos técnicos y legales sino abordar un enfoque amplio que contemple también los aspectos éticos como sesgo y equidad, impacto en el empleo, sostenibilidad, etc.

Por su parte, Jedličková (2024, 10-11) señala que los riesgos éticos son multidimensionales y para mitigarlos de manera efectiva, es necesario adoptar un enfoque holístico que integre soluciones técnicas, políticas claras y un marco robusto que asegure la incorporación de principios éticos en cada etapa del ciclo de vida de la inteligencia artificial.

En este contexto, la ISO/IEC 42001:2023 (2023, 19) acota que la naturaleza y nivel de riesgo tienen relación con el contexto específico de la organización y su entorno operativo. Para su identificación, las organizaciones deben establecer mecanismos de monitoreo continuo e implementar acciones sistemáticas que involucren el análisis de las posibles afectaciones a la privacidad, seguridad y ética en relación con sus partes interesadas. Los riesgos identificados deben ser evaluados en función de su probabilidad de ocurrencia y el impacto potencial, con el objetivo de priorizar estrategias y medidas de mitigación hacia los que requieren inmediata atención, estas pueden incluir controles técnicos y organizativos, capacitación del personal o creación de políticas que aborden riesgos específicos. La efectividad de estas medidas debe ser revisada continuamente para garantizar la adaptación a los cambios en el entorno operativo de las organizaciones (ISO/IEC 2023, 19).

Las organizaciones deben fomentar una cultura de gestión de riesgos que se manifieste en el comportamiento y prácticas de todos sus miembros, para esto es necesario el compromiso de la alta dirección (ISO/IEC 2023, 18), la conciencia y la formación de los colaboradores (ISO/IEC 2023, 33), una comunicación abierta que garantice la transparencia en la gestión organizacional (ISO/IEC 2023, 33), un sentido de responsabilidad compartida

de los miembros de la organización, la clara sistematización de la gestión de riesgos en los procesos de decisiones, la evaluación y mejora continua y por último un enfoque de ética y responsabilidad social que permita concientizar sobre los impactos y afectaciones que podrían ocasionar los sistemas de IA a la sociedad en general (ISO/IEC 2023, 16).

A pesar de los numerosos beneficios, la adopción de nuevos estándares implica cambios significativos en la cultura empresarial, así como la adaptación de la infraestructura existente y una capacitación adecuada del personal para cumplir con los requisitos de la norma. Además, los costos asociados pueden representar un obstáculo considerable para las pequeñas y medianas empresas, lo que podría dificultar su adopción efectiva.

5.2. Norma ISO/IEC 27001:2022

Entre los desafíos importantes de la IA en las prácticas sanitarias está la gestión de la privacidad y la protección de datos personales. En este sentido, la creciente evolución de esta tecnología y su necesidad de datos masivos para su optimización aumenta el riesgo de violaciones de privacidad. De acuerdo a Murdoch (2021, 1), ni siquiera las asociaciones público-privadas han sido suficientes para garantizar la privacidad de los datos de pacientes. El riesgo se agudiza con la existencia de algoritmos sofisticados capaces de vincular datos de salud anónimos con individuos específicos desde bases de datos públicas y privadas. En el año 2019 un estudio determinó la re identificación por medio de marcos de ataque de vinculación *linkage attack framework* de 85,6 % adultos y 69,8 % niños de un estudio de actividad física, pese a la eliminación previa de información protegida (Murdoch 2021, 3); este hecho demuestra lo vulnerable de la información que consideramos confidencial.

Ante esta realidad las organizaciones de salud deberían implementar y promover prácticas que garanticen la privacidad de la información de los pacientes en proporción a los riesgos que enfrentan. Esto incluye la adopción de métodos de protección actualizados que aseguren su efectividad frente a sofisticados ataques de vulneración, especialmente tomando en cuenta que la información médica se considera muy sensible y está protegida legalmente en todo el mundo.

Existen varios estándares de seguridad de información que pueden apoyar a las organizaciones en la protección de sus activos digitales. Sin embargo, según Yungán-Cazar (2022, 3) la mayoría de ellos proporcionan solo requisitos sobre lo que se requiere, mientras

que la norma ISO/IEC 27001:2022 ofrece también una guía sobre cómo implementarlos y gestionarlos.

La norma ISO/IEC 27001:2022 se publica por primera vez en 2005 y se la revisa en 2013. Su tercera edición, lanzada el 25 de octubre de 2022, introdujo cambios significativos en ciertas cláusulas, como la 6 y la 9. No obstante, el cambio más notable se encuentra en el Anexo A, que ha sido reorganizado y reestructurado. Las 14 categorías anteriores se redujeron a solo 4, y el número de controles se disminuyeron de 114 a 93, de esta forma: 57 controles se consolidaron en 24, 58 fueron actualizados con modificaciones contextuales, y se añadieron 11 controles completamente nuevos (Protiviti 2022, 2).

El estándar ISO/IEC 27001:2022 en su tercera edición fue preparado por el Comité Técnico ISO/IEC JTC1, Subcomité SC27 (ISO/IEC 2022a, 4) y especifica los requisitos para establecer, mantener y mejorar en forma continua un sistema de gestión de seguridad de la información SGSI; se enfoca en la seguridad de la información, ciberseguridad y protección de la privacidad (ISO/IEC 2022a, 5). Su adopción es una decisión estratégica para las organizaciones, esta debe ser influenciada por sus necesidades, requisitos de seguridad, objetivos y estructura. Su implementación busca gestionar la preservación de la confidencialidad, integridad y disponibilidad de la información.

La norma detalla parámetros que deben ser establecidos, mejorados y mantenidos dentro del contexto de la organización sin importar su tamaño o sector al que pertenezca (ISO/IEC 2022a, 5). En consecuencia, esta debe determinar su estructura, propósito, partes interesadas y sus requisitos, además de límites y aplicabilidad del SGSI[17] para establecer su alcance (ISO/IEC 2022a, 7-8).

El compromiso de liderazgo garantizará el cumplimiento de las políticas y objetivos de la seguridad de la información y avalará su compatibilidad con la dirección estratégica, además de asegurar la integración de los requisitos del SGSI a los procesos de la organización y la disponibilidad de los recursos necesarios para el funcionamiento del sistema de gestión (ISO/IEC 2022a, 8-9).

En este sentido, la organización debe identificar riesgos y oportunidades que garanticen lograr los resultados previstos mediante acciones planificadas que los aborden y aseguren resultados consistentes y medibles (ISO/IEC 2022a, 9–11).

[17] Por sus siglas, Sistema de Gestión de Seguridad de la Información.

El tratamiento de los riesgos es imperativo, esta normativa ofrece en el Anexo A una tabla de controles organizacionales que deben ser considerados cuando las organizaciones definan su política de seguridad, además de formular un plan de tratamiento de riesgos en donde se incluya la aprobación de los propietarios de estos riesgos y la aceptación de riesgos residuales. Los controles definidos en el Anexo A se dividen en organizacionales, de personas, físicos y tecnológicos. En dicha sección, se incluyen controles para políticas de seguridad, roles y responsabilidades, segregación de deberes, gestión de riesgos, incidentes, activos, control de acceso, seguridad física, seguridad de la red, desarrollo seguro y pruebas de seguridad (ISO/IEC 2022a, 17–24).

En la determinación de los objetivos de seguridad, las organizaciones deben definir que se hará, los recursos requeridos, los responsables, el tiempo necesario y la forma en la que se evaluarán los resultados. Estos objetivos deben ser coherentes con sus políticas de seguridad y deben tener en cuenta los requisitos de seguridad de la información aplicables y los resultados de evaluaciones previas (ISO/IEC 2022a, 11).

Para el correcto desempeño del sistema de gestión, la organización debe determinar el perfil de competencia necesario para el personal involucrado en la seguridad de la información y garantizar que este sea conscientes de su política de seguridad y de su contribución al SGSI (ISO/IEC 2022a, 12).

Es necesario también para la organización garantizar la evidencia de su gestión mediante información documentada debidamente revisada y actualizada, además de asegurar la evaluación de su desempeño mediante monitoreo de los indicadores definidos, con el objetivo de mejorar continuamente la eficacia de su SGSI (ISO/IEC 2022a, 14–16).

Sin embargo, existen importantes retos en la implementación de esta normativa, según Fidel Maldonado (2024, 1–3), la carencia de un claro entendimiento de las implicaciones y requisitos de la norma puede convertirse en un obstáculo especialmente para las MIPYMES. Las organizaciones deben tomar en cuenta las inversiones no contempladas que podrían también afectar la adecuada gestión de los controles que son requisitos de la norma. Además, es necesario acotar el costo que implica para las organizaciones el disponer de personal capacitado con perfiles especializados en seguridad de la información que sean conscientes de la administración continua de riesgos. Para esto, es necesaria la comprensión y adaptación de la cultura organizacional en conjunto con un adecuado liderazgo y

compromiso de la alta dirección y de sus colaboradores para implementar un entorno favorable que brinde soporte a la continuidad del SGSI.

5.3. Normativas de soporte

En el ámbito del tratamiento de riesgos asociados a sistemas de inteligencia artificial, surge el estándar ISO/IEC 23894:2024. Este estándar ofrece directrices específicas para la gestión de riesgos en organizaciones que producen, implementan o utilizan productos, sistemas o servicios que integran inteligencia artificial. La norma ISO/IEC 23894:2024 amplía y adapta las directrices generales del ISO 31000:2018 para abordar las consideraciones particulares relacionadas con la IA. El documento establece los principios fundamentales de la gestión de riesgos, presenta un marco para integrar este enfoque en actividades clave de la organización y se centra en la implementación sistemática de políticas, procedimientos y prácticas para la gestión de riesgos en los distintos procesos organizacionales (ISO/IEC 2024).

En este contexto también es importante mencionar el estándar ISO/IEC 23053, el cual busca establecer un marco de terminología común y proporcionar directrices que faciliten la compresión de sistemas de IA y aprendizaje de máquina, el estándar brinda una estructura común para describir componentes y funciones de estos sistemas, aplicables a organizaciones públicas o privadas y de cualquier tamaño. La norma incluye definiciones y términos específicos e incluso herramientas que se relacionan con el desarrollo y uso de modelos de aprendizaje de máquina, el documento describe estructuras y detalla enfoques de aprendizaje automático y su dependencia con los datos de entrenamiento así como flujos de trabajo que abarcan etapas del ciclo de vida de un sistema de IA, además recalca la necesidad de gestionar en forma adecuada el uso de datos en los procesos de aprendizaje de estas herramientas (ISO/IEC 2022b).

De la misma forma, el estándar ISO/IEC 22989:2022 establece conceptos relacionados con la IA con el fin de facilitar su comprensión y comunicación entre las partes interesadas de las organizaciones, el documento incluye definiciones, términos clave y nociones que exploran la interdisciplinariedad de la IA en áreas como la informática, matemáticas, filosofía y ciencias cognitivas. Por otro lado, facilita la clasificación y comparación de distintas soluciones tecnológicas de IA en términos de confiabilidad,

seguridad y precisión en base a un enfoque práctico que pretende ser de utilidad tanto para expertos como para no expertos (ISO/IEC 2022c).

6. Ética en la IA: Responsabilidad y confiabilidad

Debido a que las leyes son específicas de cada región y considerando la rapidez con que evoluciona la tecnología de inteligencia artificial, es crucial que las organizaciones sean conscientes de los desafíos éticos y riesgos morales de estos sistemas hacia la sociedad. Según Benraouane (2024, 106) en las situaciones en las que no existe claridad legal, los marcos éticos deben convertirse en referencia para la toma de decisiones y en una guía entre lo que se debe y no se debe hacer. En este contexto, varias organizaciones internacionales han publicado marcos éticos, entre ellas la ISO con sus estándares: ISO/IEC TR 24028:2020 y el ISO/IEC TR 24368:2022, el grupo experto de alto nivel de la Comisión Europea, la UNESCO, la OCDE y también multinacionales como Microsoft, IBM, PWC, etc. Estos marcos coinciden en la necesidad de identificar e implementar principios éticos en toda la cadena de valor de las aplicaciones de IA (Benraouane 2024, 107).

La OMS (2024, 23) por medio de su grupo de expertos ha llegado a un consenso de principios éticos en el uso de IA en el área de la salud, estos se detallan a continuación:

Protección de la autonomía: los humanos deberían mantener el control de las decisiones médicas en el área de la salud y asegurar la privacidad y confidencialidad de la información del paciente a través de marcos regulatorios. Benraouane (2024, 111) sostiene que la inteligencia artificial debe orientarse hacia la autonomía humana en todos los aspectos de su diseño e implementación, enfocándose a empoderar a las personas y complementar su labor, en lugar de ejercer control, engañar o manipular su voluntad.

Promoción del bienestar humano, la seguridad y el interés público: los sistemas de IA en el área de la salud deben mantener altos estándares de seguridad, exactitud y eficacia mediante el uso apropiado de indicadores que permitan su mejora continua. Además, no deben ser usados en prácticas que resulten en algún tipo de daño físico o mental. En este contexto Benraouane (2024, 108) detalla tres tipos de daños, el individual que puede ocurrir por ejemplo en el caso de afectaciones particulares de reconocimiento facial, el daño colectivo que puede afectar a un grupo de personas de similares características y el daño social que puede ocurrir por ejemplo en la generación masiva de información falsa.

Aseguramiento de la transparencia, explicabilidad e inteligibilidad: se refiere a que los resultados y el proceso de toma de decisiones de los sistemas de IA en el área de la salud deben ser entendibles no solo por los desarrolladores sino también por los profesionales, pacientes, usuarios y reguladores. De acuerdo con Benraurane (2024, 112), la explicabilidad requiere cuatro principios: explicación, que se refiere a la evidencia y razonamiento sobre la manera como el sistema eligió una respuesta específica; significancia , que implica abordar la explicación en diferentes perspectivas para cada grupo de interés; precisión, la cual fomenta la confianza en las decisiones de estos sistemas; y, por último, conocimiento del límite, que se basa en la premisa de que una decisión no se considera explicable si se fundamenta en información que está fuera del alcance del sistema.

Asegurar la inclusividad y equidad: el diseño de los sistemas de IA debe considerar una amplia inclusión y equidad en características como edad, sexo, género, ingresos, raza, etnicidad, etc. Además, la disponibilidad de estos sistemas debe ser equitativo, con el fin de eliminar los sesgos y minimizar las disparidades de poder. Según Benraouane (2024, 109), el *Acta de IA de la UE* define la equidad como un compromiso para asegurar la distribución justa de costos y beneficios y a su vez proteger a individuos y grupos de sesgos y discriminación. El autor destaca que el sesgo no proviene de los algoritmos sino de los datos suministrados en su entrenamiento, este puede ser abordado con métodos de justicia algorítmica como la equidad contrafactual[18] o la paridad predictiva[19].

Promover la IA adaptable y sostenible: se refiere a que el uso de la IA en el presente no signifique el detrimento de los recursos del futuro, la OMS enfatiza que las tecnologías de IA deben promover la sostenibilidad en los sistemas de salud, el medio ambiente y los lugares de trabajo.

Fomentar la responsabilidad y rendición de cuentas: los sistemas de IA en el área de la salud deben ser usados bajo condiciones apropiadas y por personal capacitado, los desarrolladores deben incluir en sus algoritmos puntos de control para supervisión humana.

[18] La equidad contrafactual se basa en cuestionarse si la decisión del algoritmo hubiera sido la misma al pertenecer el individuo a un grupo diferente, pero todas las demás características relevantes se mantuvieran igual.

[19] La paridad predictiva consiste en asegurar las mismas condiciones de predicción para los diferentes grupos considerados en el algoritmo.

Por otro lado, las organizaciones deben implementar mecanismos de compensación para individuos o grupos que pudieren resultar afectados debido a las decisiones de la IA.

En este sentido, el NIST empezó en el año 2023 a trabajar en el desarrollo de regulaciones rigurosas que garanticen sistemas de IA seguros y confiables para sus aplicaciones en áreas críticas como la salud (Benraouane 2024, 105). De acuerdo a Bürger (2024, 1), pese a la transformación disruptiva de la IA en esta área, existe una brecha significativa entre la investigación y práctica clínica, el autor sugiere que esto se debe en gran parte a la falta de definiciones claras en términos como confiabilidad y confianza ,cuya malinterpretación conduciría a la desarticulación entre principios éticos y prácticas concretas, así como a la deficiente validación de estudios exploratorios en aplicaciones clínicas.

El valor de la confianza depende de las circunstancias bajo las cuales se la otorga, lo que podría ser perjudicial sin una exhaustiva reflexión sobre las capacidades del sistema de IA. Por otro lado, la confiabilidad debe entenderse como una propiedad observable, medible y verificable, resulta entonces importante establecer definiciones operacionales que incluyan orientaciones en la producción, medición y evaluación de la confiabilidad de la IA en una perspectiva socio técnica con colaboración interdisciplinaria (Bürger et al. 2024, 3).

La norma ISO/IEC 22989:2022 define la confiabilidad como la habilidad de los sistemas de IA para satisfacer las expectativas de las partes interesadas de forma verificable. Por su parte, Simion y Kelp (2023, 1) señalan que la confiabilidad de la IA se fundamenta en que esta permanezca a la altura de sus obligaciones funcionales durante el ciclo de vida del sistema. Esto implica que cada tipo de IA, especialmente cuando se enfoca en aspectos críticos como el cuidado de la salud, tendrán un conjunto de obligaciones que satisfacer para garantizar su credibilidad. En contraste, Burr y su equipo (2024, 15, 25) señalan que la confiabilidad en un sistema de IA no se limita a aspectos funcionales sino que debe abarcar valores éticos en su desarrollo e implementación mediante la conformación de comunidades que fomenten el diálogo sobre las mejores prácticas y estándares compartidos que integren la ética en el proceso de aseguramiento. En consecuencia, un sistema de IA confiable según la ISO_IEC 22989_2022 (2022c, 29–31) podría construirse sobre un marco coherente con los siguientes componentes clave:

Robustez: se refiere al sistema que mantiene un nivel de desempeño esperado bajo diversas circunstancias. Esto implica la capacidad de evitar tanto el *sub-ajuste*[20] como el *sobre entrenamiento*[21]. Un sistema robusto debe demostrar un buen rendimiento incluso con datos atípicos, asegurando que la desviación de su salida se mantenga dentro de un rango aceptable. En el contexto de un modelo de regresión, la robustez se traduce en la habilidad de manejar métricas de amplitud aceptables para cualquier entrada válida. Por otro lado, en un modelo de clasificación, significa la capacidad de evitar errores al clasificar entradas que son atípicas.

Fiabilidad: Se refiere a la capacidad de un sistema para proporcionar resultados confiables durante su operación, incluso cuando las entradas no coinciden exactamente con aquellas utilizadas en su entrenamiento. Esta característica puede verse afectada por la robustez, la generalización, la consistencia y la resiliencia del sistema. Además, la fiabilidad también está relacionada con la seguridad funcional del sistema. En este sentido, implementar respaldos de información o copias de seguridad aumenta la fiabilidad, ya que permite mantener la continuidad del sistema ante posibles fallos.

Resiliencia: Es la capacidad del sistema para recuperarse de un incidente, el sistema resiliente puede trabajar en un nivel reducido durante una falla menor, pero debe considerar estrategias para su recuperación completa, especialmente en áreas vitales como la atención médica o seguridad pública.

Controlabilidad: Es la propiedad del sistema que permite su manejo por un agente externo, esto implica contemplar mecanismos confiables de seguridad que determinen los niveles de acceso y autentificaciones.

Explicabilidad: Como se mencionó anteriormente, esta característica se refiere a la capacidad de un sistema para justificar los factores que influyen en sus salidas, de tal forma que los usuarios humanos puedan entender el razonamiento detrás de sus decisiones y es fundamental para asegurar que dichas decisiones no se basen en correlaciones espurias derivadas del proceso de entrenamiento.

---

[20] Se refiere a la situación en la que el modelo es demasiado simple para responder adecuadamente a situaciones de la vida real.

[21] Se produce cuando el modelo se desarrolló de forma rígida sobre sus datos de entrenamiento, pero presenta falencias al tratar con datos de la vida real.

En general, los algoritmos basados en reglas ofrecen un alto nivel de explicabilidad gracias a su estructura transparente, mientras que las redes neuronales profundas tienden a ser menos comprensibles debido a su complejidad inherente. Goktas y Grzybowski (2025, 15) proponen la utilización de técnicas como *Shapley Additive exPlanations SHAP* que pondera un valor a cada característica de entrada del modelo de IA en relación con su contribución a la salida. Por su parte, Bonacina (2017, 7) plantea la XAI[22], cuyo enfoque propone modelos y algoritmos más explicables y transparentes para los usuarios tanto en las predicciones como en la toma de decisiones. Sin embargo, la ISO 23894:2023 (2024, 21–22) acota que el exceso de explicabilidad y transparencia puede devenir en riesgos para la privacidad, confidencialidad y propiedad intelectual.

Predictibilidad: se refiere a la capacidad de un sistema para generar respuestas fundamentadas en suposiciones razonables. Esta propiedad de inferencia desempeña un papel crucial en la confianza y la aceptación por parte de los grupos de interés. Una mejora en la exactitud del sistema puede contribuir a reducir su nivel de impredecibilidad, lo que a su vez fortalece la confianza de los usuarios en sus resultados.

Transparencia: se refiere a la claridad y accesibilidad que el sistema brinda a sus grupos de interés con relación a sus metas, limitaciones, definiciones, diseño, asunciones, métodos de entrenamiento, procesos de aseguramiento de calidad, métodos de protección de datos, etc.

Sesgo y equidad: El concepto de sesgo se refiere a las variaciones en el comportamiento de un sistema de IA en diferentes contextos, es la tendencia repetible de un sistema de IA a reproducir resultados que favorecen en forma consistente a ciertos grupos sin fundamentarse en evidencias sino en errores inherentes al diseño. De acuerdo con Mittermaier et al (2023, 1–3), el sesgo impacta principalmente a poblaciones desfavorecidas ya que pueden ser objeto de predicciones algorítmicas poco precisas o subestimar sus necesidades de atención. En este contexto, Goktas y Grzybowski (2025, 9) mencionan un estudio que habría evidenciado diagnósticos erróneos en ciertos clasificadores utilizados en rayos X de tórax en poblaciones menos representadas en su entrenamiento, principalmente mujeres. Por consiguiente, es imperativo convertir la mitigación del sesgo en una prioridad, estableciendo controles en el desarrollo de los modelos de IA desde el entrenamiento hasta

[22] XAI Inteligencia Artificial Explicable por sus siglas en inglés.

su implementación práctica en ambientes clínicos. Esto requiere de esfuerzo y colaboración entre diseñadores, instituciones de salud y entes de regulación.

Capítulo segundo
Implicaciones y desafíos de las herramientas de IA en las áreas de diagnóstico automático, diagnóstico por imagen, salud mental y desarrollo de medicamentos

La aplicación de la inteligencia artificial (IA) en el ámbito de la salud presenta numerosas oportunidades prometedoras. Krishan et al. (2021, 176–83) clasifican las principales aplicaciones de la IA en dos grandes áreas: la investigación biomédica, que incluye el desarrollo de nuevos medicamentos, y la práctica clínica, que entre otros incluye el diagnóstico automático, la salud mental y el diagnóstico por imagen. En las secciones siguientes, se explorarán los desafíos éticos asociados con la IA en general y en particular en las áreas de diagnóstico automático, diagnóstico por imagen, salud mental y desarrollo de medicamentos. Este análisis busca esclarecer las desafíos y riesgos éticos que surgen de la integración de estas tecnologías en la atención sanitaria.

1. Desafíos, riesgos y limitaciones de la IA en el área de la salud

Según la OMS (2024, 11), numerosas herramientas de IA utilizadas en el área de la salud forman o formarán parte de los modelos multimodales de gran escala o LMM, los cuales se caracterizan por procesar y comprender múltiples formatos de datos simultáneamente y se adaptan a gran variedad de tareas y contextos, como por ejemplo analizar textos e imágenes para emitir una recomendación clínica. En consecuencia, la complejidad inherente de estos sistemas profundiza las implicaciones éticas y los hace propensos a mayores regulaciones debido a los riesgos sistémicos que pueden emerger con su uso. En este contexto, según Longo y O'Reilly (2023, 1662:166–67) el riesgo aumenta en los sistemas desbloqueados[23], aunque estos poseen la ventaja de optimizar su rendimiento a lo largo del tiempo, su comportamiento se torna estocástico y sus decisiones inesperadas, lo que podría resultar potencialmente peligroso para los pacientes. Por lo tanto, los riesgos de

[23] Los algoritmos bloqueados no se actualizan tras su implementación, mientras que los desbloqueados se adaptan y mejoran con el aprendizaje continuo(Longo y O'Reilly 2023, 1662:167)

estos sistemas están presentes en todo su ciclo de vida y deben ser asumidos por los diversos actores involucrados en el, así:

En la etapa de desarrollo, la responsabilidad de los sistemas de IA recae en los desarrolladores, quienes deben abordar aspectos como la calidad de los datos, la seguridad y privacidad de la información (World Health Organization 2024, 16).

En la fase de provisión del servicio, según la OMS (2024, 16), es el gobierno el responsable de definir los requerimientos y obligaciones tanto del desarrollador como de los proveedores, quienes deben contemplar los posibles riesgos asociados con los sistemas de IA destinados a prácticas sanitarias.

Finalmente, en la etapa de implementación, los riesgos se materializan en el uso de la aplicación ya sea por factores de impredecibilidad de la herramienta, por manejo inadecuado por parte del usuario o por variaciones en la salida del sistema a lo largo del tiempo (World Health Organization 2024, 16). Zhang y Zhang (2023, 6) sugieren asumir una responsabilidad compartida entre los médicos que están en la obligación de supervisar el desempeño de estos sistemas y los desarrolladores, fabricantes e investigadores que son responsables ante fallos inherente en su entrenamiento o programación.

Según la OMS (2024, 38), los riesgos de esta tecnología emergente que pueden convertirse en sistémicos se dividen en: riesgos para el sistema de la salud, riesgos sociales y riesgos en el cumplimiento con requisitos legales.

