

Revisión de uso de tecnologías éticas para una buena gobernanza de la inteligencia artificial

Review of the use of ethical technologies for good governance of artificial intelligence

José Yesán-Luján¹, Marcos Diaz-Tomas²,
Alberto Mendoza-De los Santos³

Yesán-Luján, J; Diaz-Tomas, M; Mendoza-De los Santos, A.
Revisión de uso de tecnologías éticas para un buen gobernanza de la inteligencia artificial . *Tecnología en Marcha*. Vol. 39 N° especial sobre Inteligencia Artificial. Febrero, 2026. Pág. 16-27.

 <https://doi.org/10.18845/tm.v39i5.8526>



1 Facultad de Ingeniería, Escuela de Ingeniería de Sistemas, Universidad Nacional de Trujillo. Perú.

 t1013300321@unitru.edu.pe

 <https://orcid.org/0009-0003-1406-8735>

2 Facultad de Ingeniería, Escuela de Ingeniería de Sistemas, Universidad Nacional de Trujillo. Perú.

 t1533300221@unitru.edu.pe

 <https://orcid.org/0009-0006-0210-1288>

3 Facultad de Ingeniería, Escuela de Ingeniería de Sistemas, Universidad Nacional de Trujillo. Perú.

 amendozad@unitru.edu.pe

 <https://orcid.org/0000-0002-0469-915X>

Palabras clave

Gobernanza de la IA; tecnologías éticas; inteligencia artificial; sesgo algorítmico; transparencia; auditorías éticas.

Resumen

Este artículo de revisión bibliográfica descriptiva examina la implementación de tecnologías éticas en el marco de la gobernanza de la inteligencia artificial, analizando los avances en la integración de principios normativos en sistemas computacionales. Ante la expansión de la IA en sectores críticos, se identifican desafíos como el sesgo algorítmico, la privacidad y la explicabilidad. Mediante un análisis de 19 artículos científicos (2021-2025) seleccionados de bases de datos académicas bajo criterios de actualidad y relevancia temática, se sistematizan enfoques como el pragmatismo sociotécnico y las auditorías éticas. El principal aporte de este estudio es la identificación de la carencia de un modelo integrado que armonice los requisitos técnicos con las demandas sociales. Se concluye que una gobernanza efectiva requiere una regulación flexible y multiactor que incorpore la ética en todo el ciclo de vida del desarrollo, garantizando la equidad y la protección de los derechos humanos.

Keywords

AI governance; ethical technologies; artificial intelligence; algorithmic bias; transparency; ethical audits.

Abstract

This descriptive literature review examines the implementation of ethical technologies within the framework of artificial intelligence governance, analyzing recent advances in integrating ethical principles into AI systems. As AI expands across critical sectors, significant challenges emerge, such as algorithmic bias, privacy, and explainability. Through an analysis of 19 peer-reviewed articles (2021-2025) selected based on currency and thematic relevance, this study systematizes frameworks like sociotechnical pragmatism and ethical audits. The primary contribution of this article is identifying the lack of an integrated approach that harmonizes technical requirements with social demands. It concludes that effective governance requires flexible, multi-stakeholder regulation that embeds ethics throughout the AI lifecycle, ensuring equity and the protection of human rights.

Introducción

La inteligencia artificial (IA) ha revolucionado una amplia gama de sectores, desde la salud y la educación hasta la gobernanza pública y la industria, generando tanto promesas como preocupaciones. Gracias a su capacidad para procesar grandes volúmenes de datos y realizar tareas complejas de manera autónoma, la IA ha abierto nuevas posibilidades para mejorar la eficiencia y efectividad de los servicios. Sin embargo, con el aumento de su adopción, también han surgido retos éticos y sociales que requieren atención urgente. Estos desafíos deben ser gestionados para garantizar que la IA se implemente de manera responsable, equitativa y alineada con los valores humanos fundamentales. A medida que los sistemas de IA toman decisiones autónomas, desde la asignación de recursos hasta el diagnóstico médico, es esencial que sus algoritmos operen bajo principios de justicia, equidad y transparencia para evitar la perpetuación de sesgos y desigualdades preexistentes [1].