Entre los riesgos para el sistema de la salud están: la sobrestimación de los beneficios y desestimación de las amenazas, lo que resulta en que muchos productos sean recibidos en el mercado sin contar con evaluaciones rigurosas, esto se conoce como solucionismo tecnológico (World Health Organization 2024, 38) y puede incidir en que ciertos productos del mercado no solo incumplan con las expectativas sino que afecten la calidad del servicio ofrecido y la salud del consumidor. En este sentido, Elhaik (2022, 1) examina que ciertas herramientas utilizadas en el desarrollo de los modelos de IA, como el análisis de componentes principales (PCA) podrían no ser confiables, lo que plantea interrogantes sobre la validez de numerosos estudios en el ámbito de la investigación biomédica. En consecuencia, el autor señala que se requiere más rigor en el análisis, contrastar los resultados con diversas técnicas, además de la interpretación crítica y cautelosa que reconozca ante todo los posibles sesgos y limitaciones de las herramientas antes de realizar inferencias.

Por otro lado, la accesibilidad y asequibilidad hacia los LMMs puede aumentar la brecha digital y representar una barrera entre países de bajos, medios y altos ingresos (World Health Organization 2024, 39) o entre diversos sectores de la población . Según Jiménez y Torras (2020, 38–40) refiriéndose a las prótesis potenciadas con IA señala que al menos en un principio solo están al alcance de minorías con suficiente poder económico o su distribución es desigual, por ejemplo entre un militar y un ciudadano común, responsabilidad que recae en los estados.

Los sesgos sistémicos representan otro riesgo importante en el proceso de diseño y entrenamiento de estos sistemas y al no ser mitigados pueden derivar en la exclusión de grupos minoritarios y vulnerables (World Health Organization 2024, 39), Giovanola y Tribelli (2023, 2) van más allá en este contexto y señalan que estas herramientas no solo deben garantizar la ausencia de discriminación, sino también el respeto y reconocimiento por la individualidad de cada persona.

El impacto al empleo es también un riesgo que se debe considerar, debido a que si bien es cierto que las herramienta de IA pueden abordar la escasez de médicos, especialmente en lugares remotos, también pueden provocar la pérdida de millones de empleos (World Health Organization 2024, 39). Además, la dependencia hacia estas herramientas puede llevar a la vulnerabilidad de los sistemas de salud e incidir en la afectación de su confianza (World Health Organization 2024, 40).

Por otro lado, la OMS (2024, 42–47) señala que los riesgos sociales de los modelos a gran escala LMM van más allá del sistema de salud y no pueden ser abordados por una sola ley o política. Entre estos se puede señalar:

La concentración de poder de un pequeño número de empresas tecnológicas, lo que podría exacerbar desigualdades y limitar alternativas de desarrollo en beneficio de sectores sociales.

El impacto ambiental resultante de las grandes cantidades de energía y agua que demanda el entrenamiento de los LMM, lo que aumenta la huella de carbono y afecta en forma desproporcionada a comunidades vulnerables.

El desplazamiento de la autoridad epistemológica debido a la información falsa o sesgada sin contextos morales o razonamiento crítico que podrían presentar los LMM.

Las deudas éticas acumuladas por la priorización de las empresas hacia la rapidez de lanzamiento de estos productos pese a las consideraciones éticas y de seguridad.

El dominio empresarial que podría derivar en la falta de inversión pública, Matheny et al. (2019, 23) señalan que el desarrollo de estas tecnologías se ha enfocado más en el tratamiento de enfermedades que en la prevención. Esta tendencia, según los autores, se debe a que las innovaciones tecnológicas suelen estar impulsadas por criterios empresariales que priorizan la rentabilidad, la eficiencia y el retorno de la inversión.

La falta de transparencia en la evaluación y mitigación de los riesgos asociados a las limitaciones inherentes de estas herramientas, lo cual dificulta el cumplimiento de principios éticos. Zhang y Zhang (2023, 4) en este contexto identifican como un riesgo importante la opacidad de los sistemas de IA, y lo atribuyen a tres razones principales: secretos comerciales, complejidad técnica que privilegia el entendimiento a expertos en estos sistemas y la razón más crítica es la característica de caja negra de los algoritmos de aprendizaje profundo, limitación inherente por la naturaleza de estas tecnologías que afecta su credibilidad y genera desconfianza por parte de los médicos para su adopción. De acuerdo con Longo y O'Reilly (2023, 1662:168) la implementación de atributos de explicación y trazabilidad sobre los sistemas de IA es esencial para su aceptación y validación.

Según la OMS (2024, 40) entre los riesgos en el cumplimiento con requisitos legales están: los riesgos de ciberseguridad a los que se expone la información del paciente que puede comprometer su privacidad e incluso su salud mental. En este ámbito, Zhang y Zhang (2023, 5) además de la ciberseguridad identifican en la vulnerabilidad de los sistemas de IA factores como: errores en programación, falta de pruebas adecuadas o dificultades en la certificación de software.

Pablo Jiménez (2020, 38–40) acota que la seguridad debe ser impecable en el área de la salud debido a que un mal funcionamiento en los dispositivos médicos podría acarrear consecuencias graves en los pacientes. En este contexto, Zhang y Zhang (2023, 5) mencionan que entre el 2000 y el 2013 los robots quirúrgicos en EEUU estuvieron involucrados con al menos 1391 incidentes adversos que incluyeron 144 muertes. En este punto es pertinente mencionar los riesgos asociados a las capacidades sobrehumanas que pueden aportar ciertas prótesis potenciadas con IA a sus usuarios como son mayor fuerza, velocidad, resistencia y destreza, un ejemplo de estos son las prótesis cerebrales que podrían mejorar capacidades

cognitivas o las próstesis sensoriales que mejorarían inmensamente las percepciones. Estas podrían sentar una brecha entre humanos biológicos y robóticamente mejorados (Jiménez Schlegl y Torras Genís 2020, 40).

Existen leyes regulatorias y de protección de datos vigentes a nivel mundial que privilegian el principio de autonomía (Farhud y Zokaei 2021, 3). Sin embargo, según la OMS (2024, 41–42), se identifican en los LMMs violaciones recurrentes al reglamento General de Protección de Datos (GDPR) de la Unión Europea en aspectos como: el uso de datos personales sin consentimiento, la falta de transparencia hacia los usuarios, la inadecuada protección de la información, la publicación de información inexacta y el incumplimiento al derecho a la explicación del consumidor. Según Zhang y Zhang (2023, 11) el 15 % de violaciones de datos documentados en 2017 provinieron del área de la salud.

La mitigación de los riesgos presentes en todo el ciclo de vida de la inteligencia artificial en el sector salud conduce a importantes desafíos y limitaciones que deben ser abordados de manera adecuada para garantizar su efectividad, seguridad y uso ético (Pradhan 2023, 1). Aunque los términos desafíos y limitaciones se utilizan generalmente de forma indistinta, según Pradhan (2023, 2-3) los desafíos de la IA se refieren a barreras que pueden ser superadas mediante políticas, reglamentos o directrices, mientras que las limitaciones se centran en restricciones inherentes en la naturaleza de estas tecnologías.

En este contexto, la OMS(World Health Organization 2024, 2–17) identifica riesgos y desafíos según los siguientes tipos de aplicaciones en el área de la salud: Diagnóstico y cuidado clínico, aplicaciones centradas en el paciente, tareas administrativas, educación médica e investigación científica y médica.

Entre los riesgos en el área del diagnóstico y cuidado clínico se puede mencionar las respuestas inexactas, incompletas o sesgadas, especialmente por los sistemas con funciones conversacionales que reflejan sesgos provenientes de su entrenamiento, el cual habitualmente se lo realiza con datos más representativos de países con altos ingresos, entonces su respuesta será generalizada hacia países con diferentes realidades; también podemos mencionar las respuestas falsas que se conocen como alucinaciones cuyo peligro consiste en que son indistinguibles debido a las características de estos sistemas de producir información que parece real. Según la OMS (2024, 28) un estudio habría demostrado que los modelos de LLM realizan alucinaciones en un rango que va del 3 al 27 % incluso cuando el modelo es

entrenado específicamente con información médica. Además, muchos de estos sistemas no ofrecen pruebas de confirmación para verificar su precisión incluso en el área médica en donde la exactitud es crítica.

Por otro lado, la calidad de datos y los resultados sesgados también son un riesgo debido a que constituyen el resultado de un entrenamiento con información parcializada en cuanto a raza, etnia, sexo, edad, etc., según Zhang y Zhang (2023, 3) los grandes volúmenes de datos que se utilizan para entrenamiento generalmente no son estructurados ni heterogéneos. Además, los registros electrónicos de la salud suelen estar plagados de errores o información inexacta de examinaciones físicas. Olewade et al. (2023, 8) destacan que otra causa del sesgo es la utilización de datos insuficientes o no representativos de la población general. De igual manera, BaHammam y Chee (2022, 1-2) identifican riesgos en ciertos sets de datos, especialmente los de acceso público, como son: la baja calidad que podría derivar en investigaciones de bajo valor clínico, la falta de variables críticas al no estar siempre diseñados para responder preguntas específicas de investigación y por último el desconocimiento del contexto por falta de comprensión de detalles, lo que afectaría su interpretación.

En este sentido, según la OMS, el sesgo automatizado es consecuencia de la excesiva confianza por parte de los profesionales médicos hacia los resultados de sistemas automáticos LMMs, lo cual puede llevar a desatención de errores, deterioro de la evaluación crítica, reforzamiento de sesgos existentes y además influir negativamente en las decisiones clínicas, perjudicando el desarrollo de habilidades profesionales.

Un desafío importante en esta área es la diferenciación que debería existir en el resultado de LMMs en adultos y niños. Según la OMS (2024, 29), no es clara la generalización de los LMMs en la salud pediátrica, los desarrolladores deberían incluir en sus herramientas la información demográfica y edades de la población utilizada en el entrenamiento. Asimismo, los pacientes deberían ser informados sobre el papel de la IA en el diagnóstico, es decir si esta realizará el rol de asistente en sus respuestas o las generará emulando el criterio de un especialista. La OMS (2024, 30) también recalca que el paciente debería tener la potestad de retener parte o toda la información previamente compartida si es que así lo considera.

Por otra parte, las aplicaciones centradas en el paciente están encaminadas a orientar el auto cuidado del paciente en tareas como la administración de fármacos, nutrición, dieta, actividad física o incluso salud mental.

Entre los riesgos importantes en estas aplicaciones tenemos: las afirmaciones incompletas o inexactas proporcionadas por los LMMS a gente común (World Health Organization 2024, 32) que por su formación no tiene manera de contrastarla. Además, la manipulación que puede ocasionar el diálogo constante con los chatbots hasta el punto de persuadir al individuo a realizar acciones contrarias a su interés o bienestar (World Health Organization 2024, 32).

Por otro lado, según la OMS(2024, 32), la privacidad de los pacientes en estos sistemas está lejos de ser garantizada debido a que la información recibida por los LMMs generalmente se utiliza para mejorar los modelos de IA. A su vez, el deterioro de la interacción con terapeutas es un riesgo constante en el uso de LMMs debido a que esto reduce las oportunidades de los médicos en la promoción de la salud y podría significar un impacto en la empatía humana, especialmente considerando que este tipo de aplicaciones al no ser usadas en hospitales o centros de salud suelen simplemente ser introducidas al mercado sin pasar por las regulaciones clínicas convencionales (World Health Organization 2024, 33), lo que podría, entre otras cosas, profundizar la inequidad en el acceso a salud de calidad. En este contexto, según Farhud y Zokaei (2021, 3)la empatía y el entendimiento humano son cualidades que la IA no puede replicar, lo que constituye una limitación importante en estos sistemas, los cuales, de acuerdo con Tavory (2024, 2), mantienen un enfoque de responsabilidad que se encamina demasiado en la rendición de cuentas, pero este enfoque no aborda adecuadamente cuestiones críticas como la vulnerabilidad y manipulación emocional que resultan cruciales en el ámbito de la salud mental.

En cuanto a las labores administrativas y financieras en el área médica, los sistemas de IA se enfocan en las tareas rutinarias y otorgan al profesional de la salud más tiempo para la atención médica, entre sus riesgos están las inexactitudes o errores en transcripción, traducción o simplificación que podrían derivar en diagnósticos incorrectos y la inconsistencia en la generación de registros de salud resultante de ligeras variaciones de las instrucciones dadas por el usuario (World Health Organization 2024, 34-35). Además, Peco (2024, 5–12) identifica los siguientes riesgos de la IA generativa en las labores

administrativas: la desconexión entre prácticas y políticas corporativas que puede llevar a la falta de control sobre el contenido generado o el deterioro de la calidad de trabajo de los empleados que puede ocurrir cuando estos perciben que las herramientas de IA arrebatan el valor de su trabajo o no son adecuadas para las tareas que realizan; en este sentido, la falta de reconocimiento hacia los empleados por parte de la organización aun cuando se obtienen resultados eficientes puede llevar a sentimientos de frustración y desconfianza, Ante esta realidad, constituye un desafío desarrollar una cultura organizacional que fomente la sinergia entre empleados, procesos y herramientas de IA.

Siguiendo con los riesgos en los diferentes tipos de aplicaciones, la educación médica con LMMs se orienta a crear textos dinámicos, simular conversaciones que mejoran o simulan una comunicación médico-paciente, simulan escenarios de diagnóstico clínico y explican opciones de tratamiento. Esta práctica puede incidir negativamente en la comunicación entre profesionales y afectar la calidad de educación médica (World Health Organization 2024, 35).

Por último la investigación médica, en donde la OMS (2024, 37) identifica varios riesgos como: el vacío en la responsabilidad de una investigación que no puede ser asumida por una herramienta de IA, el sesgo generado por datos de entrenamiento provenientes solo de países con altos ingresos, la alucinación de estos sistemas al resumir o citar artículos académicos, el deterioro de la confianza en el uso de LMMS para actividades como la generación de revisión por pares, y la accesibilidad que puede afectar a científicos con menor poder adquisitivo que buscan participar en investigación médica.

De acuerdo con la OMS (2024, 50), los modelos de lenguaje a gran escala LMM en salud deben ser regulados por un sistema de gobernanza que asegure un equilibrio entre gobiernos y empresas. Sin embargo, la implementación de controles adecuados podría ser un desafío debido a que ambas partes enfrentan presiones competitivas.

En este sentido, la OMS (2024, 16-17) propone que los gobiernos colaboren a nivel global bajo el auspicio de las Naciones Unidas para desarrollar un sistema de gobernanza que regule estas tecnologías. Este sistema no solo debe abordar los riesgos presentes en la cadena de valor de las soluciones de IA, sino que también debe asegurar la responsabilidad de los gobiernos de cada país en la implementación de regulaciones adecuadas que incluyan principios éticos, derechos humanos y normativas internacionales.

La gobernanza internacional debería involucrar a todos los países y no solo a los que albergan a las grandes industrias de IA con el fin de integrar a todas las partes interesadas en una red multilateral (World Health Organization 2024, 17). Según Zhang y Zhang (2023, 9) se requiere que los marcos regulatorios aseguren la protección de los derechos de los pacientes incentivando la innovación responsable y evitando el temor a represalias legales desmesuradas hacia los desarrolladores.

2. Herramientas de IA en las áreas de diagnóstico automático, diagnóstico por imagen, salud mental y desarrollo de medicamentos y sus desafíos éticos particulares

2.1. Diagnóstico automático

Krishnan et al. (2021, 2) afirman que el campo del diagnóstico médico enfrenta varios obstáculos como la insuficiencia de profesionales y altos tiempos de espera para los pacientes. En este sentido, la implementación de herramientas de diagnóstico automático como son los asistentes médicos virtuales basados en IA aparecen como una posible solución, estos sistemas brindan apoyo para el diagnóstico clínico. Aunque según Krishnan et al. (2021, 2) su precisión sigue siendo un desafío en la actualidad, su potencial para democratizar el acceso al diagnóstico médico es considerable y prometedor.

De acuerdo a Wind y Kenny (2023, 1), los asistentes de salud virtuales basados en IA son programas que utilizan técnicas como procesamiento de lenguaje natural, aprendizaje de máquina y analítica de datos para proporcionar información de salud personalizada y monitoreo continuo, además brindan facilidades para la comunicación entre pacientes y proveedores de atención médica. Estos asistentes pueden interactuar con el paciente a través de múltiples plataformas como páginas web o aplicaciones móviles y ofrecer soporte 24/7, además pueden servir como puente de atención médica hacia poblaciones remotas emulando interacciones humanas.

En sus primeras etapas los asistentes médicos operaban con sistemas simples de telemedicina, limitándose a consultas telefónicas o a través de plataformas rudimentarias en donde un profesional de la salud ofrecía a los pacientes consejos básicos, la eficiencia de este servicio estaba restringida por la disponibilidad de profesionales, la conectividad y la

tecnología, lo que repercutía en la pobre experiencia del paciente y en consecuencia limitaba su adopción.

El avance de la tecnología facilitó el desarrollo de plataformas de telemedicina más eficientes en donde se implementó video consultas, lo que a su vez permitió interacciones más complejas y enriqueció la experiencia del usuario. No obstante, la introducción de la IA en los asistentes de salud virtual permitió el desarrollo de asistentes automáticos que no necesitan personal médico atendiendo en la plataforma, tienen la capacidad de analizar datos de salud del paciente en tiempo real, pueden trabajar 24/7 y ofrecer recomendaciones personalizadas de salud y monitoreo continuo de condiciones crónicas en combinación con dispositivos *IoT*; lo que representa una evolución proactiva en la atención médica (Wind y Kenny 2023, 4-5).

Según Wind y Kenny (2023, 9-11) estos asistentes mejoran la comunicación entre el paciente y proveedores de la salud de las siguientes formas:

Facilitan de consulta virtual y seguimiento optimizando la comunicación al coordinar citas médicas, enviar recordatorios e incluso realizar evaluaciones iniciales preconsulta con el objetivo de reducir tiempos y mejorar la eficiencia de entrega del servicio de salud.

Mejoran la adherencia al tratamiento del paciente por medio de explicaciones claras de su condición médica, tratamiento y medicación; ofreciendo al paciente incluso recursos interactivos para ayudarle a rastrear su progreso, lo que contribuye a mejores resultados, reducción de tratamientos fallidos y deserción.

Optimizan las tareas administrativas de proveedores de la salud mediante el manejo automático de tareas como agendamiento de citas, registro de pacientes y otros datos, ayudando a reducir la carga administrativa para liberar tiempo valioso que se puede emplear en atención médica, lo que deriva en aumento de la eficiencia del servicio y satisfacción del paciente.

Entre las aplicaciones que utilizan IA para la asistencia médica virtual están:

Ada Health, se trata de un asistente de salud que cuenta con una gran base de datos médicos que permite analizar los síntomas que proporcionan los pacientes desde su aplicación, misma que proporciona posibles causas y recomendaciones en base a los síntomas descritos; posee una interfaz amigable con enfoque conversacional que interactúa con el paciente como lo haría un médico. La aplicación se adapta a las necesidades individuales de

los usuarios teniendo en cuenta factores relevantes como su historial médico o antecedentes familiares y está disponible como aplicación móvil, lo que facilita su acceso en cualquier momento. Ada proporciona acceso 24/7 a su plataforma, lo que facilita la información médica inmediata sin tiempos de espera prolongados; en consecuencia, empodera al paciente ya que le permite tomar decisiones más comprensibles e informadas sobre su salud. La aplicación no solo la utilizan los pacientes para la autoevaluación de síntomas sino también en programas de triaje en hospitales con el fin de optimizar la clasificación de pacientes y la priorización de consultas (Wind y Kenny 2023, 11).

Por otro lado está Florence que fue diseñado específicamente para apoyar a los pacientes en la gestión de sus medicamentos y condiciones de salud, el paciente ingresa sus medicamentos y prescripción médica a la aplicación y esta envía recordatorios en el momento de suministrarlos, la aplicación permite al paciente registrar sus síntomas a lo largo del tiempo para que el sistema los monitoree e identifique patrones que representen posibles riesgos, además ofrece al paciente información fácil de entender sobre sus condiciones de salud y medicamentos en una interfaz amigable de interacción sencilla incluso para personas mayores o con dificultades tecnológicas. De esta forma, Florence facilita la adherencia al tratamiento en los regímenes de medicamentos y en consecuencia previene complicaciones mediante la identificación temprana de factores de riesgo que permite una intervención rápida de profesionales de la salud, además empodera al paciente en la gestión de su salud creando conciencia sobre su tratamiento y condiciones. Por otro lado, representa un ahorro en los costos de atención médica al reducir la necesidad de visitas médicas frecuentes y hospitalizaciones. El soporte que brinda Florence ha demostrado su eficiencia en pacientes con condiciones crónicas como la hipertensión, diabetes y otras enfermedades que requieren monitoreo permanente y suministro continuo de medicación (Wind y Kenny 2023, 5).

Otro proyecto del área es el sistema *Doctor Chatbot: Heart Desease Prediction* que consiste en un chat potenciado con IA que interactúa con el usuario por medio de algoritmos de procesamiento de lenguaje natural para recolectar información sobre el peso, talla, enfermedades preexistentes, antecedentes familiares síntomas, etc. Esta información es procesada por un algoritmo de clasificación con el modelo *máquina de soporte vectorial*[24]

[24] Es un algoritmo de aprendizaje supervisado que busca el mejor espacio que actúe como frontera de decisión (hiperplano) para separar diversas clases.

(SVM) previamente entrenado que determina la probabilidad de que el usuario padezca una enfermedad cardíaca. El resultado que despliega al usuario es probabilístico y va acompañado con sugerencias acerca del estilo de vida deseable del paciente. El sistema incluso puede ayudar a los usuarios a separar una cita médica al identificar riesgos significativos. La aplicación no receta fármacos y despliega un aviso de que su resultado no corresponde a un diagnóstico médico (Fernández et al. 2020, 1–10). En este sentido, esta herramienta no reemplaza la atención médica, sino que realiza un diagnóstico preliminar para advertir al usuario si su probabilidad es alta para el padecimiento de enfermedades cardíacas. Es importante mencionar que el sistema no brinda un 100 % de exactitud, lo que podría derivar en gastos innecesarios de dinero y tiempo (Fernández et al. 2020, 10). Esta aplicación actualmente se encuentra en fase experimental y ha sido entrenada utilizando un conjunto de datos limitados de acceso público. Para que esta herramienta pueda ser utilizada de manera comercial, sería necesario llevar a cabo un entrenamiento con un conjunto de datos considerablemente más amplio que incluya tanto imágenes como gráficos. Esto implicaría una inversión significativa y las licencias legales correspondientes (Fernández et al. 2020, 10).

Otra herramienta destacada en este campo es el Agente de Empoderamiento inteligente (IEA), un sistema diseñado para recopilar información relevante de la salud del usuario a través de un formulario (Longo y O'Reilly 2023, 1662:175), el sistema busca información en la web mediante algoritmos heurísticos de IA para luego evaluar la información de acuerdo a la calidad de su contenido, mismo que es ponderado según su confiabilidad y actualidad para obtener un índice de información de calidad. La salida del sistema es información comprensible y confiable acerca del problema consultado por el paciente, el sistema no pretende convertirse en un doctor sustituto sino empoderar al paciente en la obtención de información confiable disponible en la web (Longo y O'Reilly 2023, 1662:185).

Entre los riesgos, limitaciones y desafíos éticos en este campo están: el tratamiento al individuo en su singularidad y contexto. Las aplicaciones de IA podrían convertir al paciente en un *consumidor de atención*, planteando un dilema que refleja otro desafío en la salud pública: el equilibrio entre el bienestar individual y el de la comunidad (CCNE y CNPEN 2022, 49).

Por otro lado según Savulescu et al (2024, 5), la obsolescencia y deshumanización representan un gran riesgo entre los profesionales de la salud, ya que la adopción paulatina de esta tecnología podría llevar al reemplazo de sus roles por herramientas de inteligencia artificial, tal situación repercutiría en la descalificación de muchos médicos y especialistas. Sin embargo, el cumplimiento de estándares y la responsabilidad son fundamentales en este ámbito y al momento no pueden ser asumidos únicamente por sistemas de IA.

Por su parte, Wind y Kenny (2023, 12) identifican en este tipo de sistemas otros riesgos como la privacidad y seguridad de datos, debido a la sensibilidad de la información de los paciente. En este contexto, de acuerdo con Borna y su equipo (2024,6) la naturaleza de los modelos LLM son incompatibles con las políticas de protección de datos HIPAA[25], mientras que la información médica es sensible y debe cumplir con estas políticas. Por lo tanto, estos sistemas plantean preocupaciones en el manejo de datos confidenciales.

En este sentido, es imperativo que las recomendaciones generadas por los asistentes de salud sean precisas y alineadas con las pautas médicas para garantizar la confiabilidad, esto requiere de monitoreo y actualizaciones continuas de algoritmos, así como un entrenamiento riguroso que evite sesgos e inexactitudes que representen riesgo para los pacientes, quienes deben comprender y consentir el uso de sus datos mediante políticas transparentes que se evidencien en las operaciones de estos asistentes. Borna at al (2024, 1) subrayan la importancia de adoptar herramientas para evaluación de la comprensibilidad y utilidad de la información proporcionada por los asistentes virtuales como el PEMAT[26] con el fin de mejorar la calidad de sus respuestas.

A nivel social es un desafío ético importante asegurar el beneficio de estas tecnologías hacia comunidades desatendidas (Kraaijeveld et al. 2025, 1) para evitar el crecimiento de la disparidad en la atención médica. Además, Kraaijeveld at al (2025, 4) mencionan que la introducción de asistentes virtuales puede también influir negativamente en la relación paciente proveedor de salud y afectar la comunicación y confianza. Por último, la responsabilidad legal en caso de recomendaciones incorrectas que afecten al paciente, el

[25] Las normativas HIPAA, o Ley de Portabilidad y Responsabilidad del Seguro de Salud (Health Insurance Portability and Accountability Act, en inglés), son regulaciones en Estados Unidos que protegen la privacidad y la seguridad de la información médica de los pacientes.

[26]PEMAT en inglés Patient Education Materials Asssessment Tool evalúa la comprensibilidad de la información y la facilidad de acción que deriva de esta información (Borna et al. 2024, 1).

desafío surge en identificar el responsable entre el desarrollador del sistema de IA, el proveedor de atención médica o incluso el propio paciente. En este sentido, Borna y su equipo (2024, 1) recalcan que las interacciones con IA carecen de estándares rigurosos, a diferencia de los profesionales de la salud, quienes están obligados a seguir pautas éticas estrictas.

Ante estas consideraciones, Borna et al. (2024, 9) plantean la necesidad de implementar en estos algoritmos sistemas *clasificadores de la seguridad de diálogo* que sirven para detectar problemas contextuales en tiempo real cuando una respuesta es potencialmente perjudicial o inexacta para el paciente, además pueden ser programados para identificar sesgos en las respuestas generadas. Los autores consideran además imperativo incorporar principios éticos en el desarrollo de estos clasificadores mediante la implementación de directrices que disciernan entre consejos médicos o simples respuestas informativas, además enfatizan que estos sistemas deben actualizarse continuamente con nueva información y retroalimentación de las interacciones de usuarios con el fin de aprender de errores pasados; por otro lado, recomiendan la investigación multidisciplinaria en su desarrollo con la colaboración de expertos en IA, psicología, medicina y leyes con el fin de garantizar el abordaje de aspectos relevantes y garantizar el equilibrio entre innovación y responsabilidad.

2.2. Salud mental

Según la OMS (2022, 7) más de 1000 millones de personas sufren trastornos mentales, lo que representa uno de cada ocho personas a nivel mundial. Sin embargo, solo el 25 % de estas personas reciben atención adecuada en países con ingresos bajos y medios. Se estima además que la pandemia del COVID 19 ha incrementado la incidencia de ansiedad y depresión alrededor de 25 % a nivel mundial. En consecuencia, la salud mental que antes fue estigmatizada es ahora crucial y suscita una demanda inusitada en los sistemas de salud, creando un desequilibrio entre oferta y demanda de estos servicios (Olawade et al. 2023, 2). Ante esta necesidad, la IA ofrece herramientas innovadoras que prometen compensar ese desequilibrio, reducir estigmas y mejorar los resultados de tratamientos.

De acuerdo con Olewade et al. (2023, 3), las herramientas de IA utilizadas en salud mental pueden clasificarse en terapia basada en chatbots, aplicaciones de salud emocional y herramientas inteligentes de salud mental.

Los terapeutas virtuales basados en chatbots presentan las mismas ventajas que los asistentes médicos virtuales descritos anteriormente como accesibilidad y disponibilidad más allá de las barreras geográficas, escalabilidad en su interacción simultánea con múltiples usuarios que deriva en disminución de costos respecto a los de una terapia convencional e intervención personalizada con el paciente mediante tratamientos individualizados según el historial clínico. Estos asistentes pueden también interactuar con sistemas de monitoreo para la detección de patrones conductuales que representen alertas, además de proveer soporte continuo post terapia que resulta útil en vigilar el progreso de la recuperación del paciente y prevenir recaídas (Olawade et al. 2024, 1–7).

Ejemplos de aplicación representan Woebot y Wysa, estos son asistentes conversacionales basados en principios de psicología cognitivo-conductual que ofrecen apoyo en salud mental. A través de diversas técnicas y ejercicios, estas aplicaciones ayudan a los usuarios a comprender y gestionar sus pensamientos, emociones y comportamientos, abordando problemas como la depresión, el estrés y la ansiedad. Estos sistemas interactivos proporcionan un entorno de diálogo amigable y empático, lo que permite explorar el estado emocional del usuario y adaptar sus respuestas a las necesidades individuales. Además, registran patrones y avances en la salud mental del paciente, facilitando un seguimiento continuo y la identificación de posibles complicaciones. Entre sus principales ventajas, destaca su accesibilidad para aquellas personas que, por motivos de estigmas, barreras geográficas o económicas, no tendrían la oportunidad de acceder a servicios de salud mental. Otras aplicaciones conversacionales en el área son Talklspace y BetterHelp que conectan mediante AI pacientes con los terapistas alrededor del mundo que mejor se adapten a sus necesidades individuales (Olawade et al. 2023, 3).

Entre las herramientas de salud emocional Olewade et al (2023, 3) mencionan: Moodifit que ayuda al usuario a identificar patrones y estrategias para manejar sus emociones, Happify que ofrece variedad de actividades y juegos para mejorar el ánimo del usuario y en consecuencia apoyar en su bienestar y resiliencia, Headspace personaliza experiencias de meditación para cada usuario, Shine utiliza IA con terapia dialéctica de comportamiento para aprender las necesidades e intereses del usuario y brindarle inspiración diaria, DBT Coach utiliza terapia dialéctica de comportamiento para enseñar al usuario sobre cómo manejar sus emociones, pensamientos y conducta, Companion utiliza terapia cognitiva

de comportamiento para enseñar al usuario a identificar y cambiar patrones y comportamientos conductuales negativos, PTSD Coach que ayuda al usuario a manejar el desorden de estrés post traumático mediante varios recursos y herramientas. Por último, SuperBetter que apoya al usuario en la construcción de resiliencia y cumplimiento de objetivos mediante recursos lúdicos.

Por otro lado, las herramientas inteligentes de salud mental proveen diversos recursos innovadores a terapeutas y pacientes en aplicaciones variadas como visión artificial para la detección de emociones o predicción de progresión de enfermedades mentales en base a un tratamiento, además están las aplicaciones que monitorean ciertas actividades del paciente con el fin de identificar en forma anticipada patrones de comportamiento que podrían significar riesgos en su bienestar emocional. En este contexto, Rogan, Bucci y Firth (2024, 11) acotan que el comportamiento cotidiano del individuo es el reflejo de su estado mental, por lo tanto, su monitoreo fuera del consultorio resulta efectivo para detectar anomalías conductuales que preceden al deterioro cognitivo.

El monitoreo continuo se lo realiza mediante sensores pasivos que aprovechan la tecnología existente sea en teléfonos inteligentes con sus funciones de GPS, cámara, micrófono, acelerómetro, etc., o en pulseras biométricas para obtener información en tiempo real y sin requerir interacción directa del usuario. Estos dispositivos almacenan diversidad de datos como ubicación, actividad física o parámetros fisiológicos con el objetivo de inferir patrones de comportamientos a partir de la información recopilada (Nepal et al. 2024, 2-7), de esta forma, variables como frecuencia cardíaca, conteo de pasos, actividad electro dermal, temperatura de la piel, conductancia de la piel, actividad física, uso del teléfono, profundidad de respiración, tiempo sentado, eficiencia del sueño, etc., pueden coadyuvar en la detección de anomalías relacionadas al estrés, ansiedad, alteración del sueño, depresión, etc. (Nepal et al. 2024, 3).

Para procesar los datos extraídos por los sensores pasivos se utilizan varios algoritmos de aprendizaje de máquina (Shvetcov et al. 2024, 4) como regresiones lineales, KNN, árboles de clasificación o bosques aleatorios. Es necesario considerar en estas herramientas la posible pérdida de datos debido al movimiento u olvido a causa de la naturaleza de los dispositivos portátiles. Tomando en cuenta que la información perdida no es completamente aleatoria,

ignorarla podría resultar en una muestra sesgada, ante esto, un set de datos con amplia variedad en el tiempo podría mitigar el problema (Nepal et al. 2024, 6).

Un ejemplo representativo de este tipo de aplicaciones fue Mindstrong Health (Olawade et al. 2024, 3) , la cual ofrecía una plataforma móvil para apoyo en la salud mental de los pacientes. Los usuarios nuevos realizaban una evaluación inicial en donde la aplicación determinaba su estado de salud mental y sus necesidades individuales, luego, mediante sensores pasivos el sistema registraba variables como las interacciones del teclado en el teléfono inteligente o hábitos de sueño y actividad física con el objetivo de identificar patrones que representen alarmas en la salud mental del paciente y en base a esto ofrecer al usuario intervenciones personalizadas como ejercicios de terapia cognitivo conductual, técnicas de manejo de estrés, herramientas de atención plena y meditación o también proporcionaba recursos educativos sobre salud mental. La plataforma estaba diseñada para ser utilizada en conjunto con la atención médica, por lo que brindaba la posibilidad a los usuarios de compartir la información y progreso con sus proveedores de salud. Sin embargo, esta empresa fue adquirida recientemente por Otsuka Pharmaceutical quien viene desarrollando soluciones de salud mental. Actualmente existen otras aplicaciones similares como HumanITcare o Feel Therapeutics.

Otra herramienta inteligente de salud mental que sirve como apoyo en tele terapia es Kintsugi, la cual según Olewade et al. (2024, 5-6) utiliza análisis facial y de voz en tiempo real para ayudar a los profesionales de la salud a entender el estado emocional del paciente y mejorar la terapia. Kintsugi monitorea las variaciones en el comportamiento y emociones del paciente para detectar signos tempranos de angustia emocional. La aplicación además puede interactuar con los pacientes en un entorno de apoyo que fomenta la comunicación abierta.

Las aplicaciones de IA en la salud mental no están exentas de consideraciones éticas. El uso de sensores pasivos para el monitoreo puede invadir la privacidad de los pacientes, hacerlos sentir constantemente vigilados y afectar de alguna forma su bienestar psicológico. Ante esto, el paciente debe estar plenamente informado sobre los datos que se recopilarán, incluyendo quién los recogerá, cómo se utilizarán y con qué propósito. Es fundamental que el procedimiento sea claro, transparente y basado en el consentimiento voluntario (Nepal et al. 2024, 6).

El sesgo en el entrenamiento de los algoritmos puede aumentar la disparidad en el diagnóstico mental, incidir en el acceso a la salud mental e incluso influir en los resultados del tratamiento; ante lo cual, se debería profundizar en el desarrollo de estas aplicaciones y ampliar la participación de profesionales en salud mental y comunidades marginadas (Olawade et al. 2024, 7). Además, implementar medidas de transparencia como por ejemplo comunicar al usuario acerca de los grupos poblacionales usados en los entrenamientos para que este pueda inferir sobre el nivel de efectividad que la herramienta podría tener en su tratamiento y tomar decisiones informadas.

Otra de las posibles implicaciones de esta tecnología es el impacto en la relación terapéutica entre el médico y el paciente, lo que puede llevar a una pérdida del contacto humano y de la empatía. Según Olawede et al. (2024, 7), las herramientas de IA están desprovistas de empatía y comprensión , lo cual es vital en interacciones terapéuticas. Además, el usuario podría adoptar comportamientos deliberados con el fin de influir en los resultados, lo que resultaría en diagnósticos y tratamientos inadecuados. Una posible solución a este problema es la implementación de revisiones periódicas de los datos junto con el acompañamiento continuo por parte de un especialista, con el objetivo de fortalecer la relación terapéutica y garantizar una atención más integral (Rogan, Bucci, y Firth 2024, 10).

2.3. Diagnóstico por imagen

Existen soluciones de IA para analizar prácticamente todos los tipos de imágenes médicas y apoyar en el diagnóstico de casi todas las patologías (Maria 2023, 14). Entre los principales beneficios de la IA en esta área, está la rapidez que esta tecnología aporta en el procesamiento y análisis de las imágenes médicas frente a los métodos tradicionales que toman demasiado tiempo y son propensos a errores; este factor es crítico en situaciones de emergencia en los que cada segundo es importante. Por otro lado, es significativo destacar que los algoritmos basados en AI pueden identificar patrones y anomalías que permanecen ocultas para el ojo humano (Khalifa y Albadawy 2024, 2). Al respecto, Ablameyko (2023, 14) señala que los algoritmos de detección de cáncer en imágenes médicas desarrollados por la empresa Ezra están incrementando la capacidad de detectar tumores en etapas tempranas, lo que constituye un hito, puesto que así aumentan las expectativas de vida del paciente en

un 80 % en comparación con el 20 % de probabilidades de curación cuando el cáncer es detectado en estados tardíos.

De acuerdo con Eldin y Kaboudan (2023, 5) los algoritmos de IA para diagnóstico por imagen pueden cumplir las siguientes funciones:

- La clasificación, que facilita la identificación de la categoría a la que pertenecen las imágenes, como tomografías computarizadas (CI), resonancias magnéticas (MRI) y radiografías, al tiempo que permite distinguir las distintas partes del cuerpo que representan, como cerebro, pulmones, rodilla, etc.
- La detección de objetos, que permite localizar objetos de interés dentro de las imágenes médicas, tales como anomalías en las imágenes de RX o tumores en imágenes de MRI.
- La segmentación, que separa regiones específicas de interés en las imágenes médicas, esto permite delimitar el tamaño y forma de anomalías como tumores y monitorizarlos a lo largo del tiempo.

Por otro lado, Khalifa y Albadwy (2024, 4) identifican cuatro sectores en los que la IA muestra potencial en el ámbito del diagnóstico por imagen como son: el análisis de imagen, la eficiencia operacional y el análisis predictivo.

El análisis de imagen e interpretación, que optimiza el estudio de imágenes médicas y la identificación de patrones patológicos y anomalías, reduciendo al mismo tiempo errores humanos debido a la fatiga o el descuido y garantizando la exactitud de la interpretación. En este sentido, las redes neuronales convolucionales juegan un rol crucial en la detección de anomalías mediante la segmentación y posterior identificación de patrones difíciles de discernir para el ojo humano, como por ejemplo la identificación de márgenes de resección en el cáncer rectal en imágenes de resonancia magnética o la detección de nódulos pulmonares en TAC (Khalifa y Albadawy 2024, 4). Respecto a la reducción de errores humanos, Khalifa y Albadawy (2024, 4) recalcan que los diagnósticos de varios sistemas como el GRAIDS[27] superaron la tasa de exactitud de endoscopistas expertos en la identificación de lesiones cancerosas y tumores estomacales. Cabe acotar que el análisis automático de lesiones con asistencia de la IA generalmente se limita a áreas anatómicas

[27] GRAIDS: Gastrointestinal Artificial Intelligence Diagnostic System.

determinadas y se enfoca en tipos de lesiones específicas. En consecuencia, podría pasar por alto lesiones importantes. Ante lo cual se presenta el concepto de *Análisis Universal de Lesiones* (ULA) que busca identificar múltiples tipos de lesiones de diferentes órganos en una sola imagen. Una base de datos para este propósito es *DeepLesion dataset* que contiene un conjunto de lesiones a gran escala y permite el desarrollo de modelos capaces de detectar, clasificar y segmentar varios tipos de lesiones en diversas ubicaciones del cuerpo (Jin et al. 2021, 18–20). Sin embargo, este concepto aun es un objetivo en desarrollo debido a la dificultad del entrenamiento, la complejidad de su estructura y a los desafíos técnicos y éticos que se enfrentaría en la universalidad de la detección de anomalías.

La eficiencia operacional, debido al incremento de velocidad y precisión en el proceso de interpretación de imágenes médicas, este factor no solo es crítico por salvar vidas sino también por optimizar costos al reducir la necesidad de repetir exámenes y minimizar diagnósticos errados. En este aspecto, dichos modelos CNN están ayudando a incrementar la eficiencia del análisis de imágenes médicas, según Khalifa y Albadawi (2024, 10) un estudio detectó 8,4 % de nódulos pulmonares que no se hubieran diagnosticado sin IA, Por otro lado, Fan at al (2022, 7) encuentran que los algoritmos de IA mejoran el efecto de segmentación en las imágenes de tomografía en estudios de hernia de disco lumbar en comparación con métodos tradicionales, optimizando los costos debido a la rapidez y certeza del diagnóstico. En contraste, un análisis demostró que el estudio de interpretación radiológica asistido por IA no mejora significativamente la tasa de falsos positivos pero potencia enormemente el flujo de trabajo y reduce el tiempo de diagnóstico (Khalifa y Albadawy 2024, 10), por lo tanto, resulta crucial en situaciones de emergencia.

El análisis predictivo mediante datos históricos ayuda a pronosticar enfermedades potenciales, facilitando las estrategias de salud proactiva. Al respecto, un sistema de IA permitió diferenciar el cáncer de mama negativo de los fibroadenomas benignos en imágenes de RMN por medio de imágenes y características clínicas particulares del paciente (Khalifa y Albadawy 2024, 8), contribuyendo a la prescripción de tratamientos personalizados, tempranos y específicos.

Por último, el soporte de decisiones clínicas que contempla dos componentes: la asistencia en procedimientos complejos, en los cuales las aplicaciones de análisis de imágenes médicas con IA son capaces de ofrecer guías y recomendaciones basados en

análisis en tiempo real, estas herramientas resultan cruciales en la mejora de sistemas de navegación quirúrgica, lo cual incrementa la tasa de operaciones exitosas; por otro lado, la integración con otras tecnologías como los registros de salud electrónicos EHR, cuya consolidación proporciona al profesional de la salud una visión holística del paciente y de sus condiciones, lo que resulta clave en la personalización y efectividad de tratamientos (Khalifa y Albadawy 2024, 4).

Además de los CNN, la IA generativa también juega un rol muy importante en el campo del diagnóstico de imágenes médicas, especialmente con técnicas como las redes adversariales generativas y los autoencorders varicionales (VAE). La IA generativa se centra principalmente en la síntesis de nuevas imágenes médicas para aumentar las bases de datos que sirven de entrenamiento de las aplicaciones con el objetivo de mejorar su precisión y también ha mostrado excelentes resultados en tareas como segmentación de imágenes, predicción de resultados y detección de anomalías (Musalamadugu y Kannan 2023, 1–6).

Según el hospital de Houston (2022), las tres principales técnicas de imágenes médicas son: Rayos X, resonancia magnética y tomografía computarizada. A continuación, se describirá cómo la inteligencia artificial, a través de funcionalidades como el análisis de imagen, la eficiencia operacional, el análisis predictivo y el soporte en la toma de decisiones clínicas, está transformando estas áreas de imagen médica. Además, se incluirá la mamografía debido a su relevancia en la detección del cáncer de mama, que constituye la segunda causa de muerte de mujeres a nivel mundial (Khurshid et al. 2024, 2), el estudio identificará también los principales riesgos y desafíos éticos asociados con estas aplicaciones:

2.3.1. Radiología general

Las fracturas de cadera y pelvis son muy frecuentes en el área radiológica, sin embargo, factores como las complejidades anatómicas o distorsiones en la imagen resultan en una alta tasa de errores de diagnóstico que incrementan la morbilidad y mortalidad. En consecuencia, se han entrenado varios modelos CNN para el análisis de imagen y posterior detección de fracturas pélvicas, uno de estos estudios logró un área bajo de curva[28] de 0.980

[28] AUC Área bajo la curva por sus siglas en inglés, implica la capacidad del modelo para discriminar correctamente entre imágenes normales y patológicas.

en la detección de fracturas de cadera (Jin et al. 2021, 16). Según Jin et al (2021, 17) las comparaciones de rendimiento entre modelos de IA y radiólogos mostraron claras ventajas de la IA al identificar sitios de fracturas que fueron pasados por alto por los médicos.

Por otro lado, cada toma radiográfica a la que se somete el paciente, lo expone a altas dosis de radiación, lo que podría afectar potencialmente su salud, ante esto muchos profesionales deciden no repetir los estudios radiográficos incluso pese a la mala calidad de ciertas imágenes (Potočnik, Foley, y Thomas 2023, 3), situación que podría derivar en un mal diagnóstico. Ante esta realidad, la optimización de tomas radiográficas resulta crucial. Existen varias aplicaciones destinadas para este propósito, entre ellas Critical Care Suite, desarrollada por GE para sus equipos de RX móviles. Estos algoritmos son capaces de detectar patologías como el neumotórax, evaluar la correcta colocación de un tubo endotraqueal y realizar clasificaciones de triaje. Además, proporcionan un entorno de software intuitivo que permite la rotación de imágenes y la identificación de problemas como la subexposición y sobreexposición, sugiriendo factores óptimos al operador. Los estudios que se consideran críticos son etiquetados y priorizados automáticamente en la lista de trabajo del especialista ("Critical Care Suite 2.0" 2025).

Con el objetivo de optimizar la eficiencia de sus equipos de rayos X, Siemens desarrolló YSIO X.pree, que consiste en un sistema de radiografía digital que utiliza cámaras 3D para detectar automáticamente por ejemplo la región del tórax sin necesidad de que esté visible y realiza una colimación automática de esta región para limitar al máximo la dosis absorbida por el paciente (Rasche, ten Cate, y Habecker 2023, 1–5). Por su parte, Philips desarrolló el Radiology Smart Assistant que consiste en un software potenciado por IA que analiza en tiempo real las imágenes antes de ser almacenadas o enviadas a su interpretación, esto evita que el paciente deba regresar para una repetición de un examen fallido, también brinda retroalimentación a los operadores en la colimación del área a irradiarse y el correcto posicionamiento del paciente. Además, el sistema también arroja estadísticas del rendimiento del operador por medio de índices de aceptación y conformidad de las imágenes tomadas por ellos, lo que promueve la mejora continua de su práctica ("Radiology Smart Assistant | Philips Healthcare" 2025). En conclusión, la IA en la radiología general no solo apoya en el diagnóstico, sino que también contribuye a la optimización de la calidad de las imágenes radiográficas, juega un papel muy importante en la reducción de dosis de radiación absorbida

por el paciente y generan un ahorro significativo de tiempo para el operador, el paciente y el radiólogo, permitiendo una mayor eficiencia en la gestión de los servicios de salud.

Según Geis at al. (2019, 2) varias sociedades radiológicas incluyendo la RSNA y la ACR realizaron un documento de declaración de ética de la IA en radiología, la cual plantea tres ejes:

La ética en los datos que se centra en la manera en que estos se adquieren, gestionan y utilizan, de acuerdo con Geis at al. (2019, 4) el consentimiento informado es fundamental para su uso, además se debe asegurar su privacidad, protección y propiedad, especialmente en el trabajo interinstitucional. Los datos que se utilizan para los sistemas de IA deben ser objetivos, neutrales y transparentes para evitar sesgos y asegurarse de garantizar el acceso equitativo a estos por parte de todos los involucrados.

La ética de los algoritmos y modelos entrenados se basa en que los algoritmos toman decisiones basadas en su entrenamiento y las propiedades de los datos, por lo tanto, es fundamental para el diseñador reconocer y evaluar las causas de los posibles sesgos y preferencias humanas para mitigar su impacto. Geis y su equipo (2019, 4) recalcan que los modelos de IA no poseen capacidad de razonamiento humano. Por lo tanto, recae en los diseñadores y operadores la responsabilidad de evitar consecuencias no deseadas. En este contexto, es fundamental asegurar la transparencia en los sistemas de IA, lo que permitiría auditar la posible causa de un error o un resultado clínico adverso. Es crítico también asegurar el control y prevención de ataques cibernéticos a estos sistemas y evitar el riesgo de alteraciones mal intencionadas o afectación a la privacidad del paciente. Además, se acota que los algoritmos deben ser monitoreados minuciosamente en su flujo clínico para asegurarse que su rendimiento es continuo y estable en el tiempo (Geis, Brady, Wu, Spencer, Ranschaert, Jaremko, Langer, Kitts, et al. 2019, 5).

Por último, la ética en la práctica destaca la importancia de que los valores éticos guíen el cómo y cuándo aplicar la IA en la práctica radiológica. Geis et al. (2019, 5) exhortan a la comunidad radiológica a desarrollar códigos de ética que promuevan el uso de los datos clínicos hacia el beneficio de pacientes y al bien común y lo limiten cuando su fin es meramente financiero. Los autores también recalcan la responsabilidad del radiólogo en el cuidado del paciente mediante el entendimiento cabal de las herramientas de IA y su aplicación efectiva (Geis, Brady, Wu, Spencer, Ranschaert, Jaremko, Langer, Kitts, et al.

2019, 6). Además, Geis et al. (2019, 6) enfatizan en la necesidad de contar con procedimientos de monitoreo post implementación para identificar consecuencias no deseadas, asegurar la calidad y disponer de procedimientos para investigar causas y la aplicación de acciones correctivas que eviten la recurrencia.

2.3.2. Tomografía computarizada

La segmentación de estructuras pulmonares es esencial para el diagnóstico y tratamiento de ciertas patologías, en tomografía esta acción es relativamente sencilla en condiciones normales. Sin embargo, se vuelve mucho más complicada en presencia de patrones anormales como: consolidaciones, derrames pleurales o nódulos pulmonares, esto lleva a la necesidad de un enfoque de segmentación más sofisticado como el PHNNs[29] que mediante técnicas de aprendizaje profundo clasifica cada voxel de la tomografía de abajo hacia arriba de acuerdo a su nivel de detalle y contexto, logrando una precisión del 98,5 % (Jin et al. 2021, 7).

Por otro lado, las aplicaciones abdominales asistidas con IA juegan un rol crucial en la detección de varios tipos de cáncer como el pancreático, carcinoma hepatocelular, colorectal, adenocarcinomas, etc. El cáncer pancreático PDAC corresponde al 85 % de los cánceres pancreáticos y emerge como la segunda causa de muerte por cáncer en los EEUU con un pronóstico de vida del 9 % en 5 años (Jin et al. 2021, 13). La localización automática del páncreas es por sí misma un reto en el análisis de imágenes médicas debido a que varía de sujeto a sujeto en su forma, tamaño y localización en el abdomen; un desafío aun mayor es la localización de tumores pancreáticos, los cuales se presentan tamaños y formas distintas (Jin et al. 2021, 14). Sin embargo, de acuerdo con Jin et al. (2021, 14-15), el desarrollo de modelos de aprendizaje profundo está ayudando a mejorar la evaluación y tratamiento de este tipo de patologías, no solo a través de la segmentación del tumor sino también mediante el análisis predictivo que evalúa en los pacientes la probabilidad de metástasis y resecabilidad de los tumores, lo que sirve de apoyo en la planificación de tratamientos quirúrgicos. En este contexto, la plataforma Zebra Medical Vision ofrece herramientas basadas en IA que permiten identificar varias patologías como fracturas óseas, enfermedades cardíacas e incluso el cáncer en etapas tempranas. El sistema es capaz de procesar grandes volúmenes de datos

[29] PHNNs Progressive Holistically-Nested Networks.

y disminuir el margen de error humano, la aplicación ofrece una solución accesible para hospitales que no pueden disponer de radiólogos especializados (Iglesia 2024), lo cual podría incidir en la democratización de esta tecnología hacia áreas apartadas u hospitales con bajos recursos económicos.