En particular, las cuestiones de responsabilidad algorítmica, la explicabilidad y la transparencia de los sistemas de IA son clave para asegurar que estos operen de manera justa y que las decisiones tomadas por los algoritmos puedan ser comprendidas y cuestionadas por las personas afectadas. Los sistemas de IA pueden presentar sesgos algorítmicos que reflejan prejuicios inherentes en los datos utilizados para entrenarlos. Además, la falta de transparencia en los modelos de IA puede generar desconfianza, ya que los usuarios y los reguladores no tienen acceso a los criterios y procesos de toma de decisiones utilizados por la IA. Estos problemas no solo afectan la credibilidad de la tecnología, sino que también pueden perpetuar y amplificar las desigualdades sociales y económicas existentes. Por lo tanto, resulta urgente desarrollar marcos de gobernanza ética que aseguren que la IA no solo sea efectiva, sino también justa, equilibrada y respetuosa con los derechos fundamentales de las personas [2].

La gobernanza de la IA implica la creación y la implementación de marcos regulatorios, principios éticos y mecanismos de supervisión para garantizar que los sistemas de IA se desarrollen de manera alineada con los valores fundamentales de la sociedad. Estos marcos deben incluir tanto los aspectos técnicos de la IA, como la precisión y eficiencia de los algoritmos, como los aspectos sociales y éticos, como la responsabilidad, la equidad y el respeto a la dignidad humana. Las decisiones tomadas por sistemas autónomos deben ser explicables y transparentes para que los usuarios y reguladores puedan entender cómo y por qué se tomaron determinadas decisiones. Sin embargo, a medida que los sistemas de IA se vuelven más complejos y diversificados, la supervisión de su funcionamiento se convierte en un desafío aún mayor para los reguladores. Esto requiere la implementación de mecanismos regulatorios y auditorías éticas que aseguren que los sistemas de IA sigan principios éticos sólidos y que sean responsables ante los usuarios y las comunidades que afectan [3].

Este artículo tiene como objetivo realizar una revisión crítica de las investigaciones más recientes sobre el uso de tecnologías éticas en la gobernanza de la IA. A través de un análisis de 19 artículos clave publicados entre 2021 y 2025, se exploran los enfoques más innovadores y las mejores prácticas que buscan mitigar los riesgos éticos asociados con la IA, asegurando su implementación de manera responsable y justa. La revisión cubre tanto las estrategias regulatorias como los marcos técnicos que buscan integrar principios éticos en todas las fases del ciclo de vida de la IA. En particular, se revisan enfoques que integran tanto los aspectos técnicos (como la optimización de los algoritmos para evitar sesgos) como los aspectos sociales de la IA, con el objetivo de construir confianza en la tecnología y garantizar que no perpetúe desigualdades ni cause daños no deseados en la sociedad. Los resultados de esta revisión proporcionan una visión integral sobre las prácticas actuales y las áreas de mejora en la gobernanza ética de la IA, ofreciendo recomendaciones clave para el desarrollo de políticas que garanticen un futuro en el que la IA sea una herramienta para el bien común [4].

Metodología

El objetivo primordial de este estudio es desarrollar una revisión sistematizada sobre la implementación de tecnologías éticas en el marco de la gobernanza de la inteligencia artificial. Se priorizaron marcos regulatorios y principios normativos que sustentan el desarrollo responsable de sistemas autónomos. El enfoque metodológico adoptado es de carácter descriptivo-analítico, fundamentado en la revisión de literatura científica de alto impacto publicada entre 2021 y 2025, garantizando así la vigencia de los hallazgos en un campo de rápida evolución técnica.

Selección de Bases de Datos y Fuentes:

Se seleccionaron varias bases de datos académicas de alto impacto que proporcionan acceso a artículos revisados por pares en áreas de tecnología, ética, y gobernanza. Las bases de datos utilizadas fueron las siguientes:

- Scopus: Considerada una de las principales fuentes para acceder a artículos científicos sobre tecnología y ética en la IA.
- Google Académico: Para obtener una visión más amplia y accesible de artículos relevantes que puedan no estar indexados en otras bases de datos.
- IEEE Xplore: Para acceder a artículos de conferencias y revistas sobre IA y sus implicaciones éticas.
- Springer Nature: Para incluir literatura sobre marcos regulatorios, principios éticos y aplicaciones de la IA en el sector público y privado.