Por otra parte, un estudio de la Universidad de Melbourne (2020, 4) identifica tres características con las que la IA puede apoyar la eficiencia de las técnicas de tomografía computarizada, estas son: velocidad de reconstrucción, reducción de ruido y reducción de artefactos. En este sentido, la aplicación de GE TrueFidelity es la primera aprobada por la FDA que utiliza un modelo de aprendizaje profundo denominado DLIR para reconstruir imágenes tomográficas de alta calidad a partir de imágenes tomadas con bajas dosis de radiación (Hsieh et al. 2019, 2). Además, ciertos sistemas como Siemens FAST Integrated Workflow, Philips CT Precise Suite y GE Revolution Maxima utilizan aprendizaje profundo para apoyar al operador en la identificación de la localización óptima del paciente dentro del gantry mediante cámaras 3D. La empresa Aidoc (2019, 1–6), por otro lado, ofrece gran cantidad de aplicaciones diseñadas para optimizar el análisis y el flujo de trabajo. Sus herramientas incluyen detección automática de lesiones en áreas sospechosas, optimización del flujo de trabajo con integración a sistema PACS y EMR y el monitoreo de anomalías. El sistema no solo ha mostrado su utilidad en la detección de nódulos que puedan significar cáncer, sino también en su clasificación, evolución y explicación.

Pese a las grandes ventajas que la IA aporta a la tomografía computarizada, se necesita más desarrollo, diversificación de datos de entrenamiento y validación efectiva de los algoritmos (Lee y Seeram 2020, 2–5) para garantizar su confiabilidad y eficacia en entornos clínicos . Según Yan et al. (2023, 9), los métodos de reconstrucción tomográfica basados en IA son no solo capaces de generar imágenes de alta calidad con mínima dosis de radiación, sino también de reducir el tiempo de reconstrucción y de aprender de los datos de entrada. Sin embargo, necesitan de grandes volúmenes de datos para su entrenamiento, su rendimiento puede mostrar inestabilidades y presentan baja capacidad de generalización hacia nuevos datos o diferentes escáneres. Al respecto Jin et al. (2021, 12) señalan que un gran obstáculo en el desarrollo de la IA en este campo es la falta de datos etiquetados a gran escala para entrenamiento y evaluación. Los datos públicos existentes son limitados y carecen de anotaciones completas, los autores sugieren el trabajo conjunto de una comunidad de IA en

la que los radiólogos colaboren para definir patrones patológicos y los etiqueten correctamente. Además, el desarrollo de enfoques que permitan el entrenamiento a partir de datos etiquetados débilmente o la exploración de minería de datos para extraer instancias no etiquetadas de múltiples conjuntos de datos heterogéneos con el fin de mejorar la generalización de los sistemas en el reconocimiento de patrones.

Pietrowski et al. (2021, 7) revelan en su estudio sobre aspectos éticos de la IA en radiología que sobre todo se observa la carencia de directrices rigurosas acerca de la inclusión y exclusión de pacientes y la privacidad de su información en la investigación y desarrollo de estas herramientas. Además, muy pocas investigaciones discuten los patrones de predicciones incorrectas y las posibles razones detrás de estos errores. Pietrowski et al. (2021, 8) proponen el establecimiento de directrices uniformes para la elaboración y reporte de investigaciones sobre las herramientas que utilicen IA para garantizar estudios robustos que permitan aumentar la confiabilidad de la comunidad científica y de los profesionales médicos en estas tecnologías emergentes.

2.3.3. Resonancia magnética

Conforme a Hernanez-Trinidad et al. (2024, 3) el diagnóstico por RMN requiere por lo general el análisis de gran cantidad de imágenes cuya interpretación manual resulta laboriosa y propensa a variabilidad inter observador. En este aspecto, la IA tiene la capacidad de procesar grandes volúmenes de imágenes con rapidez y precisión, lo cual puede facilitar la detección temprana y monitoreo de diversas patologías como por ejemplo tumores cerebrales y otros tipos de cáncer, también se ha demostrado su utilidad en la detección y diagnóstico de enfermedades neurológicas como el Alzheimer o patologías cardíacas (Hernandez-Trinidad et al. 2024, 12). Por otro lado, la IA también ha mostrado precisión en la segmentación de lesiones en imágenes RMN, lo que resulta crucial para su correcto diagnóstico y tratamiento, también se puede mencionar el análisis de funcionalidad cerebral en el que las herramientas potenciadas con IA son capaces de mapear en forma dinámica las regiones del cerebro que se activan durante tareas específicas (Hernandez-Trinidad et al. 2024, 6).

Por otro lado, Según Potočnik et al. (2023, 5) los altos tiempos de escaneo de resonancia magnética plantearon la necesidad del desarrollo de algoritmos de aprendizaje

profundo que los aceleren. En este sentido, General Electric y Philips han logrado reducir el tiempo de escaneo entre 50 % y 66 % con la implementación de sistemas de aprendizaje profundo en sus resonadores. Por su parte, Siemens ha logrado escaneos 70 % a 80 % más rápidos. En equipos tradicionales el examen de resonancia de una rodilla tomaba aproximadamente 10 minutos mientras que en un modelo de esta marca potenciado por IA toma menos de dos minutos. Esto se logra en parte debido a algoritmos de IA como los GAN empleados para corregir artefactos derivados del movimiento junto con otros algoritmos entrenados con grandes sets de datos de imágenes MRI para reconocer puntos de referencia anatómicos que ayudan a optimizar el posicionamiento del paciente en el escáner con el objetivo de minimizar la variabilidad de la calidad de imagen por dependencia del operador y estandarizar la configuración inicial del paciente (Foti y Longo 2024, 4). Sin embargo, Foti y Longo (2024, 6) consideran que si bien las altas velocidades de escaneo facilitan enormemente el proceso y lo hacen menos incómodo para el paciente, pueden dar lugar a la pérdida de información crítica en las imágenes, comprometiendo la exactitud del diagnóstico. Además, los autores en sus estudios identificaron metástasis milimétricas creadas artificialmente en el proceso de reconstrucción con aprendizaje profundo, lo que podría generar falsos positivos que alterarían la evaluación del paciente, situación que es especialmente grave en pacientes oncológicos, cuyo mal diagnóstico puede incidir en un drástico e innecesario cambio en su tratamiento.

Según Hernandez-Trinidad (2024, 15), el etiquetado de imágenes de MRI resulta crucial para el entrenamiento de los modelos de IA, esta actividad demanda la participación de radiólogos y otros expertos. Sin embargo, la variabilidad inter observador podría estar presente en los datos de entrenamiento de los modelos de IA, afectando inadvertidamente el diagnóstico y tratamiento de pacientes. Por otro lado, la integración de los diferentes tipos de MRI: anatómica, funcional y de difusión resulta compleja debido a las diferencias de formatos y características, lo que representa un desafío en la confiabilidad de su representación y análisis holístico (Hernandez-Trinidad et al. 2024, 12). Además, la privacidad y seguridad en el manejo de los datos personales del paciente también representa un desafío importante en los sistemas de MRI, esta situación plantea un dilema debido a que es ingente la cantidad de datos requerida para el entrenamiento de los modelos de IA. Ante lo cual, en la práctica se utilizan herramientas para la creación de nuevos datos de

entrenamiento a partir de imágenes ya existentes (Hernandez-Trinidad et al. 2024, 14). Sin embargo, las nuevas imágenes obtenidas de esta forma mantienen las principales características de la primera y varían generalmente solo en aspectos de posición u orientación, más no eliminan la posibilidad de sesgo que también podría estar presente en esos sistemas de IA si es que los conjuntos de datos utilizados para su entrenamiento no son representativos, lo que perjudicaría la equidad en el diagnóstico y tratamiento.

Al respecto, Hernandez-Trinidad at al. (2024, 15) proponen el desarrollo y estandarización de métodos y protocolos utilizados para el etiquetado de imágenes y la anotación de información que sirve de entrenamiento a los modelos de IA o el uso de herramientas de anotación. Además, la implementación de auditorías de calidad que brinden retroalimentación y apoyen al mejoramiento continuo de la precisión en esta etapa de desarrollo.

2.3.4. Mamografía

Según el INEC, el cáncer de mama ocupó en el año 2018 el lugar número 11 en la lista de causas generales de muerte femenina en el Ecuador (2018), también es importante acotar que de acuerdo con la Organización Mundial de la Salud, la proporción de morbilidad por cáncer de mama está en función del índice de desarrollo humano. De este modo, en países con IDH alto una de cada 71 mujeres muere por esta enfermedad, mientras que en países con IDH bajo una de cada 48 mujeres muere por esta patología (OMS 2024).

En este contexto, la democratización en el uso de nuevas herramientas de IA en mamografía, podría en un futuro cercano incidir en la reducción de costos tanto para las instituciones de salud como para sus pacientes. Según Nafissi y Nahid (2024, 8-9) las herramientas de IA pueden facilitar la identificación temprana de cáncer de mama incluso en entornos donde los servicios médicos son escasos o inadecuados. Además, la capacitación de más profesionales de la salud respecto a las tecnologías de IA tendría como consecuencia una mayor concientización, educación y participación de comunidades vulnerables en programas de detección y prevención del cáncer de mama (Nafissi 2024, 2).

En la actualidad, el desarrollo de herramientas de análisis de mamografía basadas en IA avanza exponencialmente (Lamb et al. 2022, 1) debido al soporte que estas herramientas brindan al radiólogo principalmente en las siguientes aplicaciones interpretativas: Detección

de lesiones y diagnóstico, triaje, evaluación de la densidad de la mama; y en las siguientes aplicaciones no interpretativas: evaluación de riesgo, control de calidad de imagen, adquisición de imagen y reducción de dosis absorbida por el paciente.

La detección temprana del cáncer de mama o cribado es esencial para mejorar la esperanza de vida de las pacientes que lo padecen, los exámenes mamográficos tradicionales han cubierto de cierta forma esta necesidad pero con limitaciones en la dificultad a la que se enfrentan los radiólogos en el análisis de tejido mamario denso o la variabilidad en el desempeño de cada profesional, lo que puede resultar en diagnósticos erróneos y falsos positivos (Dave et al. 2025, 13, 27). Según Lamb y su equipo (2022, 2), la primera gran ventaja en el uso de técnicas de aprendizaje profundo es la capacidad de detección de características sutiles de la imagen y la relación entre características que no son percibidas por el ojo humano.

En este contexto, Escalante González (2023, 2) acota que la IA ha mostrado su efectividad en la detección del cáncer de mama, mejorando la tasa de detección y reduciendo la carga laboral de los radiólogos. De acuerdo con un estudio sueco, los algoritmos de IA detectaron, sin la existencia de falsos positivos, 20 % más de tumores de mama en comparación con los diagnósticos médicos (Escalante González 2023, 5), disminuyendo los procedimientos innecesarios que pueden no solo producir efectos secundarios en el paciente, sino también generar ansiedad. En esta línea , Ahmad y Alquarshi (2024, 11) detallan un modelo que logró una precisión del 98,61 % en sus resultados en la detección de cáncer de mama por medio de un sistema de aprendizaje profundo basado en RNN.

Según Dave at al. (2025, 15) algunos modelos de IA han alcanzado niveles de sensibilidad y especificidad mayores al 90 %, lo que evidencia su capacidad para detectar la gran mayoría de casos positivos de cáncer de mama, así como garantizar un menor número de mamografías clasificadas incorrectamente como positivas. Un ejemplo de software para detección temprana de cáncer es *Lunit Insight*, que ofrece una exactitud de 96 %, la aplicación está diseñada para detectar y clasificar varias anomalías en mamografías y señalar a los radiólogos las áreas que requieren su atención, este software funciona como herramienta de soporte para el médico y ha mostrado excelentes resultados en la aceleración de los procesos de diagnóstico y aumento de la eficiencia en el trabajo (Lunit 2025). Sin embargo,

aún necesita validación adicional en diferentes poblaciones (Dave et al. 2025, 18) para evitar el sesgo de sus resultados.

Otra aplicación de la IA en mamografía es el triaje, que permite priorizar pacientes con hallazgos sospechosos, demarcándolos en las listas de trabajo de los médicos. El objetivo del triaje es mejorar la eficiencia del flujo de trabajo permitiendo a los radiólogos que se enfoquen en los casos que requieren atención inmediata. Entre las herramientas comerciales que cumplen con esta función tenemos Siage-Q, cmTriage y HealthMammo, cuyos desempeños respectivamente corresponden a un AUC de 0,97,0,95 y 0,97 (Lamb et al. 2022, 4). En consecuencia, según Lamb et al. (2022, 9), la utilización de herramientas de triaje potenciadas por IA pueden contribuir a una atención más ágil y oportuna, los autores citan un estudio de caso que redujo de 9,6 días a 3,9 días el tiempo transcurrido desde la finalización del estudio hasta la entrega del informe. Por su parte, Yon y Kim (2021, 7) registran reducciones de tiempo de lectura de hasta 52.7 %.

La densidad del tejido mamario es un factor considerable en el riesgo de cáncer de mama y en la visibilidad de las lesiones en imágenes radiológicas. La densidad de mama se refiere a la cantidad de tejido glandular y fibroso en relación con el tejido graso. Al respecto, Bergan et al (2024, 2) aseveran que las mujeres con mamas extremadamente densas presentan un riesgo entre 4 y 6 veces mayor a aquellas con mamas menos densas. Sin embargo, existe amplia variabilidad entre radiólogos en la determinación de densidad mamaria, situación que puede afectar los resultados del diagnóstico. Su correcta valoración resulta importante porque puede influir en la determinación de posteriores exámenes como: resonancia magnética, tratamientos de quimio prevención o pruebas genéticas. En este contexto, la IA ofrece una oportunidad de mejora y estandarización de la evaluación de la densidad mamaria (Lamb et al. 2022, 12). Entre las aplicaciones aprobadas por la FDA que brindan esta funcionalidad están: DM-Density, Densityai, DenSeeMammo, Insight BD, PowerLook Density Assessment, Quantra, Visage Breast Density y Volpara Imaging Software (Lamb et al. 2022, 5).

Varias características del paciente son consideradas como factores de riesgo frente al cáncer de mama, entre las cuales figuran el segmento poblacional, antecedentes familiares, densidad mamaria, así como elementos presentes en la mamografía como masas, microcalcificaciones y diferencias entre ambas mamas. Estas variables constituyen la entrada

de modelos de IA predictivos que pueden determinar el porcentaje de riesgo a corto plazo del paciente. Existen algunas aplicaciones que proveen este soporte para predicciones de cribado individualizado en intervalos de uno o dos años (Lamb et al. 2022, 6). En contraste, el estudio de Gjesvik et al (2024, 7) sugiere que los algoritmos de IA pueden predecir el cáncer futuro de mama de 4 a 6 años antes del diagnóstico humano. Aunque estos autores aseveran que aún son necesarias más investigaciones para validar su precisión. Una aplicación de estas características es iBRISK (Breast Imaging Reporting and Data System*)* desarrollado para evaluar el riesgo de cáncer de mama en mujeres con hallazgos mamográficos sospechosos sin que deban pasar por biopsias (Ezeana et al. 2023, 1). El modelo utiliza aprendizaje profundo para analizar factores de riesgo como: antecedentes personales y familiares, obesidad, diabetes, hipertensión, edad, raza y etnicidad y descripción de lesiones observadas en las mamografías, el software permite clasificar a los pacientes en grupos de baja, moderada y alta probabilidad de malignidad, lo que ayuda al médico a determinar los pacientes que podrían evitar la biopsia o los que requieren aún más intervenciones (Ezeana et al. 2023, 2). Según Ezeana et al. (2023, 3) el software tiene la posibilidad de proporcionar un análisis de ahorros de costos potenciales al evitar biopsias innecesarias en grupos de baja o moderada probabilidad. La evaluación del software alcanzó una precisión del 89.5 % y un AUC de 0,93 con un intervalo de confianza del 95 % y se demostró que el modelo es capaz de reducir biopsias innecesarias en un 50 % en pacientes con riesgo bajo y moderado (Ezeana et al. 2023, 1), lo que implica un ahorro significativo en costos y la mejora en la calidad de atención.

Por último, la posición del paciente junto con la compresión adecuada de la mama en el examen y la selección de factores correctos son también en mamografía fundamentales para lograr buenas imágenes diagnósticas y disminuir la dosis de radiación absorbida por el paciente. Algunos equipos de mamografía como el modelo *Mammomat Inspiration* de Siemens, proporcionan al operador sugerencias e información sobre estas variables y visualización anatómica de la mama para evitar las repeticiones de imágenes, además proveen informes automáticos con métricas de calidad (Lamb et al. 2022, 6).

Entre los desafío éticos de la implementación de algoritmos de IA en mamografía de acuerdo con Yoon y Kim (2021, 11) están: la responsabilidad legal de los resultados, al respecto existe incertidumbre sobre cómo se manejarían las consecuencias legales y la

pertinencia de incluir en los registros médicos los datos analíticos proporcionados por la IA. En este sentido, Ezeana et al et al. (2023, 2) argumentan que la precisión en estos sistemas debe estar garantizada para evitar el riesgo de omitir diagnósticos inapropiados. Yoon y Kim (2021, 11) también plantean el dilema de considerar a la IA como lector independiente en el proceso de diagnóstico y la influencia de sus recomendaciones en las decisiones clínicas y el riesgo de caer en una dependencia excesiva hacia los sistemas de IA que podría resultar en la disminución de habilidades interpretativas de los radiólogos y afectar la calidad del diagnóstico.

Por otro lado, la integración apropiada de los nuevos sistemas de IA con otras herramientas tradicionales, sin comprometer la eficiencia del diagnóstico también es considerada por Ezeana at al. (2023, 8) como un desafío importante, sin olvidar el sesgo que siempre está presente en algún grado en los sistemas de IA. En este sentido, el iBRISK fue entrenado y validado con pacientes de la ciudad de Houston y sus pruebas restringidas a 3 grandes hospitales en Texas (Ezeana et al. 2023, 10), situación que pone en manifiesto la necesidad de un estudio más amplio para evaluar su desempeño. Al respecto, Dave at al. (2025, 3) señalan que la falta de representatividad en los conjuntos de datos, sesgados hacia ciertos grupos demográficos, puede limitar la generalización de los algoritmos de IA. Los autores sugieren ampliar el universo de entrenamiento de estos modelos con datos provenientes de múltiples instituciones y ubicaciones geográficas para mejorar su robustez y aplicabilidad (Dave et al. 2025, 14).

Según Yoon y Kim (2021, 11) la IA tiene el potencial para revolucionar las técnicas de mamografía. Sin embargo, esta tecnología aún se encuentra en etapas iniciales, su éxito en la práctica dependerá de la suficiente validación clínica que permita generalizar los diagnósticos, consensos sociales, el desarrollo de estándares tanto éticos como técnicos y la aceptación por parte de los profesionales de la salud. En este contexto, Dave et al. (2025, 14) destacan la importancia de desarrollar métodos que conviertan el proceso de toma de decisiones de la inteligencia artificial en información clara y accesible para los radiólogos. El objetivo de estas iniciativas es aumentar la confianza en los sistemas de IA y proporcionar transparencia en los procedimientos algorítmicos.

2.4. Desarrollo de medicamentos

El desarrollo de medicamentos puede ser radicalmente transformado mediante técnicas de aprendizaje automático, lo que impacta directamente en la reducción del tiempo de desarrollo y en la disminución de los costos asociados. En consecuencia, estas tecnologías pueden resultar significativas para acelerar la atención al paciente y derivar más inversión y recursos hacia enfermedades huérfanas[30] (Persons y MCGinnis 2019, 42).

Entre las técnicas de IA más utilizadas en este campo están el aprendizaje de máquina supervisado, no supervisado, semi supervisado y el aprendizaje reforzado, además se utiliza visión artificial para el procesamiento de imágenes biológicas e incluso PLN en el procesamiento de literatura científica , procesos moleculares o códigos de moléculas de ADN (Persons y MCGinnis 2019, 52).

En este sentido, las herramientas que utilizan PLN pueden acelerar significativamente la revisión de vasta literatura científica y priorizar la información relevante para identificar patrones (Persons y MCGinnis 2019, 56). Las técnicas de aprendizaje profundo se emplean para procesar grandes cantidades de información y datos complejos como imágenes, además muestran en general gran capacidad de predicción, empero, requieren de una gran cantidad de datos para su entrenamiento, esto suele ser un limitante para su uso, en contraste, los modelos de aprendizaje de máquina pueden ser entrenados con menor cantidad de datos (Persons y MCGinnis 2019, 56).

De acuerdo con Wu et al. (2024, 4) las redes neuronales recurrentes son particularmente importantes en esta área debido a su capacidad de analizar información en secuencia o en series de tiempo, así como por su cualidad de aprender de datos que contienen correlaciones hacia adelante y hacia atrás. Es menester mencionar que las técnicas de aprendizaje profundo pueden ser usadas también para generar nuevas moléculas en base a ciertas propiedades predefinidas. Un modelo de GAN que se basa en dos redes neuronales que compiten entre sí para generar datos auténticos fue capaz de producir potentes inhibidores para una enfermedad en particular en menos de dos meses, cuando este proceso toma habitualmente dos o tres años (Persons y MCGinnis 2019, 58–59).

En este contexto, las herramientas de IA pueden desempeñar un papel crucial en diversas etapas en el desarrollo de fármacos, partiendo desde la investigación, donde se

[30] Son enfermedades raras que no reciben mucha atención, en muchos casos debido a su costo.

exploran nuevas aplicaciones terapéuticas a partir de componentes ya conocidos, pasando por la etapa preclínica y hasta en las posteriores pruebas con humanos mediante la selección, reclutamiento y estratificación de pacientes para ensayos clínicos, optimizando así la eficacia de los estudios y la seguridad de los participantes (Persons y MCGinnis 2019, 42).

La etapa de investigación se centra en entender los mecanismos básicos de las enfermedades para luego en ellas identificar y validar objetivos biológicos como proteínas o genes, cuya reacción será estudiada con miles de compuestos químicos hasta obtener el resultado terapéutico deseado (Persons y MCGinnis 2019, 49). En este ámbito Chen et al. (2023, 10) mencionan que varias herramientas de IA como DLEPS proporcionan una visión holística de las vías biológicas involucradas en una enfermedad y sugieren nuevos objetivos terapéuticos en lo que se conoce como *análisis de datos multi-ómicos*. Por otro lado, también se utilizan modelos de aprendizaje profundo para la generación de nuevos compuestos farmacéuticos sin un esquema base previo en lo que se conoce como *novo drug design* (Chen et al. 2023, 9). Además, según Chen et al. (2023, 9) considerando que muchos de los objetivos de las medicinas son proteínas, la IA puede cumplir una importante función de predicción de su estructura mediante herramientas de aprendizaje profundo como el AlphaFold de DeepMind que predice la estructura 3D de las proteínas desde sus secuencias de amino ácidos.

Es relevante en este punto mencionar a la nanotecnología, la cual se está utilizando actualmente en el desarrollo de nanomedicinas debido a que por su tamaño pueden atravesar barreras infranqueables para los fármacos tradicionales. Como por ejemplo, el modelo de IA diseñado por Muñiz Castro (Chen et al. 2023, 9) que consiste en la impresión 3D de nanomateriales realizando predicciones de temperatura de extrusión, características mecánicas del filamento, tiempo de disolución y efectividad. Además otros modelos son capaces de predecir la efectividad de la absorción celular (Chen et al. 2023, 9) puesto que esta característica es crucial en la efectividad de la acción de las nanomedicinas.

En la etapa de investigación, la IA también se aplica para predecir propiedades de solubilidad, estabilidad y bioactividad entre varios compuestos (Chen et al. 2023, 10), lo que resulta determinante para la elaboración de fármacos.

Probablemente la etapa de investigación en la que la IA ofrece más potencial es el cribado o screening, que consiste en identificar la acción de compuestos o biomoléculas hacia

un objetivo biológico. El cribado de alto rendimiento se utiliza para analizar la acción de un fármaco frente a un gran número de compuestos, los ensayos se realizan en matraces de micro titulación y se identifica automáticamente su respuesta para luego seleccionar los compuestos más prometedores (Wu et al. 2024, 4). En la actualidad, de acuerdo con Wu et al. (2024, 4-5) se utiliza más el aprendizaje de máquina para el cribado virtual de compuestos con dos técnicas en particular: el cribado basado en el objetivo que mide el efecto del compuesto sobre el objetivo seleccionado y el cribado fenotípico que mide la acción holística del compuesto. Las técnicas de IA ayudan a identificar compuestos para pruebas posteriores y evitan el costoso cribado de extensas librerías del laboratorio. La efectividad en la selección de compuestos usando técnicas de aprendizaje de máquina es muy relevante. En una investigación sobre compuestos que activan una proteína cuya disfunción se asocia con fallas cardíacas, se comprobó que la tercera parte de las selecciones resultantes de modelos de aprendizaje de máquina fueron exitosas mientras que en un cribado masivo tradicional se obtiene una tasa de éxito aproximada de 0,01 %. Además, en este caso en particular el tiempo de desarrollo del fármaco se redujo seis veces (Persons y MCGinnis 2019, 58).

Por otro lado, en la investigación pre clínica, se comprueba la toxicidad del compuesto utilizando técnicas in vitro con tejidos o células en un tubo de ensayo in vivo con animales, si las pruebas son exitosas se puede pasar hacia las pruebas con humanos luego de obtener los permisos requeridos por ley (Persons y MCGinnis 2019, 50). En esta etapa , según Chen et al. (2023, 11), la IA presenta gran potencial de evaluar la toxicidad de compuestos químicos mediante modelos computacionales predictivos de redes neuronales como *DeepTox*, estos modelos se alimentan con datos históricos, características moleculares, propiedades fisicoquímicas y se entrenan con bases de datos de toxicidad como la Tox21.

El uso de estas herramientas ayuda a descartar compuestos altamente tóxicos antes de asumir los altos costos en pruebas in vitro o in vivo. Además, la optimización de ensayos clínicos se presenta como un ámbito propicio para la inteligencia artificial, especialmente considerando que la tasa de finalización de este tipo de ensayos con métodos tradicionales es de tan solo el 10 % (Gnanasambandam 2023, 2). Herramientas como Medidata Solutions permiten mejorar el diseño del ensayo clínico, determinar el tamaño de muestra ideal y definir criterios de inclusión y exclusión para la selección de la población. Además, facilitan el análisis de grandes volúmenes de datos en tiempo real, lo cual enriquece la toma de

decisiones mediante un monitoreo activo que identifica desviaciones del ensayo de manera inmediata. También contribuyen a la logística de recursos y ofrecen métodos estadísticos avanzados que se adaptan a la complejidad de los datos, garantizando resultados robustos que consideran los efectos secundarios y la relación riesgo-beneficio (Gnanasambandam 2023, 2–5).

Entre las principales aplicaciones comerciales descritas por Wu et al. (2024, 8) para el desarrollo de fármacos están las siguientes: DeepDrug que se utiliza para detectar el genoma de patógenos, Reinvent que se utiliza principalmente para generación molecular, AlphaFold por su parte se utiliza para predecir la estructura de las proteínas, SwissADME predice la bioavilidad oral de los compuestos, Admestar predice la toxicidad de moléculas de fármacos, AutoDockVina determina la conformación y afinidad de uniones, ADMETsar predice las propiedades de los compuestos, DeepPurpose asiste en la exploración de las potenciales características de los fármacos, PathPred predice la ruta sintética de compuestos y Deep 6 IA que asiste en la rápida identificación de potenciales participantes en pruebas.

Cabe mencionar que pese a las grandes ventajas que ofrecen la IA en el desarrollo de fármacos existen también importantes desafíos en su implementación y uso, entre ellos se puede señalar: los vacíos en la investigación, los cuales se originan en la falta de conocimiento sobre el mecanismo de las enfermedades o de la interacción de los medicamentos con ciertas proteínas debido a que estos objetivos muestran un comportamiento dinámico en el organismo (Persons y MCGinnis 2019, 64). Por otro lado, las librerías químicas contienen alrededor de 11 billones de compuestos sintetizables que apenas representan una pequeña fracción de la infinidad de compuestos posibles (Persons y MCGinnis 2019, 65). Además, la extrapolación de nuevos compuestos con características biológicas predefinidas creados con modelos generativos y usados para alimentar modelos predictivos podría producir resultados poco fiables (Persons y MCGinnis 2019, 65).

Otro de los principales obstáculos en esta área es la estandarización de los datos para el entrenamiento de los modelos de aprendizaje de máquina ya que estos provienen de diversas fuentes y en consecuencia se encuentran en diferentes formatos. Por lo tanto, normalizarlos y estandarizarlos requiere grandes recursos y un proceso intensivo (Persons y MCGinnis 2019, 66).

En consecuencia, según Pearsons y MCGinnis (2019, 67) la disposición limitada de datos puede ser una de las causas del sesgo algorítmico que alterara la confiabilidad de los resultados. Es relevante destacar que los datos pueden mostrar sesgos en función de etnia, género o clase social, lo que podría generar ineficiencias en la efectividad del fármaco resultante u orientar su eficacia hacia subpoblaciones específicas. Otra limitación poco considerada pero que influye en el sesgo es que la gran mayoría de información disponible proviene de ensayos que arrojaron resultados positivos, mientras que aquellos con resultados negativos permanecen ocultos. Además, la información sobre los pacientes también presenta sesgos, ya que se basa principalmente en individuos que reciben tratamiento, sin considerar a aquellos que carecen de acceso a atención médica.

De acuerdo con Persons y McGinnis (2019, 67), el acceso a los datos constituye también un reto muy importante en el desarrollo de fármacos debido al costo y a los trámites legales. Por otro lado, los pacientes suelen mostrar cautela en consentir el uso de sus datos para fines investigativos. Además, las compañías farmacéuticas consideran sus bases de datos como grandes activos que les brindan ventajas competitivas. Wu et al. (2024, 9) acotan que incluso el uso de tecnologías como el reconocimiento facial para la selección de participantes en pruebas pueden infringir la privacidad de los participantes. Estas restricciones podrían aumentar los sesgos en los modelos de IA. Sin embargo, existen formas legales para acceder a la información, principalmente para propósitos de salud pública.

Es importante tomar también en cuenta que la mano de obra calificada es escaza en el campo interdisciplinario de la ciencia de datos y la ingeniería biomédica. Por lo tanto, es un reto encontrar profesionales idóneos para la innovación, incluso las agencias reguladoras enfrentan este desafío al requerir personal con habilidades para entender y regular el desarrollo de medicamentos con tecnologías de IA (Persons y MCGinnis 2019, 68).

Aunque varios modelos de IA han mostrado gran eficiencia y confiabilidad comparados con técnicas tradicionales como la experimentación animal, consolidar esta innovación en la práctica requiere cambios legales (Persons y MCGinnis 2019, 69). De acuerdo con varias empresas farmacéuticas de EE. UU. las regulaciones en el uso de IA en el desarrollo de fármacos no están claras y no existe suficiente guía. Esta incertidumbre impide que aumente la inversión en estas tecnologías. En contraste, Brady (2024, 8) señala que la FDA y la EMA en sus respectivas jurisdicciones se encuentran realizando encuestas a

los desarrolladores con el fin de regular adecuadamente este campo sin llegar a sofocar la innovación. Además, estos organismos han publicado guías de orientación sobre el uso de IA en el ciclo de vida del desarrollo de fármacos. La prioridad de los organismos reguladores es asegurar altos estándares de seguridad y calidad mediante pautas que involucren la protección de los individuos que son parte de los estudios clínicos (Brady 2024, 6). Por otro lado, según Brady (2024, 5), la transparencia de los sistemas de IA utilizados en esta práctica también es de gran interés por parte de los reguladores, dada la alta complejidad de los modelos de IA, se considera crucial que los desarrolladores sean capaces de demostrar el proceso de toma de decisiones de sus modelos en contextos críticos como los ensayos clínicos.

Según lo descrito anteriormente, la IA brinda diversas soluciones en las diferentes áreas involucradas, así como también se identifican desafíos éticos importantes que deben abordarse para evitar consecuencias negativas y garantizar que esta tecnología se utilice en beneficio de la sociedad.

La siguiente tabla muestra las diferentes soluciones identificadas en las áreas referentes a esta investigación y las vincula con los desafíos éticos encontrados en la bibliografía consultada.

Tabla 1

Desafíos éticos de la IA identificados en las aplicaciones de las áreas de salud propuestas

Soluciones	**Aplicativos**	**Áreas involucradas**	**Desafíos éticos**
Atención virtual al paciente	-Ada Health -Doctor Chatbot: Heart Desease Prediction -Talkspace -Betterhelp -Moodifit -Happify -Shine -Companion -SuperBetter	Asistentes médicos virtuales Terapia basada en chatbots Salud emocional	-Mitigar el Sesgo y asegurar la equidad (Borna et al. 2024, 1) -Fomentar la Transparencia (Borna et al. 2024, 1) - Evitar la mercantilización de la salud(CCNE y CNPEN 2022, 49) - Evitar la obsolescencia y deshumanización(Savulescu et al. 2024, 5) -Asegurar la privacidad y seguridad de los datos(Wind y Kenny 2023, 12) -Evitar la disparidad en la atención médica (Kraaijeveld et al. 2025, 1) -Prevenir la erosión de la relación paciente-médico(Kraaijeveld et al. 2025, 4) -Evitar los resultados erróneos y su impacto en el paciente (Borna et al. 2024, 2)
Apoyo al paciente en la adherencia al tratamiento	-Florence	Asistentes médicos virtuales	
Monitoreo en el comportamiento del paciente	-Mindstrong Health	Herramientas inteligentes de salud mental	-Prevenir la invasión a la privacidad del paciente(Nepal et al. 2024, 6) -Reducir el sesgo algorítmico (Olawade et al. 2024, 7) -Fomentar la transparencia (Olawade et al. 2023, 7)

			-Evitar el deterioro en la relación paciente-médico(Olawade et al. 2023, 7) -Fomentar la privacidad y seguridad de la información(Olawade et al. 2023, 7)
Análisis de imágenes médicas para detección de patologías, triage y cribado.	-Critical Care Suite -Zebra Medical Vision iBRISK (Breast Imaging Reporting and Data System)	-Radiología convencional -Tomografía -Resonancia magnética -Mamografía	-Asegurar la privacidad y seguridad de la información(Geis, Brady, Wu, Spencer, Ranschaert, Jaremko, Langer, Borondy Kitts, et al. 2019, 4) -Promover la transparencia de los resultados(Geis, Brady, Wu, Spencer, Ranschaert, Jaremko, Langer, Borondy Kitts, et al. 2019, 5) -Asegurar la integridad en el diseño o errores de omisión (Geis, Brady, Wu, Spencer, Ranschaert, Jaremko, Langer, Borondy Kitts, et al. 2019, 5) -Prevenir la disparidad en el acceso a la tecnología(Geis, Brady, Wu, Spencer, Ranschaert, Jaremko, Langer, Borondy Kitts, et al. 2019, 5) -Minimizar el sesgo algorítmico (Lee y Seeram 2020, 5) -Prevenir la subjetividad clínica en el etiquetado de imágenes para entrenamiento(Hernandez-Trinidad et al. 2024, 15) -Asumir responsabilidad legal de los resultados(Yoon y Kim 2021, 11) -Prevenir la dependencia excesiva del profesional hacia los sistemas de IA (Yoon y Kim 2021, 11) -Evitar la insuficiente validación clínica(Yoon y Kim 2021, 11) -Garantizar la transparencia de los resultados(Dave et al. 2025, 14)
Reducción del tiempo de escaneo	-Deep Resolve de Siemens Healthineers -Sonic DL de GE Healthcare	Resonancia magnética	-Evitar la pérdida de información clínica y artefactos creados artificialmente en la reconstrucción de imágenes(Foti y Longo 2024, 6)
Cribado, determinación de toxicidad	-SwissADME -ADMETSAR -DeepPurpose	Desarrollo de fármacos	-Prevenir vacíos en la investigación (Persons y MCGinnis 2019, 64) -Gestionar las limitaciones de las librerías químicas(Persons y MCGinnis 2019, 66) -Asegurar la normalización efectiva de los datos de entrenamiento(Persons y MCGinnis 2019, 66) -Prevenir el sesgo algorítmico por disposición limitada de datos de entrenamiento(Persons y MCGinnis 2019, 67) -Gestionar la dificultad en el acceso a los datos debido a costo y trámites legales (Persons y MCGinnis 2019, 67) -Abordar la escasez de mano de obra calificada tanto para el desarrollo como para la regulación (Persons y MCGinnis 2019, 68) -Abordar la falta de claridad en las regulaciones legales(Persons y MCGinnis 2019, 69)

			-Asegurar la transparencia de los resultados(Brady 2024, 5) -Prevenir errores de programación en el desarrollo de aplicaciones (Wu et al. 2024, 9)
Selección de individuos para estudios clínicos	-Deep 6 IA -Trial Pathfinder -Criteria2Query -DQueST -TrialGPT	Desarrollo de fármacos	-Prevenir la transgresión de la privacidad del paciente con tecnologías como el reconocimiento facial(Wu et al. 2024, 9) -Prevenir el sesgo en la selección(Hutson 2024, 5) -Gestionar los resultados difíciles de reproducir(Hutson 2024, 5) -Evitar la alta dependencia de los investigadores hacia estos sistemas(Hutson 2024, 5) -prevenir la falta de transparencia en los resultados(Hutson 2024, 5)

Fuente: Borna et al., Brady, CCNE y CNPEN, Dave et al., Foti y Longo, Geis et al., Hernandez-Trinidad et al., Hutson, Kraaijeveld et al., Lee y Seeram, Nepal et al., Olawade et al., Persons y MCGinnis, Savulescu et al., Wind y Kenny, Wu et al., Yoon y Kim.
Elaboración propia

3. Gestión de los desafíos éticos de la IA en el área de la Salud

Existen diversas maneras de gestionar los desafíos éticos de la IA y diferentes normativas que orientan las prácticas de las organizaciones para lograr calidad, seguridad, cumplimiento normativo y ético, al tiempo que aseguran los requerimientos de los usuarios y de sus grupos de interés.

El estándar ISO/IEC 42001:2023 tiene como premisas dotar a las organizaciones que lo adoptan de las herramientas para generar confianza entre sus partes interesadas, garantizar el cumplimiento de otros marcos de AI y establecer una robusta estructura de gestión de riesgos. (Deloitte & Touche LLP 2023, 4). La RSS y el IFoA[31] (2019, 6) señalan que la IA tiene el potencial de actuar en beneficio y en detrimento de los individuos y de la sociedad. Por lo tanto, las organizaciones deben entender su trabajo y el impacto que este podría ocasionar a sus grupos de interés. En este contexto, el estándar ISO/IEC 23894:2023 (2024, 11) señala que las organizaciones inmiscuidas en la IA deben considerar otros aspectos externos e internos aparte de los estipulados tradicionalmente.

Por otro lado, según Benraouane (2024, 140) la planeación es la sección más crítica del estándar ISO/IEC 42001 debido a que contiene requerimientos importantes en referencia a la gestión de riesgos y se orienta a tres acciones principales: el proceso de evaluación del

[31] RSS es la sociedad Real de Estadística, IFoA es el Instituyo y Facultad de Actuarios.

riesgo y como llevarlo a cabo, opciones para manejarlo, la forma de abordarlo y por último, el proceso de evaluación del impacto en el sistema de gestión. En este sentido, Steimers y Sheneider (2022, 2) señalan que las metodologías tradicionales no bastan para gestionar los nuevos riesgos derivados de la IA. Ante esto, Banrauane (2024, 146) acota que la evaluación de riesgos adquiere mayor relevancia con la utilización del estándar ISO/IEC 23894. Además, Simonetta y Paoletti (2024, 4) destacan la importancia de integrar en este marco otras normativas de soporte como la ISO 27001 para asegurar la calidad y confianza.

De acuerdo con Benraouane (2024, 148) los datos estadísticos de implementación de marcos de gestión de riesgo en las organizaciones involucradas con la IA son alarmante debido entre otras cosas a la incertidumbre y ambigüedad de las regulaciones. En este sentido, para lograr una evaluación de riesgos significativa, es necesario analizar los grupos de interés que resultarían afectados con los sistemas de IA, así como sus preocupaciones, necesidades y el nivel de impacto (Benraouane 2024, 150). El RSS y el IFoA (2019, 11) mencionan que la confiabilidad de los modelos de IA debe construirse y mantenerse mediante un compromiso con las partes interesadas en el que incluya un diálogo honesto que ayude a la organización a comprender las fuentes potenciales de riesgo.

Benraouane (2024, 148) expone además que resulta muy complejo diseñar un marco de gestión de riesgo ad hoc debido a que el riesgo está presente en toda la cadena de valor de las organizaciones. Ante esto existen una variedad de marcos para la gestión de riesgos de la IA como el AI RMF 1.0, ERM Framework y otros de organizaciones privadas como Microsoft, BCG, Deloite, etc. Según Benraouane (2024, 148), el marco de gestión de riesgos de NIST se basa en principios similares a las ISO/IEC 42001 como son:

- La integración de los riesgos de la IA dentro del sistema de gestión de riesgos de la organización
- Implementar un enfoque integral que permita alinear las herramientas y prácticas de gestión de riesgo con la visión de la organización
- Personalizar el marco de gestión de riesgos en el contexto organizacional.
- Asegurarse de la inclusión efectiva de los grupos de interés internos y externos.
- Dotar de agilidad al sistema de gestión de riesgos para que responda apropiadamente al dinamismo regulatorio, económico y tecnológico

- Prestar atención al sesgo derivado de factores como percepción, interpretación, diferentes puntos de vista, aproximaciones, culturas, etc.
- Mejorar continuamente el marco de gestión de riesgos con el aprendizaje.

En cuento a los recursos para la gestión, de acuerdo con Ranjbar et al. (2024, 3) las organizaciones que utilizan conscientemente los sistemas de IA según el estándar ISO/IEC 42001:2023 necesitan mantener una infraestructura adecuada no solo con personal debidamente capacitado en tópicos como ciencia de datos, seguridad de la información y aseguramiento de calidad sino también mediante software de control certificado con posibilidades de actualización que permita validar sus soluciones de IA en el contexto de la organización para garantizar resultados adecuados. Para este propósito, también son importantes recursos tangibles conforme a las necesidades particulares de las organizaciones como: capacidad computacional, capacidad de almacenamiento e infraestructura comunicacional. En este sentido, el RSS y el IFoA (2019, 9-10) enfatizan la necesidad de minimizar la incertidumbre y el riesgo inherentes al desempeño del personal por medio de regulaciones rigurosas, capacitación y evaluación continua.

La planeación descrita anteriormente, está íntimamente relacionada con el control operacional que permite mantener, administrar y monitorear los diferentes procesos del sistema de gestión. Según Benraouane (2024, 180), el control operacional responde a la pregunta ¿Cómo la organización se asegura que los procesos y procedimientos implementados funcionen como se pretendía?, en consecuencia, requiere tres pasos fundamentales:

- Definir los controles necesarios para el funcionamiento del sistema de gestión. El anexo A proporciona una buena referencia de controles orientados a políticas, organización interna, recursos, evaluación de impacto, ciclo de vida, datos, información para partes interesadas, utilización y relaciones con terceras partes. La organización puede utilizar según sus objetivos varios de estos o elaborar sus propios controles.
- Establecer los criterios de monitoreo, control, y acciones correctivas.
- Documentar y planear el proceso de control para ganar confianza en el sistema.

Benraouane (2024, 180) recalca que el propósito de un proyecto de IA es solventar un problema. Sin embargo, la complejidad inherente en los sistemas de IA amerita la implementación de un marco de gestión en todo el ciclo de vida del proyecto. El ciclo de vida incluye el diseño, desarrollo e implementación del sistema. Una de las características del estándar ISO 42001 es que demanda un entendimiento profundo del contexto de la organización, debido a que coloca el ciclo de vida de los proyectos individuales en una relación holística con el liderazgo, políticas, recursos y compromisos (Ranjbar et al. 2024, 2).

En la fase de diseño, el desarrollador debe identificar el problema a resolver y alinear el desarrollo con el contexto de la organización mediante atributos de responsabilidad y confiabilidad.

Según Benraouane (2024, 185), es de vital importancia en esta fase tomar en cuenta la calidad de los insumos. Para los modelos de IA el principal insumo son los datos, por lo tanto, se los debe depurar y estandarizar con el fin de evitar problemas como pérdida de registros, valores extremos y valores atípicos que pueden dar lugar a respuestas incorrectas. El RSS y el IFoA (2019, 7) acotan que el riesgo en el uso de datos proviene de la utilización de información personal sensible o en el manejo de grandes volúmenes de datos sin consentimiento. En consecuencia, la organización debe definir como recogerá los datos y sus fuentes, además del tamaño apropiado del set de datos para entrenar el modelo de IA.

En este contexto, el anexo A del estándar ISO/IEC 42001 (2023, 29) demanda controles para determinar:

- En el anexo A.7.2: como la organización planea usar los datos para enriquecer su sistema de gestión.
- En el anexo A.7.3: como la organización obtiene los datos y como los selecciona.
- En el anexo A.7.4: como la organización controla la calidad de los datos y que procesos utiliza para asegurar los criterios de calidad.
- En el anexo A.7.5: de dónde vienen los datos.
- En el anexo A.7.6: cual es proceso en la preparación de los datos.

De esta forma, la ISO/IEC 4200:2023 se encamina a garantizar la calidad de los datos que la organización utiliza, porque estos inciden directamente en el comportamiento del

modelo y en consecuencia en sus resultados que podrían derivar en: falta de representatividad, sesgo, ruido o deficiencia en su calidad. Por su parte, la El RSS y el IFoA (2019, 7) identifican posibles consecuencias éticas en el mal manejo de datos como: pérdidas financieras, daños a la reputación, privacidad o psicológicos y exclusión de beneficios o servicios. En este sentido, Ranjbar et al. (2024, 4) señalan que las instituciones de salud están en esencia impulsadas por datos. Sin embargo, por lo general estas adolecen de prácticas efectivas para su gestión.

De acuerdo con Benraouane (2024, 155) los datos de calidad deben tener los siguientes atributos: sensibilidad, representatividad, precisión, accesibilidad, cobertura geográfica y temporal. En este sentido, Benraoune (2024, 155–57) recalca que distorsiones como el sesgo pueden aparecer en cualquiera de las etapas de diseño y desarrollo del modelo de IA:

- En la etapa de recolección de datos se puede presentar sesgo cuando estos son seleccionados mayormente de un grupo específico.
- En la etapa de preparación de datos el sesgo puede aparecer en la utilización de atributos de segmentación de la muestra como edad, sexo, nivel de educación, etc.
- En la etapa de planteamiento del problema puede aparecer el sesgo en forma subconsciente en la forma en que se plantea la pregunta, puesto que esta puede ser influenciada por contextos culturales y sociales.

También es importante tomar en cuenta la etapa de evaluación del modelo debido a que en esta se utiliza generalmente nuevos datos. Por lo tanto, resulta crítica la identificación de riesgos y la implementación de los controles.

Por otro lado, en la fase de implementación, el modelo utilizará información nueva, la procesará ya sea en una computadora, en la nube o en una combinación de ambas, lo que demanda un análisis de riesgos de seguridad y accesibilidad. En esta etapa es de especial importancia el monitoreo continuo del sistema, puesto que su desempeño puede variar en el transcurso del tiempo o desviarse de sus objetivos iniciales (Benraouane 2024, 187–88). Por lo tanto, es importante contar con indicadores de alerta que permitan identificar estas desviaciones en todas las etapas del diseño, desarrollo e implementación de los sistemas de IA.

De acuerdo con RSS y el IFoA (2019, 12) es además crucial en todo el ciclo de vida de las aplicaciones de IA garantizar la responsabilidad y rendición de cuentas nombrando responsables a cargo del desempeño del modelo y comunicar a las partes interesadas sobre la forma en que la organización asumirá dicha responsabilidad en caso de ser necesario.

Capítulo tercero
Recomendaciones basadas en la norma ISO/IEC 42001:2023 y otras normativas de soporte para la mitigación de los riesgos éticos identificados

En este capítulo se presenta un resumen de los desafíos éticos identificados en las áreas de esta investigación con el propósito de evitar la redundancia de sugerencias para gestionarlos, y se los vincula con los fundamentos éticos propuestos por la OMS (2024, 23) para validar su alineación con principios rectores internacionales.

Posteriormente, se proponen recomendaciones para la gestión de estos desafíos, fundamentadas en la norma ISO/IEC 42001:2023, con especial énfasis en los controles del tratamiento del riesgo del anexo A, tomando en cuenta las etapas de diseño y desarrollo e implementación y operación, con el fin de abarcar el ciclo de vida de las aplicaciones de IA. Estas recomendaciones no buscan fragmentar el carácter holístico del sistema de gestión, sino aportar sugerencias que suman valor y pueden alinearse e integrarse con el resto de los procesos del sistema.

Tabla 2
Resumen de desafíos éticos identificados y su relación con los principios éticos sugeridos por la OMS

Desafío Ético	**Principio OMS**
Reducir el sesgo y asegurar la equidad	Asegurar la inclusividad y equidad
Dotar de transparencia y explicabilidad a los sistemas de IA	Asegurar la transparencia, explicabilidad e inteligibilidad
Evitar la mercantilización de la salud	Promover el bienestar humano, la seguridad y el interés público
Evitar la obsolescencia y deshumanización	Proteger la autonomía
Asegurar la privacidad de la información	Promover el bienestar humano, la seguridad y el interés público Proteger la autonomía
Evitar la erosión de la relación paciente-médico y prevenir la dependencia excesiva del profesional hacia los sistemas de IA	Promover la IA adaptable y sostenible Proteger la autonomía

Prevenir la insuficiente validación clínica en los sistemas de IA y asegurar la precisión y reproductibilidad de los resultados	Promover el bienestar humano, la seguridad y el interés público

Fuente: Borna et al., Brady, CCNE y CNPEN, Dave et al., Foti y Longo, Geis et al., Hernandez-Trinidad et al., Hutson, Kraaijeveld et al., Lee y Seeram, Nepal et al., Olawade et al., OMS, Persons y MCGinnis, Savulescu et al., Wind y Kenny, Wu et al., Yoon y Kim.
Elaboración propia.