Términos de Búsqueda:

La búsqueda se realizó utilizando una combinación de términos clave relacionados con gobernanza de la IA, tecnologías éticas y los principios éticos de la IA. Los términos específicos utilizados fueron:

- Gobernanza de IA OR AI Governance
- Tecnologías éticas OR ethical technologies
- Inteligencia artificial OR artificial intelligence
- IA ética OR ethical AI
- Responsabilidad algorítmica OR algorithmic accountability
- Explicabilidad de la IA OR AI explainability
- Transparencia en IA OR transparency in AI

Se emplearon operadores booleanos (AND, OR) para asegurar que se cubrieran diversas combinaciones de estos términos y obtener artículos relevantes en las áreas de interés.

Criterios de Inclusión y Exclusión:

Para garantizar la relevancia y calidad de los artículos seleccionados, se aplicaron los criterios de inclusión y exclusión mostrados en el cuadro 1.

Cuadro 1. Criterios de inclusión y exclusion.

Criterio	Inclusión	Exclusión
Temporalidad	Artículos publicados entre 2021 y 2025.	Publicaciones anteriores a 2021.
Tipo de fuente	Artículos de revistas científicas revisados por partes.	Libros, tesis, actas de conferencias sin revisión o artículos de opinión.
Temática	Gobernanza ética de la IA, marcos regulatorios y principios éticos.	Artículos centrados exclusivamente en aspectos técnicos (hardware/código) sin enfoque ético.
Ámbito de aplicación	Salud, educación, gestión pública y administración.	Ámbitos ajenos al interés del estudio o de carácter estrictamente comercial.
Accesibilidad	Artículos con acceso gratuito o a través de repositorios institucionales.	Documentos sin acceso al texto completo o de fuentes no verificadas.

Estrategia de Búsqueda y Revisión:

El proceso de búsqueda y selección se dividió en varias etapas:

Busqueda Inicial:

Se realizaron búsquedas en las bases de datos seleccionadas utilizando los términos establecidos. Los resultados obtenidos fueron filtrados según los criterios de fecha (2021-2025) y tipo de artículo (solo artículos académicos). Se realizaron búsquedas adicionales para asegurarse de que los artículos estuvieran disponibles para su descarga o acceso gratuito.

Revisión de títulos y resúmenes:

Los artículos inicialmente seleccionados fueron revisados en función de sus títulos y resúmenes. Esta revisión preliminar permitió identificar los artículos que abordaban de manera relevante los temas de la gobernanza ética, los marcos regulatorios, y los principios éticos en la IA. Se descartaron artículos que no cumplieran con los objetivos de la revisión.

Lectura completa y selección final:

Se procedió a la lectura completa de los artículos seleccionados. En esta fase, se verificó que los artículos cumplieran con los criterios de inclusión establecidos. Después de la lectura completa, se eligieron los 18 artículos más relevantes que trataban de manera profunda los temas de gobernanza ética, tecnologías éticas en la IA, y regulación en los sistemas de IA.

Análisis de los Artículos:

Los artículos seleccionados fueron analizados en detalle según los siguientes criterios:

- Marcos regulatorios propuestos: Se identificaron los enfoques regulatorios y las mejores prácticas utilizadas para abordar los riesgos éticos asociados con la IA.
- Enfoques éticos: Se revisaron los enfoques que abordan el sesgo algorítmico, la explicabilidad, y la transparencia de los sistemas de IA.
- Recomendaciones políticas: Se analizaron las recomendaciones para políticas públicas y marcos regulatorios con el fin de asegurar un uso ético y responsable de la IA.
- Aplicaciones prácticas: Se examinaron las aplicaciones de IA ética en sectores clave como la salud y la gobernanza pública, evaluando cómo los marcos regulatorios propuestos pueden ser implementados efectivamente en la práctica.

Limitaciones:

Es importante señalar que, aunque se utilizó un enfoque sistemático para la selección y análisis de los artículos, la literatura revisada no abarca todos los aspectos posibles de la gobernanza de la IA. La revisión se centró en los marcos regulatorios y principios éticos, dejando fuera otros aspectos relevantes como la tecnología subyacente en los sistemas de IA o el impacto económico de la adopción masiva de la IA.

Resultados

En esta sección, se presentan los resultados obtenidos a partir del análisis de los 19 artículos clave. Los artículos fueron clasificados en función de sus principales aportaciones a la gobernanza ética de la IA, los marcos regulatorios y los principios éticos utilizados para mitigar los riesgos éticos asociados con la implementación de la IA (ver cuadro 2). A continuación, se presentan los resultados organizados en varias categorías clave, con las citas correctas en formato IEEE y las contribuciones clave de cada artículo.

Cuadro 2. Principales Aportaciones de los Artículos.

Nro	Autor(es)	Artículo	Principal Aportación
1	J. Mökander, L. Floridi	From algorithmic accountability to digital governance	Propuesta de un enfoque integral para la responsabilidad algorítmica en la gobernanza digital, con énfasis en la transparencia y la justicia social.
2	D.S. Watson, J. Mökander, L. Floridi	Competing narratives in AI ethics: a defense of sociotechnical pragmatism	Defensa del pragmatismo sociotécnico como un enfoque para integrar tanto los aspectos técnicos como los sociales en la ética de la IA.
3	C. Thomas, H. Roberts, J. Mökander et al.	The case for a broader approach to AI assurance: addressing “hidden” harms in the development of artificial intelligence	Argumento para un enfoque más amplio en la garantía de IA, abordando los daños ocultos durante el desarrollo de la IA y cómo la regulación puede mitigarlos.
4	J. Mökander, R. Schroeder	Artificial intelligence, rationalization, and the limits of control in the public sector	Análisis de los límites del control de la IA en el sector público, particularmente en la optimización de políticas fiscales.
5	J. Amann, D. Vetter, S.N. Blomberg et al.	To explain or not to explain?—Artificial intelligence explainability in clinical decision support systems	Discusión sobre la explicabilidad de la IA en sistemas de soporte a decisiones clínicas, destacando la importancia de la transparencia para ganar confianza pública.
6	T. Hagendorff	Blind spots in AI ethics	Identificación de las lagunas éticas en la IA y la necesidad de integrar principios éticos más sólidos en el desarrollo de sistemas de IA.
7	T. Hagendorff	A virtue-based framework to support putting AI ethics into practice	Propuesta de un marco basado en virtudes para la ética de la IA, enfatizando la responsabilidad, la justicia y la equidad como principios fundamentales.
8	R.V. Zicari et al.	Co-design of a trustworthy AI system in healthcare: deep learning based skin lesion classifier	Co-diseño de un sistema de IA confiable para el sector salud, enfocado en la creación de un clasificador ético de lesiones cutáneas utilizando deep learning.
9	R.V. Zicari et al.	On assessing trustworthy AI in healthcare. Machine learning as a supportive tool to recognize cardiac arrest	Evaluación de la confianza en IA aplicada a la salud, utilizando machine learning para el reconocimiento de paro cardíaco, y la necesidad de auditorías éticas.
10	T. Hagendorff et al.	Speciesist bias in AI: how AI applications perpetuate discrimination and unfair outcomes against animals	Discusión sobre el sesgo especista en la IA, abordando cómo los sistemas de IA pueden perpetuar discriminación contra los animales y la necesidad de integrar principios éticos.
11	R.V. Zicari et al.	Co-design of trustworthy AI in healthcare: deep learning as a tool	Investigación sobre cómo el co-diseño de IA en el ámbito de la salud asegura confianza y ética en el desarrollo de modelos predictivos para diagnóstico médico.
12	Thilo Hagendorff, S. Fabi	Why we need biased AI: How including cognitive biases can enhance AI systems	Propuesta de que la inclusión de sesgos cognitivos en la IA puede mejorar su efectividad, sugiriendo un enfoque ético controlado para su implementación.
13	Maricielo Estefany Caciano-Arroyo et al.	Integrating AI into IT Governance: A systematic review	Revisión sistemática sobre la integración de la IA en la gobernanza de TI, destacando su potencial para mejorar la eficiencia y la toma de decisiones en instituciones.
14	Jakob Mökander	Auditing large language models: a three-layered approach	Propuesta de un enfoque en tres capas para auditar modelos de lenguaje, asegurando que sean transparentes y libres de sesgo.
15	S. Taeihagh	Governance of Generative AI (Introductory Article)	Introducción a los desafíos éticos y regulatorios del uso de IA generativa, proponiendo estrategias para una gobernanza ética en este campo emergente.
16	Vinay Gali, Arpit Jain	Ethical Considerations and Responsible Governance of Generative AI: A Systematic Review	Revisión sistemática sobre la gobernanza ética de la IA generativa, proporcionando un análisis crítico de los desafíos éticos y los marcos regulatorios propuestos.
17	J. Mökander et al.	Auditing AI in decision-making systems: Accountability and transparency	Propuesta de auditoría para sistemas de IA en la toma de decisiones, destacando la importancia de responsabilidad y explicabilidad en la toma automatizada de decisiones.
18	T. Hagendorff et al.	Ethics of Autonomous Systems: Ethical frameworks for AI in governance	Análisis de los marcos éticos para los sistemas autónomos en la gobernanza pública, con un enfoque en la responsabilidad y el equilibrio entre humanos y máquinas.