Tabla 3

Recomendaciones para gestionar mediante la norma ISO/IEC 42001 y otras normativas de soporte los desafíos éticos identificados en la investigación

Desafío/Etapa	**Recomendación**
Reducir el sesgo y asegurar la equidad /Diseño y desarrollo	Las políticas de la organización deberían priorizar la inclusión y diversidad (Anexo A.2.2), lo que puede materializarse, según los controles del anexo A.4.6, a través de la capacitación y sensibilización del personal sobre la importancia de estos valores, así como la identificación y mitigación del sesgo y la consideración del impacto social en los roles y responsabilidades (literales 5.2 y 5.3). El estándar ISO/IEC 23894:2023 (literal 5.4.1) destaca que los sistemas de IA pueden transformar la cultura organizacional mediante la introducción de nuevos roles y tareas, lo cual requiere preparación y adaptación dentro de la organización. Para asegurar la equidad y relevancia del sistema, es fundamental determinar y documentar con claridad las poblaciones destinatarias de la aplicación, así como analizar las implicaciones internas y externas de su uso (capítulos 4.1 y 4.3). La evaluación de impacto podría considerar explícitamente la inclusión, o exclusión, de ciertos grupos poblacionales en el alcance del sistema, conforme al literal 6.1.4. Adicionalmente, se recomienda dialogar con los grupos de interés para identificar posibles impactos del sesgo, tal como sugiere el estándar ISO/IEC 23894 (2024, 11). Ante esto, la organización debe mantener un enfoque realista hacia la mitigación del sesgo, reconociendo que su abolición total puede no ser alcanzable (Shwartz et al., 2022). El alcance definido por la organización debe ser comunicado, según el anexo A.8.2, a su público objetivo, incluyendo advertencias sobre los riesgos de utilizar el sistema fuera de dicho alcance. Herramientas como datasheets y cartas modelo facilitan la transparencia, proporcionando información sobre la representatividad, el sesgo de los datos y el rendimiento del modelo en diferentes grupos (Schuwart et al., 2022). La gestión del ciclo de vida de los datos debe ser documentada cuidadosamente, desde su origen y actualizaciones, hasta su preparación y modificación, garantizando la calidad, relevancia y cumplimiento de los requisitos definidos. Esto implica establecer procesos verificables para la limpieza, preprocesamiento y estandarización de los datos que aseguren su uniformidad y reduzcan los sesgos, según el anexo A.7.6 del estándar ISO/IEC 42001:2023. Asimismo, según el anexo A.7.2 es necesario evaluar su representatividad y complementarlos si corresponde. Se recomienda, siguiendo el anexo B.4.6, la colaboración de expertos en diversidad, justicia social y ética, quienes pueden aportar perspectivas valiosas sobre el impacto social del sesgo. Además, según el anexo A.7.4, la validación de los modelos de IA debe asegurar que sus resultados sean apropiados para todos los segmentos poblacionales destinatarios, definiendo métricas de evaluación y empleando datos de validación representativos. Los procesos de validación, de acuerdo con el literal B.7 de la ISO/IEC 23894:2023, deberían aplicarse al ciclo de vida completo de la IA, supervisando actualizaciones para preservar la calidad, confianza y equidad del sistema. Por último, la organización debe establecer canales accesibles de comunicación para que los usuarios, según el anexo A.8.4, reporten retroalimentación sobre el desempeño y posibles sesgos algorítmicos. La información recabada, según la cláusula 6.1.1, debe analizarse por personal capacitado, conforme a los procesos de gestión de riesgos, y las acciones correctivas comunicarse a los usuarios para fortalecer la confianza. Es fundamental implementar medidas de protección de

	datos en estos canales conforme a la ISO/IEC 27001:2022, dada la sensibilidad de la información manejada.
Reducir el sesgo y asegurar la equidad/Implementación y operación	La organización podría establecer en los procesos de compra o contratación, según el anexo A.10.3, los requisitos y criterios de selección relacionados al alcance de los productos o servicios de IA para asegurarse que cubre en su desarrollo las poblaciones a las que se atenderá con el objetivo de satisfacer las necesidades y expectativas de sus grupos de interés. Para esto es importante que la organización defina políticas y objetivos de equidad y transparencia, de acuerdo con el anexo A.9.3, estas políticas deben ser socializadas a toda la organización para que sean tomadas en cuenta no solo en la adquisición de la herramienta de IA sino también en su uso y evaluación. La organización podría de esta forma de acuerdo con el anexo A.9.4. Establecer roles, responsabilidades y autoridades que en el caso de la adquisición de soluciones de IA no solo exija a los proveedores la información necesaria que demuestre o acredite la validación de los modelos en las poblaciones objetivo de la empresa, sino también asegure experticia en el uso de esta tecnología. La organización, según la sección 9.1, también podría establecer indicadores estadísticos que permitan dilucidar posibles sesgos en los resultados de estas herramientas, mediante un proceso de monitoreo que utilice parámetros que evalúen la equidad de los resultados en las poblaciones objetivo en los términos que la organización defina, como, por ejemplo: género, edad, etnia, etc. Los indicadores podrían ser: tasas de falsos positivos/negativos, distribución de las predicciones en los diferentes grupos poblacionales o cantidad de quejas que la organización catalogue como discriminativas. En este sentido, es fundamental que se preste atención a la retroalimentación de los pacientes que puedan sentirse perjudicados o discriminados en algún resultado o diagnóstico. La organización debería informar al desarrollador, según el anexo A.8.5, los sesgos algorítmicos hallados mediante los canales de comunicación que este último ha implementado.
Dotar de transparencia y explicabilidad a los sistemas de IA Diseño y /Desarrollo	Para la implementación de explicabilidad y transparencia en los modelos de IA, la organización debería considerar en el diseño el uso de algoritmos XAI, definiendo, según el literal 6.1 y el anexo A.6.1.2 , los objetivos, riesgos y oportunidades que la adopción de la XAI aportaría al desarrollo de los modelos de IA y la posible repercusión de la explicabilidad en la precisión de sus algoritmos, el desarrollador también debería definir los niveles de exhaustividad de las explicaciones y su formato de salida, además de las herramientas a utilizarse para su desarrollo, entre ellas, según Ganatra et al (2024, 10–13): SHAP, LIME, AIX360, etc. Por otro lado, la organización, según el anexo A4, también podría determinar los recursos necesarios para la implementación de explicabilidad de sus modelos de IA como por ejemplo los datos, el tipo de modelo, las herramientas utilizadas etc. En el caso de los datos, la organización, según el anexo A.7.2, debería documentar métricas para la evaluación del impacto de ciertas características de datos en la explicabilidad del modelo realizando por ejemplo ponderaciones del peso de cada variable en los resultados. La evaluación y monitoreo continuo de estos factores, según el literal 9.1 y el anexo A.6.2.6, deberían servir para que la organización pueda tomar acciones hasta lograr sus objetivos de transparencia y explicabilidad. También es importante, según el anexo A.4.6 , definir y documentar los perfiles técnicos necesarios para el desarrollo de la XAI, determinando los objetivos del modelo IA en cuanto a la explicabilidad junto a su ciclo de vida para identificar la experticia técnica necesaria en cada fase, las competencias clave en este desafío ético pueden ser: dominio de herramientas de XAI como lime y shap, experiencia en evaluación de interpretabilidad y dominio en el área de salud específica para determinar el grado de explicabilidad necesaria en las aplicaciones .
Dotar de transparencia a los sistemas de IA /Implementación y operación	La organización podría evaluar, según el anexo A.10.3, el nivel de transparencia necesario en las herramientas de IA que utilizará con el objetivo de establecer los requerimientos exigibles al desarrollador y complementar, según el anexo A.2.2, sus políticas y objetivos. En este aspecto, también es necesario, según el anexo A.4.6, que la organización cuente con el personal debidamente capacitado para

	entender los procesos de decisión de estos sistemas y garantizar la eficiencia del trabajo. En este punto, es menester mencionar que el nivel de explicabilidad debe ser diferente en las herramientas de IA cuyo cliente es el médico a aquellas cuyo cliente es el paciente. Asimismo, es fundamental incorporar la intervención humana en los procesos de IA señalados en el literal 4.4, permitiendo, según el literal 9.1 y el anexo A.6.2.6 , que un profesional supervise y, si es necesario, modifique o invalide las decisiones adoptadas por los sistemas de inteligencia artificial., en este contexto, se deberían establecer, según el literal 7.4 y el anexo A.8.5, los canales de comunicación apropiados para el respectivo reporte y gestión, tanto hacia las autoridades designadas dentro de la organización como hacia el desarrollador del modelo si fuere el caso. En este aspecto, es importante asegurarse, según el anexo A.6.2.7, que la información necesaria esté disponible para sus partes interesadas como son: pacientes, entidades reguladoras, desarrolladores, etc. La organización, según el anexo A.6.2.4, debería implementar métricas de evaluación de explicabilidad para los sistemas de IA adquiridos, estas pueden ser: encuestas de evaluación a los profesionales involucrados, porcentaje de casos en los que el profesional logró establecer las justificaciones en las decisiones, etc.
Evitar la mercantilización de la salud /Diseño y desarrollo	La organización podría establecer, según el anexo A.2.2, un equilibrio entre los intereses comerciales y el acceso justo a sus productos estableciendo en sus políticas plasmadas en el literal 5.2 precios diferenciados, dirigidos hacia diferentes segmentos comerciales o utilizando licencias tipo GPL. Esta orientación requiere el establecimiento de objetivos que guíen el uso responsable de IA según el anexo A.10.4, cumpliendo con las metas de accesibilidad y equidad. Estos objetivos, según los anexos A.5.4 y A.5.5, deben relacionarse con la evaluación de impactos hacia los grupos de interés definidos en la sección 4.2. Por ejemplo, la evaluación de impacto de las políticas de precios de la organización frente a sus diferentes segmentos comerciales, con el fin de evitar exclusión hacia otros grupos y la evaluación de impacto en la seguridad de los datos y sus acciones correctivas basadas en la norma ISO/IEC 27001, debido a que los datos médicos sensibles podrían ser explotados comercialmente. Por último, la organización, según el literal 9.1 y el anexo A.6.2.6, debería establecer mecanismos de monitoreo en el ciclo de vida de la IA, por ejemplo, en la gestión de licencias y términos de uso de sus sistemas, con el objetivo de implementar ajustes en caso de consecuencias negativas.
Evitar la mercantilización de la salud /Implementación y operación	La organización que utiliza los sistemas de IA podría alinear, de acuerdo con los anexos A.2.3,A.9.2 y A.9.3 en sus políticas y objetivos corporativos aspectos sociales y de equidad en el acceso a sus servicios como: -Evitar la recomendación de servicios que no estén plenamente justificados con evidencia clínica. -Informar a los pacientes sobre las posibles alternativas y sus costos con el fin de evitar la venta forzada de paquetes de salud innecesarios, en este sentido, la organización debería garantizar que la decisión del paciente se basa en información completa y comprensible. -Evitar la publicidad engañosa sobre productos o servicios que la organización ofrece. -Evitar los incentivos al personal basados únicamente en ventas de servicios o productos, y mas bien basarlos en indicadores de calidad y satisfacción del paciente. -Esto requiere de un compromiso de la alta dirección, según e literal 5.1, con aspectos éticos y sociales que privilegien la equidad y rompan el paradigma de la maximización de la utilidad, estableciendo una cultura organizacional enfocada en el bienestar humano, en el aspecto financiero podría la organización privilegiarse de la cantidad de pacientes atendidos en lugar del alto precio facturado a unos pocos. Este enfoque podría ayudar a la institución a posicionarse en el cumplimiento de responsabilidad social incluso en el sector privado y lograr beneficios como la mejora de su imagen institucional, atracción y retención de talento humano, fidelización de pacientes, atracción de inversiones, reducción de riesgos legales, etc.
Evitar la obsolescencia	Entre los objetivos de diseño de los modelos de IA del anexo A.9.3 puede sumarse la colaboración humano-sistema utilizando criterios de la norma ISO 9241-210

Y deshumanización/diseño y desarrollo	(2010, 20) que incluyen principios de usabilidad, entre los que se encuentran la transparencia y explicabilidad, la controlabilidad por parte del usuario y la idoneidad para la individualización, según la cláusula 6.4.2.1 de este estándar, con el fin de asegurar la interacción efectiva del modelo con los humanos. El desarrollador puede implementar ciclos de evaluación iterativa de impacto del modelo de IA, hacia personas, grupos o sociedades según el anexo A.6.1.2 en términos de relevancia laboral o deshumanización e implementar acciones para prevenirlo, como por ejemplo el establecimiento de puntos de supervisión humana. Este enfoque requiere de la participación del personal técnico con la estrecha colaboración de especialistas en ética y aspectos sociales y la formación continua del personal en el contexto de los valores humanos.
Evitar la obsolescencia Y deshumanización /implementación y operación	El usuario podría añadir en sus políticas, según el anexo A.2.2 el uso responsable de las herramientas de IA incluyendo etapas de supervisión humana en las decisiones de estos sistemas, además implementar la evaluación continua del impacto de sus herramientas de IA, según el anexo A.5.2, en aspectos como la empleabilidad y la interacción social mediante indicadores como: percepción de confianza por parte del cliente, frecuencia de interacción humana, impacto en la moral de los empleados, etc. En este punto, la transparencia y explicabilidad son muy importantes para conservar la colaboración humano-máquina por medio de la comprensión., por lo tanto, la organización podría incluir estas características, según el anexo A.10.3, en los requisitos de compra de sus herramientas de IA . La retroalimentación del cliente hacia el proveedor también es indispensable para mitigar este desafío ético mediante actualizaciones de software, según el anexo A.6.2.6, en caso de presentarse desviaciones de sus objetivos éticos.
Asegurar la privacidad de la información /diseño y desarrollo	El desarrollador podría agregar el respeto a la privacidad de la información en sus políticas de responsabilidad documentadas según el anexo A.2.2 implementando mecanismos que garanticen la privacidad en las diferentes etapas del ciclo de vida del modelo de IA en base al estándar ISO/IEC 27001 anexo A, mediante medidas como: asignar roles y responsabilidades que controlen la seguridad de la información, mantener contacto con grupos de interés y foros especializados en seguridad, tratar la información como un activo y asociarla con sus propietarios, controlar la información entregada al personal u otras partes interesadas al culminar su empleo, contrato o acuerdo, establecer controles y derechos de acceso, etc. En este contexto, el desarrollador debería analizar el impacto de los riesgos de privacidad proveniente de los datos que utiliza en el entrenamiento y validación de sus modelos, según la sección 6.1.2 y asegurarse según el anexo A.7.3 en tópicos como adquisición y preparación, implementando reglas de aceptación de datos y documentando su origen, según el anexo A.4.3 y en correspondencia con las regulaciones de las regiones en donde se implementará la solución de IA, como por ejemplo: utilizar únicamente datos anonimizados o consentidos y evitar la utilización de datos sensibles que pongan en riesgo la privacidad del individuo o datos sin autorización o consentimiento. Es importante en el desarrollo, según el anexo A.6.2.4, realizar pruebas de privacidad y definir métricas como tasa de ataques de re-identificación que consiste en cuantificar el porcentaje de registros que pueden ser re identificados a partir de las salidas del modelo.
Asegurar la privacidad de la información /Implementación y operación	La organización, según el anexo A.10.3, debería exigir al proveedor información clara sobre la forma en la que sus herramientas de IA manejan los datos que generan sus pacientes y los niveles de seguridad de esta información. En este contexto, la organización debería identificar las regulaciones referentes a la protección de datos de su localidad, según la sección 4.1 a) y hacerlas parte de sus requisitos en la adquisición de sus herramientas de IA, esto es especialmente importante cuando se trata de sistemas de aprendizaje continuo o aquellos basados en servidores remotos. La organización que utiliza la IA podría evaluar los riesgos de seguridad de la información e implementar controles basados en procesos transparentes para el manejo de los datos de sus pacientes, utilizando el anexo A de la norma ISO/IEC 27001 que detalla medidas como: establecer políticas de seguridad, roles y responsabilidades en el manejo de la información, clasificar la información, implementar reglas en la transferencia de la información, proteger los registros

	contra pérdida, destrucción, falsificación, acceso no autorizado y publicación no autorizada, además, utilizar antivirus, firewalls o sistemas de protección de intrusos y considerar elaborar copias de seguridad de la información importante. Es crucial mantener códigos de ética en la cultura organizacional, según el anexo A.2.2, y asegurarse y documentar, según el anexo A.3.2 que el personal que maneja esta información sensible es competente y consciente de sus implicaciones éticas y legales y documentarlo según el literal A.4.6
Evitar la erosión de la relación paciente-médico y prevenir la dependencia excesiva del profesional o paciente hacia los sistemas de IA/Diseño y desarrollo	El desarrollador podría evaluar la afectación potencial de sus sistemas hacia individuos o grupos de individuos, según el anexo A.5.2 y añadir en sus objetivos éticos de diseño la colaboración del juicio humano en lugar de su sustitución, según el anexo A.6.1.2, por medio de interfaces y mecanismos de auditoría que permita a los médicos la revisión, edición o anulación de decisiones tomadas por el sistema, para esto es necesario que el modelo cumpla con características de explicabilidad y transparencia. Las aplicaciones diseñadas para el paciente podrían presentar resultados en forma de sugerencias que inviten al usuario a consultar con el médico y no en forma de diagnósticos concluyentes, en este sentido, el desarrollador debería explicar al usuario claramente el alcance y limitaciones del sistema de IA, según el anexo A.8.2, con el fin de evitar la participación exclusiva de estos sistemas en su cuidado. Es necesario según el anexo A.5.4, que el desarrollador implemente mecanismos de control como entrevistas o encuestas y mantenga referencia de la percepción de sus clientes sobre el papel de los modelos de IA en la interacción paciente-médico, y según los resultados realizar las mejoras necesarias para alcanzar los objetivos éticos propuestos.
Evitar la erosión de la relación paciente-médico y prevenir la dependencia excesiva del profesional o paciente hacia los sistemas de IA/Implementación y operación	El usuario podría incluir entre sus requisitos de compra, según el anexo A.10.3, características de transparencia, explicabilidad y la existencia de puntos de control y supervisión en los resultados de los sistemas de IA. En este sentido, el usuario debería incluir en sus procesos, según el anexo A.6.2.6, la supervisión de los resultados emitidos por los sistemas de IA antes de ser entregados al paciente, para esto es necesario asumir la responsabilidad correspondiente, según el anexo A.3.2, y desarrollar competencias, según el anexo A.4.6, que permita a los médicos comprender las decisiones tomadas por estas herramientas, evaluarlas y explicarlas a sus pacientes. En este contexto, la organización podría fomentar la capacitación constante y actualización de conocimientos por parte del personal médico con el fin de fomentar su pensamiento crítico y que su rol se mantenga relevante en la práctica clínica. La organización, de acuerdo con sus objetivos de uso responsable del anexo A.9.3, también podría, según la sección 9.1, implementar indicadores como: incidencia de errores no detectados, porcentaje de decisiones no justificadas, porcentaje de decisiones corregidas por intervención humana, tiempo promedio para toma de decisiones, etc., con el objetivo de asegurar la intervención constante del profesional en el proceso de decisiones.
Prevenir la insuficiente validación clínica en los sistemas de IA y asegurar la precisión y reproductibilidad de los resultados/diseño y desarrollo	El desarrollador podría establecer políticas, según el anexo A.2.2, para el desarrollo de sistemas de IA que incorporen criterios de explicabilidad sin sacrificar su precisión, esto podría lograrse por medio de sistemas híbridos que incorporen modelos simples e interpretables junto con modelos complejos(Vaira et al. 2024, 12) Además, la organización podría implementar objetivos de precisión y reproductibilidad claros, medibles y monitoreables, según la sección 9.1, en todas las etapas del desarrollo de sus sistemas de IA . En la etapa de selección de datos para el entrenamiento y validación de los modelos de IA, según los anexos A.7.2-A.7.5, la organización podría implementar controles para garantizar su calidad, procedencia, veracidad y relevancia, asimismo, es fundamental que la organización garantice adecuados procesos de estandarización, según el anexo A.7.6, debido a que generalmente estos datos provienen de diversas fuentes y en consecuencia se encuentran en distintos formatos. También es importante que la organización implemente métodos de validación de las herramientas a utilizarse en el desarrollo de sus modelos, según el anexo A.4.4, desde las que se utilizan para el acondicionamiento y preparación de datos hasta las que se utilizan para optimización y evaluación del modelo.

	Los criterios de validación en la precisión del modelo, por ejemplo, F1 score en matrices de confusión. Podrían orientarse, según el anexo A.10.4 y A.6.2.4, hacia los segmentos poblacionales a los que este va dirigido, El desarrollador podría también implementar, según el anexo A.6.2.6, funcionalidades de monitoreo en todo el ciclo de vida de su modelo de IA y establecer, según el anexo A.8.3, canales de comunicación que permitan al cliente reportar problemas de precisión identificados durante su uso. En este contexto, el modelo debería ser actualizado periódicamente tomando en cuenta la retroalimentación de los clientes y datos con información actualizada con el objetivo de mantener su relevancia, reproductibilidad y precisión en el tiempo.
Prevenir la insuficiente validación clínica en los sistemas de IA y asegurar la precisión y reproductibilidad de los resultados/ implementación y operación	El usuario podría, según el anexo A.10.3, definir criterios de precisión orientados a sus grupos poblacionales específicos en los requisitos de compra de las herramientas de IA, por ejemplo, sensibilidad y especificidad mayores a 95 %. El usuario también debería, según el anexo A.8.2, conocer los posibles riesgos y limitaciones del sistema para ajustarlos a sus expectativas y tomar decisiones responsables También podría utilizar, según el anexo A.6.2.4, herramientas como QAMAI (Vaira et al. 2024, 1) para validar los sistemas de IA de la organización en cuestiones como precisión, claridad, relevancia, fuentes y utilidad. Es importante que la organización que utiliza el modelo, según el anexo A.4.6, cuente con el personal adecuado y capacitado para garantizar la precisión de los resultados. Parte de la responsabilidad del desempeño de las herramientas de IA , según los anexos A.8.3 y A.8.5, está en la retroalimentación que el cliente realice sobre impactos adversos o comportamientos no deseados que permitan al desarrollador gestionarlos, para esto, según el anexo A.6.2.6, es importante establecer mecanismos de monitoreo continuo de los resultados de estas herramientas, como pueden ser: tasa de aciertos= predicciones correctas/total de predicciones, precisión=verdaderos positivos/total de positivos, etc.

Fuente: Ganatra et al.; Schuwart et al.; Shwartz et al.; Vaira et al.; ISO/IEC 27001:2022; ISO/IEC 23894:2023; ISO/IEC 42001:2023; ISO 9241-210:2010.
Elaboración propia.

Conclusiones

Esta investigación identificó en las herramientas de IA orientadas al diagnóstico automático, diagnóstico por imagen, salud mental y desarrollo de medicamentos, los siguientes desafíos éticos: reducir el sesgo asegurar la equidad, dotar de transparencia y explicabilidad a los sistemas de IA, evitar la mercantilización de la salud, evitar la obsolescencia y deshumanización, asegurar la privacidad de la información. Además, evitar la erosión de la relación paciente-médico, prevenir la dependencia excesiva del profesional o paciente hacia los sistemas de IA, prevenir la insuficiente validación clínica en los sistemas de IA y asegurar la precisión y reproductibilidad de los resultados. Estos desafíos se relacionan con los principios éticos definidos por la OMS para el uso de la IA en el área de la salud. En cada uno de estos desafíos se propusieron recomendaciones para su gestión, basadas en la norma ISO/IEC 42001:2023, principalmente en el anexo A y otras normativas de soporte como la ISO/IEC 23894:2023 y la ISO/IEC 27001:2022

El Anexo A constituye el núcleo operativo de la norma ISO/IEC 42001:2023, ya que detalla los controles que posibilitan la gestión efectiva de los desafíos éticos asociados a la inteligencia artificial. Entre sus principales directrices destaca la necesidad de implementar políticas sólidas que respalden la gestión, la cual exige un compromiso decidido por parte de la alta dirección con los principios éticos de la IA. Este compromiso es fundamental debido a que las regulaciones en esta área aún son incipientes a nivel global. Por lo tanto, los marcos éticos se convierten muchas veces en la única guía para las organizaciones. El liderazgo de la alta dirección debe ir acompañado de la provisión de los recursos necesarios, que incluyen: datos confiables, herramientas tecnológicas y personal capacitado para asegurar la eficacia y sostenibilidad del sistema de gestión.

En el contexto organizativo, resulta fundamental delimitar el alcance de los sistemas de IA, un aspecto que a menudo es subestimado por los desarrolladores de soluciones para el sector salud. Además, la tendencia a la mercantilización excesiva conlleva no solo al solucionismo tecnológico sino también a la oferta indiscriminada de modelos de IA a nivel global sin la adecuada especificación de las poblaciones objetivo para las que dichos modelos fueron entrenados. Esto favorece la perpetuación de desafíos éticos como el sesgo, la

disminución de la precisión y la falta de reproductibilidad de los resultados. En este contexto, es también necesario reconocer la corresponsabilidad del consumidor, quien además de definir su propio público objetivo, debería exigir a los proveedores modelos validados en esos grupos específicos para garantizar resultados confiables e implementar métricas que permitan evaluar el desempeño de sus herramientas de IA con el objetivo de reportar al desarrollador cualquier fallo, comportamiento no deseado o desviación detectada, contribuyendo así a la mejora continua del sistema.

Por otro lado, la aparición de nuevas tecnologías suele venir de la mano con el temor a la pérdida de empleos y la deshumanización, evitarlo requiere más allá de las leyes, un compromiso entre desarrolladores y usuarios. Los desarrolladores pueden privilegiar la creación de modelos explicables e implementar puntos de control humano, mientras que los usuarios deberían mantenerse relevantes en el proceso de interacción con la IA, utilizándola como una herramienta y no delegando en ella la toma de decisiones finales.

Se sostiene que una de las principales limitaciones de los sistemas de IA radica en su incapacidad para manifestar empatía. Sin embargo, cada vez son más las personas que prefieren interactuar con chatbots en lugar de seres humanos, debido entre otras cosas, a que los primeros no emiten juicios de valor. Por lo tanto, el profesional de la salud debería actualizar sus conocimientos en forma constante y no solo interactuar más con sus pacientes sino mejorar su forma de interacción con ellos.

En consecuencia, es preciso mencionar que la confiabilidad de la IA requiere un trabajo y colaboración conjunta entre desarrolladores, consumidores, entes reguladores y académicos, que garantice los derechos de las sociedades y la sostenibilidad de la interacción humano-máquina.

Además, es fundamental destacar la oportunidad que tienen tanto el país como la región en este contexto crítico para generar las condiciones adecuadas que permitan el desarrollo y uso ético de aplicaciones de IA. Para ello, resulta clave fortalecer las bases de datos públicas y establecer normativas que, por ejemplo, promuevan el entrenamiento de sistemas de IA en áreas como la salud utilizando datos propios, con el objetivo de prevenir sesgos y, en consecuencia, garantizar la precisión y confiabilidad de los resultados. Asimismo, esta coyuntura puede ser propicia para incentivar la apertura y la cooperación regional, especialmente con países que comparten raíces étnicas similares, a fin de

implementar reglamentos y directrices comunes que faciliten el intercambio de bases de datos debidamente normalizadas. Esto permitirá desarrollar herramientas de IA confiables y robustas que contribuyan al avance y a la mejora de la salud en la región. En este sentido, cobra especial relevancia la adopción de la norma ISO/IEC 42001:2023 por parte de las organizaciones, ya que proporciona un marco ético que permite estandarizar la gestión a nivel regional y también exige el cumplimiento de los requisitos legales y reglamentarios de cada país. Aprovechar este escenario será clave para posicionar al país y a la región como referentes en el desarrollo responsable y eficaz de tecnologías de IA.

Futuras investigaciones podrían analizar o medir el impacto que tiene la implementación del sistema de gestión ISO/IEC 42001:2023 sobre los desafíos éticos asociados al desarrollo y uso de herramientas de inteligencia artificial en el ámbito de la salud. Sería especialmente relevante explorar la correlación entre el nivel de cumplimiento ético y el crecimiento organizacional, así como proponer y validar indicadores específicos para el monitoreo efectivo del sesgo, privacidad, explicabilidad, precisión y otros aspectos esenciales para una gestión ética y responsable de la IA.