Marcos Regulatorios y Estrategias de Gobernanza Ética:

Un tema central en los artículos revisados fue la necesidad de establecer marcos regulatorios que aseguren la implementación ética de la IA. La gobernanza de la IA no solo se refiere al uso de tecnologías de IA, sino también a la integración de principios éticos que guíen el desarrollo, implementación y supervisión de la IA, teniendo en cuenta sus impactos en la sociedad.

- **Enfoques Regulatorios:** Varios artículos proponen enfoques regulatorios, como el pragmatismo sociotécnico, que reconoce que la IA no debe verse solo como un conjunto de herramientas tecnológicas, sino como un sistema que afecta a las comunidades. Este enfoque destaca la necesidad de considerar no solo los aspectos técnicos de la IA, sino también sus impactos sociales y culturales. El pragmatismo sociotécnico sugiere que la IA debe ser regulada de manera que se minimicen los efectos negativos en las comunidades y se maximicen los beneficios sociales y económicos, considerando también la ética y los derechos humanos [1][2].
- **Regulaciones Flexibles:** Otro punto clave señalado en los artículos es la necesidad de regulaciones flexibles que puedan adaptarse a los rápidos avances tecnológicos. Dado el carácter evolutivo de la tecnología de IA, los marcos regulatorios deben ser adaptativos, permitiendo cambios rápidos sin comprometer los principios éticos fundamentales. Las regulaciones deben ser lo suficientemente generales como para abarcar una variedad de aplicaciones de IA, pero también lo suficientemente específicas como para abordar los riesgos emergentes y las particularidades de los diferentes sectores, como la salud, el transporte o la seguridad pública [3].
- **Políticas Públicas Globales:** La regulación a nivel internacional también se destacó como un área clave. Los artículos abogan por la creación de políticas globales que fomenten una gobernanza compartida y responsable de la IA. Esto implica la colaboración entre actores gubernamentales, académicos y el sector privado, para crear marcos normativos adecuados que puedan hacer frente a los desafíos éticos globales. A nivel internacional, se reconoce la necesidad de estandarizar las normas éticas para asegurar que la IA sea utilizada de manera justa, transparente y sin riesgo de perpetuar desigualdades [4].

Principios Éticos en el Desarrollo de la IA:

Los artículos revisados subrayan que los sistemas de IA deben desarrollarse y desplegarse bajo principios éticos claros para evitar riesgos como el sesgo, la discriminación y la falta de transparencia. En este sentido, los principios clave que guían la gobernanza ética de la IA son los siguientes:

Responsabilidad Algorítmica: La responsabilidad algorítmica es uno de los principios más discutidos en la literatura. Varios estudios destacan que los desarrolladores de IA y las organizaciones deben ser responsables de los impactos de las decisiones tomadas por la IA, especialmente cuando afectan directamente a las personas. La responsabilidad incluye la capacidad de auditar y explicar las decisiones algorítmicas, y garantizar que estas sean comprensibles para los usuarios afectados. De esta manera, se asegura que las decisiones automatizadas sean justas, verificables y transparentes, y que los errores sean corregibles de manera eficiente. Este enfoque busca evitar el sesgo y la discriminación en los sistemas de IA [5].

Explicabilidad y Transparencia: Otro principio fundamental que se abordó ampliamente en los artículos es la explicabilidad. Los sistemas de IA deben ser lo suficientemente transparentes para que los usuarios puedan comprender cómo y por qué se toman ciertas decisiones, especialmente en sectores críticos como la salud o el derecho. La explicabilidad no solo

permite a los usuarios confiar más en las decisiones tomadas por IA, sino que también facilita la rendición de cuentas ante los reguladores y la sociedad. La transparencia es crucial para prevenir que los sistemas de IA perpetúen sesgos preexistentes o que operen de manera opaca, sin un marco que permita cuestionar sus decisiones [6].