Obras citadas

Ahmad, Istiak, y Fahad Alqurashi. 2024. “Early Cancer Detection Using Deep Learning and Medical Imaging: A Survey”. *Critical Reviews in Oncology/Hematology* 204 (diciembre): 104528. doi:10.1016/j.critrevonc.2024.104528.

Aidoc. 2019. “Next Gen Radiology Ai the Journey from an Algorithm to a Clinical Solution”.

Álvarez-Maldonado, David, Carmen Pénnanen-Arias, Nicolás Barrientos Oradini, y Ximena Vega Donoso. 2024a. “Inteligencia Artificial y Emprendimiento: Una revisión sistemática desde un enfoque contextual”. *Journal of the Academy*, nº 11 (octubre): 33–52. doi:10.47058/joa11.3.

———. 2024b. “Inteligencia Artificial y Emprendimiento: Una revisión sistemática desde un enfoque contextual”. *Journal of the Academy*, nº 11 (octubre): 33–52. doi:10.47058/joa11.3.

Aravind Ayyagiri, Anshika Aggarwal, y Shalu Jain. 2024. “Enhancing DNA Sequencing Workflow with AI-Driven Analytics”. *International Journal for Research Publication and Seminar* 15 (3): 203–16. doi:10.36676/jrps.v15.i3.1484.

Asamblea Nacional del Ecuador. 2021. Ley orgánica de protección de datos personales.Registro Oficial Suplemento 459.

Azabache Santos, José Diego, Nelson Alejandro Ángeles Piedra, y Alberto Carlos Mendoza De Los Santos. 2024. “Impacto de la integración del gobierno de ti en la adopción de la inteligencia artificial”. *Investigacion & desarrollo* 23 (2). doi:10.23881/idupbo.023.2-9e.

Badaro, Sebastian, Leonardo Javier Ibañez, y Martín Agüero. 2013. “Sistemas expertos: fundamentos, metodologías Y aplicaciones”. *Ciencia y Tecnología* 1 (13). doi:10.18682/cyt.v1i13.122.

BaHammam, Ahmed S, y Michael Wl Chee. 2022. “Publicly Available Health Research Datasets: Opportunities and Responsibilities”. *Nature and Science of Sleep* Volume 14 (septiembre): 1709–12. doi:10.2147/NSS.S390292.

Bekdaş, Gebrail, Sinan Melih Nigdeli, y Melda Yücel, eds. 2020. *Artificial Intelligence and Machine Learning Applications in Civil, Mechanical, and Industrial Engineering*. Advances in Computational Intelligence and Robotics. IGI Global. doi:10.4018/978-1-7998-0301-0.

Benítez Vargas, Marlon Antonio, Laura Rico Duarte, y Fanny Patricia Niño Hernández. 2023. “El desafío que representan las obras creadas por inteligencia artificial al derecho de autor en Colombia”. Editado por Universitat Oberta de Catalunya. *IDP. Revista de Internet, Derecho y Política* 0 (38). doi:10.7238/idp.v0i38.403977.

Benraouane, Sid Ahmed. 2024. *AI Management System Certification According to the ISO/IEC 42001 Standard: How to Audit, Certify, and Build Responsible AI Systems*. 1ª ed. New York: Productivity Press. doi:10.4324/9781003463979.

Bergan, Marie Burns, Marthe Larsen, Nataliia Moshina, Hauke Bartsch, Henrik Wethe Koch, Hildegunn Siv Aase, Zhanbolat Satybaldinov, Ingfrid Helene Salvesen Haldorsen, Christoph I. Lee, y Solveig Hofvind. 2024. “AI Performance by Mammographic Density in a Retrospective Cohort Study of 99,489 Participants in BreastScreen

Norway". *European Radiology* 34 (10): 6298–6308. doi:10.1007/s00330-024-10681-z.

Bogucka, Edyta, Marios Constantinides, Sanja Šćepanović, y Daniele Quercia. 2024. "Co-Designing an AI Impact Assessment Report Template with AI Practitioners and AI Compliance Experts". arXiv. doi:10.48550/arXiv.2407.17374.

Bonacina, Maria Paola. 2017. "Automated Reasoning for Explainable Artificial Intelligence". En . doi:10.29007/4b7h.

Borna, Sahar, Cesar A. Gomez-Cabello, Sophia M. Pressman, Syed Ali Haider, Ajai Sehgal, Bradley C. Leibovich, Dave Cole, y Antonio Jorge Forte. 2024. "Comparative Analysis of Artificial Intelligence Virtual Assistant and Large Language Models in Post-Operative Care". *European Journal of Investigation in Health, Psychology and Education* 14 (5): 1413–24. doi:10.3390/ejihpe14050093.

Brady, Patrick. 2024. "Artificial Intelligence in Drug Development: Navigating Emerging Regulatory Expectations and Enhancing Clinical Trials". Iqvia.

Bürger, Valerie K., Julia Amann, Cathrine K. T. Bui, Jana Fehr, y Vince I. Madai. 2024. "The Unmet Promise of Trustworthy AI in Healthcare: Why We Fail at Clinical Translation". *Frontiers in Digital Health* 6. doi:10.3389/fdgth.2024.1279629.

Burr, Christopher, Sophie Arana, Cassandra Gould Van Praag, Ibrahim Habli, Marten Kaas, Michael Katell, Shakir Laher, et al. 2024. *Trustworthy and Ethical Assurance of Digital Health and Healthcare: Supporting an Assurance Ecosystem for Data-Driven Technologies in Health and Healthcare*. Alan Turing Institute. doi:10.5281/ZENODO.10532573.

Bussines & Decision. 2023. "Livre-blanc la generative futur numerique". www.businessdecision.com.

CCNE, y CNPEN. 2022. "Medical Diagnosis and Artificial Intelligence: Ethical Issues".

CENIA. 2024. "Informe ILIA 2024". https://www.nodal.am/wp-content/uploads/2024/09/ILIA_20242.pdf.

CEPAL. 2024. "Superar las trampas del desarrollo de América Latina y el Caribe en la era digital: el potencial transformador de las tecnologías digitales y la inteligencia artificial". *LC/CMSI.9/3*.

Cheang Wong, Juan Carlos. 2005. "Ley de Moore, nanotecnología y nanociencias: síntesis y modificación de nanopartículas mediante la implantación de iones". *Revista Digital Universitaria* 6 (7).

Chen, Wei, Xuesong Liu, Sanyin Zhang, y Shilin Chen. 2023. "Artificial Intelligence for Drug Discovery: Resources, Methods, and Applications". *Molecular Therapy - Nucleic Acids* 31 (marzo): 691–702. doi:10.1016/j.omtn.2023.02.019.

"Cifras de Ecuador – Cáncer de Mama – Ministerio de Salud Pública". 2018. https://www.salud.gob.ec/cifras-de-ecuador-cancer-de-mama/.

Contreras, Pablo. 2024. "Convergencia internacional y caminos propios: regulación de la inteligencia artificial en América Latina". En *Actualidad Jurídica Iberoamericana*, 468–93. 21.

"Critical Care Suite 2.0". 2025. Accedido marzo 21. https://www.gehealthcare.com/products/radiography/critical-care-suite?srsltid=AfmBOoqzLLmET1pLRkHBm8_dGcEbYyzO2-7b21XD7UA-TMlNpXTlFk3j.

Dai, Ling, Mingming Zhou, y Haipeng Liu. 2023. "Recent Applications of Convolutional Neural Networks in Medical Data Analysis": En *Advances in Healthcare Information*

Systems and Administration, editado por Ahdi Hassan, Vivek Kumar Prasad, Pronaya Bhattacharya, Pushan Dutta, y Robertas Damaševičius, 119–31. IGI Global. doi:10.4018/979-8-3693-1082-3.ch007.

Dave, Daksh, Adnan Akhunzada, Nikola Ivković, Sujan Gyawali, Korhan Cengiz, Adeel Ahmed, y Ahmad Sami Al-Shamayleh. 2025. "Diagnostic Test Accuracy of AI-Assisted Mammography for Breast Imaging: A Narrative Review". *PeerJ Computer Science* 11 (febrero): e2476. doi:10.7717/peerj-cs.2476.

Deloitte & Touche LLP. 2023. "Demonstrate Market Readiness of Your AI Systems with ISO 42001". https://www2.deloitte.com/us/en/pages/financial-advisory/articles/iso-42001-standard-ai-governance-risk-management.html.

Duran, Cals, Caralt Conesa, y Viladrosa Clarisó. 2013. *Introducción a la representación del conocimiento*. Catalunya: Fundación para la Universitat Oberta de Catalunya.

Eldin, Waleed Salah, y Ahmed Kaboudan. 2023. "AI-Driven Medical Imaging Platform: Advancements in Image Analysis and Healthcare Diagnosis". *Journal of the ACS* 14 (junio).

Elhaik, Eran. 2022. "OPEN Principal Component Analyses". *Scientific Reports* 12. doi:https://doi.org/10.1038/s41598-022-14395-4.

Escalante González, Melissa. 2023. "Aplicación de la inteligencia artificial para la detección del cáncer de mama". *Revista Médica Sinergia* 8 (12): e1113. doi:10.31434/rms.v8i12.1113.

Esteva, Andre, Alexandre Robicquet, Bharath Ramsundar, Volodymyr Kuleshov, Mark DePristo, Katherine Chou, Claire Cui, Greg Corrado, Sebastian Thrun, y Jeff Dean. 2019. "A Guide to Deep Learning in Healthcare". *Nature Medicine* 25 (1): 24–29. doi:10.1038/s41591-018-0316-z.

Ezeana, Chika F., Tiancheng He, Tejal A. Patel, Virginia Kaklamani, Maryam Elmi, Erika Brigmon, Pamela M. Otto, et al. 2023. "A Deep Learning Decision Support Tool to Improve Risk Stratification and Reduce Unnecessary Biopsies in BI-RADS 4 Mammograms". *Radiology: Artificial Intelligence* 5 (6). doi:10.1148/ryai.220259.

Fan, Xiaofei, Xiaoming Qiao, Zhisheng Wang, Luetao Jiang, Yue Liu, y Qingshan Sun. 2022. "Artificial Intelligence-Based CT Imaging on Diagnosis of Patients with Lumbar Disc Herniation by Scalpel Treatment". Editado por Arpit Bhardwaj. *Computational Intelligence and Neuroscience* 2022 (mayo): 1–8. doi:10.1155/2022/3688630.

Farhud, Dariush D., y Shaghayegh Zokaei. 2021. "Ethical Issues of Artificial Intelligence in Medicine and Healthcare". *Iranian Journal of Public Health* 50 (11). doi:10.18502/ijph.v50i11.7600.

FCD. 2025. "Ecuador empeora su calificación en el Índice de Percepción de Corrupción Ciudadanía y Desarrollo". febrero 11. https://www.ciudadaniaydesarrollo.org/2025/02/11/ecuador-empeora-su-calificacion-en-el-indice-de-percepcion-de-corrupcion/.

Fernández, Sherwin, Rutvij Gawas, Preston Alvares, y Macklon Fernandes. 2020. "Doctor Chatbot:Heart Disease Prediction System". *ITEE Journal* 9 (5).

Filgueira, Fernando. 2023. "Desafíos de gobernanza de inteligencia artificial en América Latina. Infraestructura, descolonización y nueva dependencia". *Revista del CLAD Reforma y Democracia*, nº 87 (diciembre): 44–70. doi:10.69733/clad.ryd.n87.a3.

Flasiński, Mariusz. 2016. *Introduction to Artificial Intelligence*. Cham: Springer International Publishing. doi:10.1007/978-3-319-40022-8.

Foti, Giovanni, y Chiara Longo. 2024. “Deep Learning and Ai in Reducing Magnetic Resonance Imaging Scanning Time: Advantages and Pitfalls in Clinical Practice”. *Polish Journal of Radiology* 89 (septiembre): 443–51. doi:10.5114/pjr/192822.

Ganatra, Amit, Brijeshkumar Y. Panchal, Devarshi Doshi, Devanshi Bhatt, Jesal Desai, Bijal Talati, Neha Soni, y Apurva Shah. 2024. “Introduction to Explainable AI”. En *Explainable AI in Health Informatics*, editado por Rajanikanth Aluvalu, Mayuri Mehta, y Patrick Siarry, 1–31. Computational Intelligence Methods and Applications. Singapore: Springer Nature Singapore. doi:10.1007/978-981-97-3705-5_1.

Garcia, Juan Camilo Ramirez, Julieth Nathalia Martinez Carrillo, y Diego Fernando Santanilla Cuellar. 2023. “Inteligencia artificial: ventajas y beneficios en los negocios”. *Aglala* 14 (2).

García, Miguel. 2018. *Fundamentos de Inteligencia Artificial*. Bogotá: Universidad Manuela Beltrán.

Garzón Espitia, Paola Andrea, y Ana María Rodríguez Padilla. 2022. “Análisis sobre marcos regulatorios internacionales sobre en la evolución de la inteligencia artificial (2008-2018)”. *Punto de vista* 13 (20): 127–44. doi:10.15765/pdv.v13i20.3459.

Geis, J. Raymond, Adrian P. Brady, Carol C. Wu, Jack Spencer, Erik Ranschaert, Jacob L. Jaremko, Steve G. Langer, Andrea Borondy Kitts, et al. 2019. “Ethics of Artificial Intelligence in Radiology: Summary of the Joint European and North American Multisociety Statement”. *Radiology* 293 (2): 436–40. doi:10.1148/radiol.2019191586.

Geis, J. Raymond, Adrian Brady, Carol C. Wu, Jack Spencer, Erik Ranschaert, Jacob L. Jaremko, Steve G. Langer, Andrea Borondy Kitts, et al. 2019. “Ethics of Artificial Intelligence in Radiology: Summary of the Joint European and North American Multisociety Statement”. *Insights into Imaging* 10 (1): 101. doi:10.1186/s13244-019-0785-8.

Giovanola, Benedetta, y Simona Tiribelli. 2023. “Beyond Bias and Discrimination: Redefining the AI Ethics Principle of Fairness in Healthcare Machine-Learning Algorithms”. *AI & SOCIETY* 38 (2): 549–63. doi:10.1007/s00146-022-01455-6.

Gjesvik, Jonas, Nataliia Moshina, Christoph I. Lee, Diana L. Miglioretti, y Solveig Hofvind. 2024. “Artificial Intelligence Algorithm for Subclinical Breast Cancer Detection”. *JAMA Network Open* 7 (10): e2437402. doi:10.1001/jamanetworkopen.2024.37402.

Gnanasambandam, Anjali. 2023. “Artificial Intelligence in Clinical Trials- Future Prospectives”. *Bioequivalence & Bioavailability International Journal* 7 (1): 1–8. doi:10.23880/beba-16000196.

Goktas, Polat, y Andrzej Grzybowski. 2025. “Shaping the Future of Healthcare: Ethical Clinical Challenges and Pathways to Trustworthy AI”. *Journal of Clinical Medicine* 14 (5). MDPI AG: 1605. doi:10.3390/jcm14051605.

González, Rodrigo. 2007. “El test de Turing: dos mitos, un dogma”. *Revista de filosofía* 63. doi:10.4067/S0718-43602007000100003.

Grover, Veena, y Mahima Dogra. 2024. “Challenges and Limitations of Explainable AI in Healthcare”: En *Advances in Healthcare Information Systems and Administration*, editado por Veena Grover, Balamurugan Balusamy, Nallakaruppan M.K., Vijay Anand, y Mariofanna Milanova, 72–85. IGI Global. doi:10.4018/979-8-3693-5468-1.ch005.

Gutiérrez Proenza, Janetsy. 2022. “datos personales en el Ecuador como un derecho humano una necesidad de mejoramiento en su regulación”. *Revista Jurídica Crítica y Derecho* 3 (5): 53–66. doi:10.29166/cyd.v3i5.3950.

Guzmán Luna, Jaime Albero. 2003. “Técnicas de planificación para la generación automática de programas de control”. *Ciencia Investigación Academia Desarrollo* 8 (2): 25–31.

Heredia Peñaloza, Luis Fernando, y María Auxiliadora Santacruz Vélez. 2023. “Responsabilidad médica administrativa, Manejo de datos personales relativos a la salud, Ecuador, 2023”. *Revista Religación* 8 (38): e2301102. doi:10.46652/rgn.v8i38.1102.

Hernandez-Trinidad, Aron, Blanca Olivia Murillo-Ortiz, Rafael Guzman-Cabrera, y Teodoro Cordova-Fraga. 2024. “Applications of Artificial Intelligence in the Classification of Magnetic Resonance Images: Advances and Perspectives”. En *New Advances in Magnetic Resonance Imaging*, editado por Denis Larrivee. IntechOpen. doi:10.5772/intechopen.113826.

Hsieh, Jiang, Eugene Liu, Brian Nett, Jie Tang, Jean-Baptiste Thibault, y Sonia Sahney. 2019. “A New Era of Image Reconstruction: TrueFidelity™: Technical White Paper on Deep Learning Image Reconstruction”.

Hutson, Matthew. 2024. “How AI Is Being Used to Accelerate Clinical Trials”. *Nature* 627 (8003): S2–5. doi:10.1038/d41586-024-00753-x.

Iglesia, Estefania Dominguez de la. 2024. “Zebra Medical Vision”. *Campus Health Tech*. junio 6. https://campushealthtech.com/blog/zebra-medical-vision/.

INEC. 2024. “Títulos Nacionales y Extranjeros”. *Servicios Senescyt*. https://siau.senescyt.gob.ec/titulados-nacionales-extranjeros/.

ISO. 2010. “ISO 9241-210 Ergonomics of Human–System Interaction Part 210: Human-Centred Design for Interactive Systems”. Geneva.

ISO/IEC. 2022a. “ISO/IEC 27001:2022 Seguridad de la información, ciberseguridad y protección de la privacidad-Sistemas de gestión de la seguridad de la información Requisitos”. Vernier,Geneva.

———. 2022b. “ISO/IEC 23053:2022 Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML).” Geneva.

———. 2022c. “ISO/IEC 22989:2022 Information Technology-Artificial Intelligence-Artificial Intelligence Concepts and Terminology”. Geneva.

———. 2023. “BS-ISO-IEC-42001-2023 Information Technology Artificial Intelligence Management System”. London.

———. 2024. “BS EN ISO/IEC 23894:2024 Information Technology - Artificial Intelligence - Guidance on Risk Management”. London.

Jagodič, Gregor, y Miloš Šinkovec. 2021. “Involvement of Artificial Intelligence in Modern Society”. *International Journal of Management, Knowledge and Learning* 10: 267–73. doi:10.53615/2232-5697.10.267-273.

Jauregui Moran. 2024. “Democratizando el Acceso a la Inteligencia Artificial”. Unpublished. doi:10.13140/RG.2.2.28804.74887.

Jiménez Schlegl, Pablo, y Carme Torras Genís. 2020. “Inteligencia artificial y robótica”. Universitat Oberta de Catalunya.

Jin, Dakai, Adam P. Harrison, Ling Zhang, Ke Yan, Yirui Wang, Jinzheng Cai, Shun Miao, y Le Lu. 2021. “Artificial Intelligence in Radiology”. En *Artificial Intelligence in Medicine*, 265–89. Elsevier. doi:10.1016/B978-0-12-821259-2.00014-4.

Kappel, Rebecca. 2024. “ISO/IEC 42001: What You Need to Know”. *Centraleyes*. julio 29. https://www.centraleyes.com/iso-iec-42001/.

Khalifa, Mohamed, y Mona Albadawy. 2024. “AI in Diagnostic Imaging: Revolutionising Accuracy and Efficiency”. *Computer Methods and Programs in Biomedicine Update* 5: 100146. doi:10.1016/j.cmpbup.2024.100146.

Khurshid, Fatima, Ieraj Zia, Ume Anum Ayesha, Fatima Yaqoob, Hafsa Khurshid, Areeba Zia, y Ayesha Khurshid. 2024. “The Significance of Screening Mammography: A Preliminary Study”, 42–48. doi:https://doi.org/10.46663%2Fajsmu.v9i2.42-48.

Kraaijeveld, Steven R, Hanneke Van Heijster, Nadine Bol, y Kirsten E Bevelander. 2025. “The Ethics of Using Virtual Assistants to Help People in Vulnerable Positions Access Care”. *Journal of Medical Ethics*, febrero. doi:10.1136/jme-2024-110464.

Krishnan, Saravanan, Ramesh Kesavan, B. Surendiran, y G. S. Mahalakshmi, eds. 2021. *Handbook of Artificial Intelligence in Biomedical Engineering*. 1ª ed. Series statement: Biomedical engineering: techniques and applications: Apple Academic Press. doi:10.1201/9781003045564.

Kumar, Vijay, y Streepathy Ramesh. 2024. “AI-Driven Healthcare: Predictive Analytics for Disease Diagnosis and Treatment”. *International Journal for Modern Trends in Science and Technology* 10 (06): 5–9. doi:10.46501/IJMTST1006002.

Lamb, Leslie R., Constance D. Lehman, Aimilia Gastounioti, Emily F. Conant, y Manisha Bahl. 2022. “Artificial Intelligence (AI) for Screening Mammography, from the *AJR* Special Series on AI Applications”. *American Journal of Roentgenology* 219 (3): 369–80. doi:10.2214/AJR.21.27071.

Lee, Theresa, y Euclid Seeram. 2020. “The Use of Artificial Intelligence in Computed Tomography Image Reconstruction: A Systematic Review”. *Radiology – Open Journal* 4 (2): 30–38. doi:10.17140/ROJ-4-129.

Lim, Chee-Peng, Ashlesha Vaidya, Yen-Wei Chen, Tejasvi Jain, y Lakhmi C. Jain, eds. 2023. *Artificial Intelligence and Machine Learning for Healthcare: Vol. 1: Image and Data Analytics*. Vol. 228. Intelligent Systems Reference Library. Cham: Springer International Publishing. doi:10.1007/978-3-031-11154-9.

Loján Alvarado, Harold Patricio, y Oscar Efrén Cárdenas Villavicencio. 2024. “Regulación del Manejo de la Inteligencia Artificial, Consecuencias y Daños a la Sociedad por su Mal Uso”. *Ciencia Latina Revista Científica Multidisciplinar* 8 (1): 1966–78. doi:10.37811/cl_rcm.v8i1.9596.

Longo, Luca, y Ruairi O’Reilly, eds. 2023. *Artificial Intelligence and Cognitive Science: 30th Irish Conference, AICS 2022, Munster, Ireland, December 8–9, 2022, Revised Selected Papers*. Vol. 1662. Communications in Computer and Information Science. Cham: Springer Nature Switzerland. doi:10.1007/978-3-031-26438-2.

Lunit. 2025. “Lunit INSIGHT MMG”. *[Lunit] Lunit Insight Mmg, an Ai Solution for Mammography*. Accedido marzo 30. https://www.lunit.io/en/products/mmg.

Maldonado, Fidel. 2024. “Retos en la Seguridad de la Información”. *SIC*.

Manmeet Kaur. 2024. “A Comprehensive Overview of Artificial Intelligence-Based Classification Techniques”. *International Journal of Science and Research Archive* 11 (2): 125–29. doi:10.30574/ijsra.2024.11.2.0387.

Maria, Ablameyko. 2023. “AI Image-Based Diagnosis Systems: How to Implement Them?” *Journal of Digital Health*, junio, 12–21. doi:10.55976/jdh.22023113912-21.

Marketing Departamento. 2021. “Historia de la inteligencia artificial: origen y auge de la IA”. *www.elternativa.com*. noviembre 23. https://www.elternativa.com/historia-inteligencia-artificial/.

Márquez, Francisco. 2025. “La batalla por la supremacía tecnológica: EE. UU. vs. China”. *ieee.es*, 23/2025, .

Marr, Bernard. 2024. “Lo que todo CEO necesita saber sobre la nueva Ley de IA de la Unión Europea”. *Forbes Ecuador*. abril 1. https://www.forbes.com.ec/innovacion/lo-todo-ceo-necesita-saber-sobre-nueva-ley-ia-union-europea-n50160.

Matheny, Sonoo Thadaney Israni, Mahnoor Ahmed, y Danielle Whicher. 2019. *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril*. Washinton D.C,.

McDonald, James. 2024. “Introduction to Artificial Intelligence (AI) Technology Guide for Travel & Tourism Leaders”. World Travel and Tourism Council. https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/final/en-us/microsoft-brand/documents/2024-wttc-introduction-to-ai.pdf.

MINTEL. 2021. “Proyecto: Diagnostico sobre la inteligencia artificial en el Ecuador”.

———. 2025. *Política Pública para la Transformación Digital del Ecuador 2024-2030 (con base en el Acuerdo Ministerial Nro. MINTEL-MINTEL-2025-0005)*.

Mittermaier, Mirja, Marium M. Raza, y Joseph C. Kvedar. 2023. “Bias in AI-Based Models for Medical Applications: Challenges and Mitigation Strategies”. *Npj Digital Medicine* 6 (1): 113, s41746-023-00858-z. doi:10.1038/s41746-023-00858-z.

Mohamad, Nor Rafidah, Norlia Yusof, y Wan Hussain Wan Ishak. 2004. “Expert System in Supporting Business: The Challenge and Future Prospect”, 279–86.

Moor, James. 2006. “The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years”. *AI Magazine* 27 (4).

Murdoch, Blake. 2021. “Privacy and Artificial Intelligence: Challenges for Protecting Health Information in a New Era”. *BMC Medical Ethics* 22 (1): 122. doi:10.1186/s12910-021-00687-3.

Musalamadugu, Tanmai Sree, y Hemachandran Kannan. 2023. “Generative AI for Medical Imaging Analysis and Applications”. *Future Medicine AI*, septiembre, FMAI5. doi:10.2217/fmai-2023-0004.

Nafissi, Nahid. 2024. “The Application of Artificial Intelligence in Breast Cancer”. *Eurasian Journal of Medicine and Oncology*, 235–44. doi:10.14744/ejmo.2024.45903.

Nathaniel Kumar Sarella, Prakash, y Vinny Therissa Mangam. 2024. “AI-Driven Natural Language Processing in Healthcare: Transforming Patient-Provider Communication”. *Indian Journal of Pharmacy Practice* 17 (1): 21–26. doi:10.5530/ijopp.17.1.4.

Nawaz, Ali, Shehroz S. Khan, y Amir Ahmad. 2024. “Ensemble of Autoencoders for Anomaly Detection in Biomedical Data: A Narrative Review”. *IEEE Access* 12: 17273–89. doi:10.1109/ACCESS.2024.3360691.

Nepal, Subigya, Gonzalo J. Martinez, Arvind Pillai, Koustuv Saha, Shayan Mirjafari, Vedant Das Swain, Xuhai Xu, et al. 2024. “A Survey of Passive Sensing in the Workplace”. arXiv. doi:10.48550/arXiv.2201.03074.

Olawade, David B., Ojima J. Wada, Aanuoluwapo Clement David-Olawade, Edward Kunonga, Olawale Abaire, y Jonathan Ling. 2023. “Using Artificial Intelligence to Improve Public Health: A Narrative Review”. *Frontiers in Public Health* 11 (octubre): 1196397. doi:10.3389/fpubh.2023.1196397.

Olawade, David B., Ojima Z. Wada, Aderonke Odetayo, Aanuoluwapo Clement David-Olawade, Fiyinfoluwa Asaolu, y Judith Eberhardt. 2024. “Enhancing Mental Health with Artificial Intelligence: Current Trends and Future Prospects”. *Journal of Medicine, Surgery, and Public Health* 3 (agosto): 100099. doi:10.1016/j.glmedi.2024.100099.

OMS. 2024. “Cáncer de mama”. marzo 13. https://www.who.int/es/news-room/fact-sheets/detail/breast-cancer.

Oppenheimer, Andrés. 2014. *¡Crear o morir! la esperanza de América Latina y las cinco claves de la innovación*. 1. ed. Ensayo. México, D.F: Debate.

Peco, Pedro Pernías. 2024. “El Impacto de la Inteligencia Artificial Generativa en la Productividad Empresarial”. *Dpto. de lenguajes y sistemas informáticos. Universidad de alicante*, junio. Unpublished. doi:10.13140/RG.2.2.15413.49126.

Persons, Timoty M, y Michael MCGinnis. 2019. *Benefits and Challenges of Machine Learning in Drug Development*. GAO-20-2155P. GAO and National Academy of Medicine.

Piotrowski, Tomasz, Joanna Kazmierska, Mirosława Mocydlarz-Adamcewicz, y Adam Ryczkowski. 2021. “Ethical Issues on Artificial Intelligence in Radiology: How Is It Reported in Research Articles? The Current State and Future Directions”. *Journal of Medical Science* 90 (2): e513. doi:10.20883/medical.e513.

PNUD. 2025. “AILA: Evaluación del Panorama de la Inteligencia Artificial en Ecuador”.

Potočnik, Jaka, Shane Foley, y Edel Thomas. 2023. “Current and Potential Applications of Artificial Intelligence in Medical Imaging Practice: A Narrative Review”. *Journal of Medical Imaging and Radiation Sciences* 54 (2): 376–85. doi:10.1016/j.jmir.2023.03.033.

Pradhan, Devasis. 2023. “A Critical Review on Challenges and Limitations in Artificial Intelligence-Based e-Health Applications”. *Juniper* 5 (1). doi:10.19080/ETOAJ.2023.05.555654.

Prensa Asociada. 2019. “YouTube Fined $170m for Collecting Children’s Personal Data”. *The Guardian*, septiembre 4, sec. Technology. https://www.theguardian.com/technology/2019/sep/04/youtube-kids-fine-personal-data-collection-children-.

Protiviti. 2022. “ISO 27001:2022 – Key Changes and Approaches to Transition”.

“Radiology Smart Assistant | Philips Healthcare”. 2025. *Philips*. Accedido marzo 21. https://www.philips.co.za/healthcare/resources/landing/radiology-smart-assistant.

Ranjbar, Arian, Eilin Mork, Jesper Ravn, Helga Brøgger, Per Myrseth, Hans Peter Østrem, y Harry Hallock. 2024. “Managing Risk and Quality of AI in Healthcare: Are Hospitals Ready for Implementation?” *Risk Management and Healthcare Policy* Volume 17 (abril): 877–82. doi:10.2147/RMHP.S452337.

Rasche, Anna, Gerben ten Cate, y Axel Habecker. 2023. “Ysio X.Pree: How Intelligent Imaging Ensures an Efficient and Patient-Centered Radiography Workflow”.

Rayan, Rehab A, Imran Zafar, y Christos Tsagkaris. 2021. “Deep Learning for Health and Medicine”. doi:https://doi.org/10.1515%2F9783110708127-001.

RICYT. 2022. “Número de patentes concedidas por la Oficina de Propiedad Intelectual de cada país según el lugar de residencia del titular 2013-2022.” https://app.ricyt.org/ui/v3/comparative.html?indicator=CPATOTOR&start_year=2013&end_year=2022.

Rini Novitasari, Dian Candra, Ahmad Zoebad Foeady, Muhammad Thohir, Ahmad Zaenal Arifin, Khoirun Niam, y Ahmad Hanif Asyhar. 2020. "Automatic Approach for Cervical Cancer Detection Based on Deep Belief Network (DBN) Using Colposcopy Data". En *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 415–20. Fukuoka, Japan: IEEE. doi:10.1109/ICAIIC48513.2020.9065196.

Rogan, Jessica, Sandra Bucci, y Joseph Firth. 2024. "Health Care Professionals' Views on the Use of Passive Sensing, Ai, and Machine Learning in Mental Health Care: Systematic Review with Meta-Synthesis". *JMIR Mental Health* 11 (enero): e49577. doi:10.2196/49577.

Royal Statistical Society (RSS) y Institute and Faculty of Actuaries. 2019. "A Guide for Ethical Data Science". https://rss.org.uk/RSS/media/News-and-publications/Publications/Reports%20and%20guides/A-Guide-for-Ethical-Data-Science-Final-Oct-2019.pdf.

Rubín, Carlos N. 2024. "La IA en la administración de negocios actual". *Cuadernos del CIMBAGE* 1 (26): 61–76. doi:10.56503/CIMBAGE/Vol.1/Nro.26(2024)/3021.

Sai, Siva, Aanchal Gaur, Revant Sai, Vinay Chamola, Mohsen Guizani, y Joel J. P. C. Rodrigues. 2024. "Generative AI for Transformative Healthcare: A Comprehensive Study of Emerging Models, Applications, Case Studies, and Limitations". *IEEE Access* 12: 31078–106. doi:10.1109/ACCESS.2024.3367715.

Savulescu, Julian, Alberto Giubilini, Robert Vandersluis, y Abhishek Mishra. 2024. "Ethics of Artificial Intelligence in Medicine". *Singapore Medical Journal* 65 (3): 150–58. doi:10.4103/singaporemedj.SMJ-2023-279.

Shenavarmasouleh, Farzan, Farid Ghareh Mohammadi, Khaled M. Rasheed, y Hamid R. Arabnia. 2023. "Deep Learning in Healthcare: An in-Depth Analysis". arXiv. doi:10.48550/arXiv.2302.10904.

Shvetcov, Artur, Joost Funke Kupper, Wu-Yi Zheng, Aimy Slade, Jin Han, Alexis Whitton, Michael Spoelma, et al. 2024. "Passive Sensing Data Predicts Stress in University Students: A Supervised Machine Learning Method for Digital Phenotyping". *Frontiers in Psychiatry* 15 (agosto): 1422027. doi:10.3389/fpsyt.2024.1422027.

Simion, Mona, y Christoph Kelp. 2023. "Trustworthy Artificial Intelligence". *Asian Journal of Philosophy* 2 (1): 8. doi:10.1007/s44204-023-00063-5.

Simonetta, Alessandro, y Maria Cristina Paoletti. 2024. "ISO/IEC Standards and Design of an Artificial Intelligence System" 3916 (7).

Solano Hernández, Janet Margarita, y José Luis Soriano Ávila. 2024. "El uso de la inteligencia artificial en la gestión empresarial para la toma de decisiones en la planeación de desplazamiento del material de lento movimiento". *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades* 5 (4). doi:10.56712/latam.v5i4.2499.

Steimers, André, y Moritz Schneider. 2022. "Sources of Risk of AI Systems". *International Journal of Environmental Research and Public Health* 19 (6). MDPI AG: 3641. doi:10.3390/ijerph19063641.

Sufyan, Nabil Saleh, Fahmi H. Fadhel, Saleh Safeer Alkhathami, y Jubran Y. A. Mukhadi. 2024. "Artificial Intelligence and Social Intelligence: Preliminary Comparison Study between AI Models and Psychologists". *Frontiers in Psychology* 15 (febrero): 1353022. doi:10.3389/fpsyg.2024.1353022.

Takale, Dattatray G. 2024. “A Study of Natural Language Processing in Healthcare Industries”. *Journal of Web Applications and Cyber Security* 2 (2). doi:https://doi.org/10.48001/JoWACS.

Tavory, Tamar. 2024. “Regulating AI in Mental Health: Ethics of Care Perspective”. *JMIR Mental Health* 11 (septiembre). doi:10.2196/58493.

Team, S. E. O. 2022. “The 5 Most Common Medical Imaging Techniques”. *Houston Physicians Hospital*. agosto 13. https://www.houstonphysicianshospital.com/the-5-most-common-medical-imaging-techniques/.

Thiers, Md PhD, Fabio A., y Zach Harned. 2024. “The Emerging Role of ISO 42001 Certification in Fostering the Deployment of Responsible Generative AI Healthcare Solutions”. doi:10.31219/osf.io/5ks37.

Trillo, Eduardo Tenés. 2023. “Impacto de la Inteligencia Artificial en las Empresas”. Madrid: Universidad Politécnica de Madrid.

Vaira, Luigi Angelo, Jerome R. Lechien, Vincenzo Abbate, Fabiana Allevi, Giovanni Audino, Giada Anna Beltramini, Michela Bergonzani, et al. 2024. “Validation of the Quality Analysis of Medical Artificial Intelligence (QAMAI) Tool: A New Tool to Assess the Quality of Health Information Provided by AI Platforms”. *European Archives of Oto-Rhino-Laryngology* 281 (11). Springer Science and Business Media LLC: 6123–31. doi:10.1007/s00405-024-08710-0.

Vivero, María Belén. 2024. “No todo aporta: regulación de la inteligencia artificial en ecuador”. *Lexis S.A.* https://www.lexis.com.ec/blog/legaltech/no-todo-aporta-regulacion-de-la-inteligencia-artificial-en-ecuador.

Wang, Dong, Yuan Zhang, Kexin Zhang, y Liwei Wang. 2020. “FocalMix: Semi-Supervised Learning for 3D Medical Image Detection”. En *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3950–59. Seattle, WA, USA: IEEE. doi:10.1109/CVPR42600.2020.00401.

Wind, Henel, y John Kenny. 2023. “Ai-Powered Virtual Health Assistants: Transforming Patient Engagement through Virtual Nursing”.

World Halth Organization. 2022. *Informe mundial sobre salud mental: transformar la salud mental para todos. Panorama general*. 1st ed. Geneva: World Health Organization.

World Health Organization. 2024. *Ethics and Governance of Artificial Intelligence for Health: Large Multi-Modal Models. WHO Guidance*. 1st ed. Geneva.

Wu, Yuyuan, Lijing Ma, Xinyi Li, Jingpeng Yang, Xinyu Rao, Yiru Hu, Jingyi Xi, et al. 2024. “The Role of Artificial Intelligence in Drug Screening, Drug Design, and Clinical Trials”. *Frontiers in Pharmacology* 15 (noviembre): 1459954. doi:10.3389/fphar.2024.1459954.

Yan, Quan, Yunfan Ye, Jing Xia, Zhiping Cai, Zhilin Wang, y Qiang Ni. 2023. “Artificial Intelligence-Based Image Reconstruction for Computed Tomography: A Survey”. *Intelligent Automation & Soft Computing* 36 (3): 2545–58. doi:10.32604/iasc.2023.029857.

Yoon, Jung Hyun, y Eun-Kyung Kim. 2021. “Deep Learning-Based Artificial Intelligence for Mammography”. *Korean Journal of Radiology* 22 (8): 1225. doi:10.3348/kjr.2020.1210.

Yungán-Cazar, Juan Carlos, y Carina Narváez-Contero. 2022. “Aplicación de la Norma ISO 27001 para la seguridad de los Sistemas de Información”. *Dominio de las Ciencias* 8 (3). doi:http://dx.doi.org/10.23857/dc.v8i3.

Zapata Cortés, Julián Andrés. 2020. “Inteligencia artificial para la toma de decisiones”. *Revista Perspectiva Empresarial* 7 (2–1): 3–5. doi:10.16967/23898186.663.

Zarghami, Anita. 2024. “Role of Artificial Intelligence in Surgical Decision-Making: A Comprehensive Review: Role of AI in SDM”. *Galen Medical Journal* 13 (marzo): e3332. doi:10.31661/gmj.v13i.3332.

Zhang, Jie, y Zong-ming Zhang. 2023. “Ethics and Governance of Trustworthy Medical Artificial Intelligence”. *BMC Medical Informatics and Decision Making* 23 (1): 7. doi:10.1186/s12911-023-02103-9.

Anexos

Anexo 1: Resumen de controles y cláusulas relevantes de la norma ISO/IEC 42001:2023 para la gestión de desafíos éticos

Tabla 4

Controles del anexo A de la norma ISO/IEC 42001:2023 mencionados en las recomendaciones para gestión de los desafíos éticos identificados

Desafío	Diseño y desarrollo	Implementación y operación
Reducir el sesgo y asegurar la equidad	A.4.6, A.8.2, A.7.6, A.7.2, A.7.4, A.8.4	A.10.3, A.9.3, A.9.4, A.8.5
Dotar de transparencia y explicabilidad a los sistemas de IA	A.6.1.2, A.4.3, A.4.4, A.4.6, A.7.2, A.6.2.6	A.10.3, A.4.6, A.8.5, A.6.2.6, A.6.2.7, A.6.2.4
Evitar la mercantilización de la salud	A.5.4, A.5.5, A.6.2.6	A.9.2, A.9.3
Evitar la obsolescencia y deshumanización	A.9.3, A.6.1.2,	A.5.2, A.10.3, A.6.2.6
Asegurar la privacidad de la información	A.7.3, A.4.3, A.6.2.4	A.10.3, A.3.2, A.4.6
Evitar la erosión de la relación paciente-médico y prevenir la dependencia excesiva del profesional o paciente hacia los sistemas de IA	A.5.2, A.6.1.2, A.8.2, A.5.4,	A.10.3, A.6.2.6, A.3.2, A.4.6, A.9.3
Prevenir la insuficiente validación clínica en los sistemas de IA y asegurar la precisión y reproductibilidad de los resultados	A.6.2.6, A.7.2, A.7.5, A.7.6, A.4.4, A.10.4, A.6.2.4, A.8.3.	A.10.3, A.8.2, A.6.2.4, A.4.6, A.8.3, A.8.5, A.6.2.6

Fuente y elaboración propias

Tabla 5

Clausulas referidas de la norma ISO/IEC 42001:2023

Anexo	Descripción detallada del control
A.3.2	Definir y asignar roles y responsabilidades específicas para la gestión de IA en la organización.
A.4.3	Documentar información sobre los datos utilizados como recurso en el sistema de IA.
A.4.4	Documentar información sobre los recursos de herramientas (software, algoritmos, etc.) utilizados en el sistema de IA.
A.4.6	Documentar información sobre los recursos humanos y sus competencias involucrados en el desarrollo, implementación, operación y mantenimiento de sistemas de IA.
A.5.2	Establecer un proceso para evaluar los posibles impactos del sistema de IA en individuos, grupos y sociedades a lo largo de su ciclo de vida.
A.5.4	Evaluar y documentar los impactos potenciales del sistema de IA en individuos o grupos durante todo su ciclo de vida.
A.5.5	Evaluar y documentar los potenciales impactos sociales que pueden ocasionar los sistemas de IA durante todo su ciclo de vida.
A.6.1.2	Identificar y documentar objetivos para el desarrollo responsable de sistemas de IA y asegurarse de que estos se integren en el ciclo de desarrollo.
A.6.2.4	Definir y documentar medidas de verificación y validación para el sistema de IA y los criterios para su uso.
A.6.2.6	Definir y documentar los elementos necesarios para la operación continua del sistema de IA, incluyendo monitoreo del desempeño, actualizaciones y soporte.
A.6.2.7	Determinar y proporcionar la documentación técnica necesaria del sistema de IA para los diferentes grupos de interés.

A.7.2	Definir, documentar e implementar procesos de gestión de datos para el desarrollo y la mejora del sistema de IA.
A.7.3	Documentar los detalles sobre adquisición y selección de los datos empleados en los sistemas de IA.
A.7.4	Definir criterios de calidad para los datos e implementarlos para garantizar que los datos usados en el desarrollo y operación del sistema de IA cumplen con los requisitos.
A.7.5	Definir y documentar el proceso de registro de la procedencia de los datos utilizados en el sistema de IA durante su vida útil.
A.7.6	Definir y documentar los criterios y métodos de preparación de los datos empleados en el sistema de IA.
A.8.2	Proporcionar a los usuarios la información necesaria, documentación e instrucciones para comprender y utilizar adecuadamente el sistema de IA.
A.8.3	Establecer canales para que partes interesadas externas puedan reportar impactos adversos del sistema de IA.
A.8.4	Definir y documentar un plan para comunicar incidentes relacionados con el sistema de IA a sus usuarios.
A.8.5	Documentar las obligaciones respecto a reportar información relevante del sistema de IA a las partes interesadas.
A.9.2	Definir y documentar los procesos para el uso responsable del sistema de IA conforme a las políticas organizacionales.
A.9.3	Identificar y documentar objetivos que guíen el uso responsable del sistema de IA dentro de la organización.
A.9.4	Asegurar que el sistema de IA se use de acuerdo con los usos previstos y la documentación asociada.
A.10.3	Establecer un proceso para garantizar que los servicios, productos o materiales de proveedores sean gestionados de acuerdo con los principios responsables de IA definidos por la organización.
A.10.4	Asegurar que el enfoque responsable adoptado por la organización para el desarrollo y uso de sistemas de IA contemple, atienda y se ajuste a las expectativas y necesidades de sus clientes.

Elaboración propia, Fuente: norma ISO/IEC 42001:2023

Anexo 2: Ejemplos de implementación de la norma ISO/IEC 42001:2023 para la reducción del sesgo en aplicaciones de IA

Tabla 6

Análisis FODA que satisface parte de la cláusula 4.1 de una organización que desarrolla aplicaciones IA cuyo propósito es disminuir el sesgo en sus modelos.

Fortalezas La organización posee un equipo multidisciplicario y con experiencia en ética y desarrollo de IA. La organización dispone de procesos documentados para el desarrollo y evaluación de sus modelos y de las herramientas que utiliza en su desarrollo. La cultura organizacional de la empresa promueve los valores éticos en todo el ciclo de vida de sus aplicaciones de IA.	*Oportunidades* La creciente regulación a nivel global de las herramientas de IA planteará nuevos retos hacia las organizaciones que desarrollan estos productos. Existe una demanda creciente de productos de IA responsables y éticos que puede incidir en la apertura de nuevos mercados. El entrenamiento adecuado de los sistemas de IA tomando en cuenta las poblaciones objetivo mejorará el desempeño de sus resultados.
Debilidades Limitación en la diversidad de datos para el entrenamiento. Recursos económicos limitados Falta de herramientas especializadas para la detección del sesgo	*Amenazas* Implicaciones legales en los posibles resultados sesgados del modelo. Presión del mercado para agilizar el desarrollo del modelo. Utilización de los modelos fuera del alcance establecido.

Fuente y elaboración propias.

Tabla 7

Identificación de partes interesadas en cumplimiento de la cláusula 4.2 de una organización que desarrolla aplicaciones IA cuyo propósito es disminuir el sesgo en sus modelos.

Partes interesadas	**Requisitos relevantes**
Equipo de desarrollo y diseño	Políticas que especifiquen el compromiso de la organización con objetivos éticos y asignación de recursos para cumplir con los objetivos propuestos
Accionistas	Mantener reputación corporativa y evitar implicaciones legales

Usuarios y clientes	Sistemas debidamente entrenados que garanticen su efectividad y precisión en las poblaciones a las que se atenderá
Reguladores y autoridades	Cumplimiento de leyes locales y la necesidad de asumir responsabilidad por los resultados de los modelos de IA
Proveedores	Políticas, acuerdos y estándares que garanticen las relaciones comerciales

Fuente y elaboración propias.

Tabla 8

Definición de roles, responsabilidades y autoridades en cumplimiento de la cláusula 5.3 en una organización que desarrolla aplicaciones IA cuyo propósito es disminuir el sesgo en sus modelos

Puesto	**Roles**	**Responsabilidades**	**Autoridades**
Responsable del sistema de gestión de IA	Se asegura de mantener en el SG las conformidades necesarias para la conformidad con el estándar ISO/IEC 42001:2023 y supervisa los procesos para identificar su eficacia en la mitigación de sesgos.	Diseñar, implementar y mantener el SG de IA Gestionar canales de comunicación con el cliente Garantizar el cumplimiento regulatorio y normativo.	Aprobar o rechazar el avance del desarrollo o implementación de los modelos de IA Pedir revisiones o auditorías de procesos.
Equipo de ética de IA	Es un equipo multidisciplinario que revisa las métricas y resultados de los modelos para identificar sesgos y discriminaciones, en base a esto propone medidas correctivas para minimizarlo.	Desarrollar y mantener actualizados los objetivos éticos. Facilitar la participación de grupos de interés Promover capacitación al personal en cuestione éticas.	Exigir acciones correctivas en los modelos de IA ante el incumplimiento de los objetivos éticos. Elevar a la alta dirección sus recomendaciones
Auditor interno	Se encarga de las auditorías periódicas con el objetivo de verificar la conformidad con el SG, reporta sus hallazgos a la alta dirección	Planificar y ejecutar auditorías Auditar la representatividad de los datos y herramientas usados en el entrenamiento de los modelos Aplicar métricas y pruebas de sesgo Evaluar procesos para mitigar el sesgo	Solicitar auditorías externas Exigir información a los involucrados en el SG. Elevar sus observaciones a la alta dirección
Personal técnico y desarrolladores	Desarrollar los modelos de IA en base a los objetivos éticos de la organización e implementar las correcciones necesarias para minimizar el sesgo	Preparar y gestionar los datos para el entrenamiento o validación de los modelos. Implementar controles de sesgo, medir y documentar los resultados. Comunicar y responder a auditores, líder de proyecto o comités éticos sobre el avance, riesgos identificados y resultados de las pruebas de sesgo. Cumplir las normas y políticas internas Implementar las acciones correctivas propuestas por el líder del proyecto	Elegir algoritmos según los requerimientos del proyecto y las políticas de la organización. Proponer innovaciones o mejoras en los procesos técnicos

Fuente y elaboración propias.

Tabla 9

Matriz de riesgos en cumplimiento del capítulo 6 de una organización que desarrolla aplicaciones IA cuyo propósito es disminuir el sesgo en sus modelos

Partes interesadas	**Requisitos relevantes**	**Riesgos asociados al sesgo**	**Oportunidades relacionadas al sesgo**	**Medidas de gestión recomendadas, basadas en el anexo A de la norma ISO/IEC 42001:2023**	**Evaluación de riesgo**	**Evaluación de impacto**
Equipo de diseño y desarrollo	Políticas claras y asignación de recursos para cumplir con los objetivos éticos porpuestos	La falta de recursos o políticas claras pueden derivar en sesgos inadvertidos	La implementación de políticas que mejoren la cultura ética organizacional	Definir políticas claras que prioricen la inclusión y diversidad Determinar y documentar claramente las poblaciones hacia las que la aplicación es destinada. Asignar responsabilidades claras al personal Establecer procesos verificables para la limpieza, preprocesamiento y estandarización de datos Capacitación constante al personal en la detección y reducción del sesgo Monitoreo Continuo en el ciclo de vida del modelo de IA	Probabilidad media alta	Alto impacto en el desempeño del modelo y la reputación de la empresa.
Accionistas	Mantener reputación corporativa y evitar problemas legales	Deterioro en la reputación de la empresa y sanciones legales por resultados erróneos o sesgos discriminatorios	Adquirir diferenciación en responsabilidad ética frente a la competencia que incremente el reconocimiento y prestigio de la organización	Monitoreo continuo y adaptación a nuevas regulaciones Formar y sensibilizar al personal sobre la importancia de la inclusión y diversidad o incluir en el personal grupos demográficos específicos priorizando la inclusión y diversidad	Baja o media	Impacto negativo alto que afecta al aspecto económico y legal.

Usuarios y clientes	Sistemas de IA que brinden resultados confiables	Utilización del modelo fuera de la población objetivo con la que fue entrenado	Diseñar sistemas con un alcance poblacional definido y resultados confiables	El alcance definido por la organización debe ser comunicado a sus clientes y advertirles sobre los posibles riesgos asociados al utilizar el sistema fuera de este. Establecer canales de comunicación para recibir retroalimentación del usuario sobre el desempeño de los modelos de IA	Alta	Impacto negativo alto porque afecta la precisión de los resultados y en consecuencia la confianza en su desempeño.
Reguladores y autoridades	Cumplimiento de leyes locales y responsabilidad ante resultados adversos	Multas, sanciones, suspenciones y el deterioro de la reputación de la empresa	Generar confianza a las partes interesadas y ganar reputación mediante el cumplimiento de las regulaciones.	Establecer procesos de monitoreo y actualización normativa	Media	Alto impacto en cuestiones económicas y legales
Proveedores	Políticas, acuerdos y estándares que garanticen las relaciones comerciales	Insumos con características fuera de lo requerido por la organización	Lograr alianzas con proveedores alineados a la gestión ética de la empresa	Establecer políticas y procedimientos para la evaluación de proveedores Establecer parámetros de calidad para los insumos y comunicar en forma clara a los proveedores Implementar métricas para la evaluación del insumo	Medio	Alto impacto en la calidad del producto final

Fuente y elaboración propias.