Justicia y Equidad: La justicia y la equidad son principios igualmente fundamentales en la gobernanza ética de la IA. Varias investigaciones coinciden en que los sistemas de IA deben ser justos y equilibrados, evitando la reproducción de desigualdades sociales. Esto implica que los datos utilizados para entrenar los modelos de IA deben ser diversos y representativos, de manera que los algoritmos no favorezcan a ciertos grupos sociales en detrimento de otros. La imparcialidad debe ser un objetivo clave en el diseño de IA, de modo que todos los usuarios se beneficien de los mismos estándares éticos, independientemente de su género, raza, origen étnico u otras características demográficas [7].

Aplicaciones de la IA Ética en Sectores Clave:

Los artículos revisados también analizan cómo los principios de gobernanza ética y los marcos regulatorios propuestos se aplican en sectores clave como la salud, la educación y la gobernanza pública. A continuación, se detallan las aplicaciones específicas en estos sectores:

IA en la Salud: En el sector salud, varios estudios destacan cómo la IA puede mejorar la atención médica, pero advierten sobre los riesgos de discriminación y sesgo si no se implementan adecuadamente los principios éticos. Por ejemplo, la IA puede ayudar a diagnosticar enfermedades con mayor rapidez y precisión, pero si los datos de entrenamiento no son representativos, los sistemas de IA pueden generar diagnósticos sesgados que perjudiquen a ciertos grupos de pacientes. Los artículos sugieren que la explicabilidad y la responsabilidad son cruciales para garantizar que los sistemas de IA sean seguros, justos y transparentes en el diagnóstico y tratamiento de pacientes [8].

IA en la Gobernanza Pública: La implementación de IA en la gestión pública y las políticas fiscales también se analiza en los artículos. Aquí, los estudios subrayan que, aunque la IA puede mejorar la eficiencia administrativa, la supervisión humana sigue siendo necesaria para evitar decisiones sesgadas o injustas, especialmente en el contexto de la distribución de recursos públicos. Además, se menciona que los algoritmos de IA no deben usarse para perpetuar desigualdades ni desinformación en la administración pública. Los marcos regulatorios deben garantizar que la IA pública sea utilizada para beneficio común, sin discriminación ni exclusión social [9].

Desafíos y oportunidades futuras:

Un tema recurrente en los artículos es la necesidad de encontrar un equilibrio entre la innovación tecnológica y el respeto a los principios éticos. A medida que los sistemas de IA se vuelven más sofisticados, los problemas de transparencia y explicabilidad continúan siendo barreras significativas para su implementación ética. Sin embargo, los artículos también destacan diversas oportunidades para mejorar la gobernanza ética de la IA:

Desarrollo de Marcos Regulatorios Globales: Un avance crucial es el desarrollo de marcos regulatorios globales para la IA, que puedan estandarizar las normas éticas y las mejores prácticas. Esto permitiría una gobernanza compartida de la IA a nivel internacional, asegurando que los principios éticos sean aplicados de manera consistente y universal. Los artículos sugieren que la cooperación internacional es esencial para crear normas éticas globales que aborden los riesgos y los beneficios de la IA [10].

Colaboración Intersectorial: Fomentar la colaboración intersectorial es otro paso clave para mejorar la gobernanza ética de la IA. Esto incluye la cooperación entre gobiernos, universidades, industria y organizaciones no gubernamentales para desarrollar marcos regulatorios que se adapten a los cambios rápidos en la tecnología, sin perder de vista los principios éticos fundamentales. Además, se destaca el papel crucial de las auditorías independientes y la transparencia para garantizar que los sistemas de IA operen de manera ética y responsable [11].

Discusión

Esta revisión ha permitido identificar y analizar los principales aspectos de la gobernanza ética de la inteligencia artificial (IA), con un enfoque en los marcos regulatorios, los principios éticos y las aplicaciones prácticas de la IA en sectores clave. A través del análisis de 19 artículos clave, se ha podido observar que, aunque se ha avanzado significativamente en la integración de principios éticos en la IA, persisten desafíos importantes que deben abordarse para asegurar una implementación justa, responsable y transparente de estos sistemas.

Un hallazgo fundamental en los artículos revisados es la necesidad de marcos regulatorios robustos y flexibles que aseguren la implementación ética de la IA. Si bien algunos países han comenzado a desarrollar regulaciones para la IA, los marcos existentes no son lo suficientemente adaptativos para manejar los rápidos avances en la tecnología. Los enfoques propuestos, como el pragmatismo sociotécnico, sugieren que la gobernanza de la IA debe integrar no solo los aspectos técnicos, sino también los impactos sociales, culturales y éticos. Es crucial que los marcos regulatorios se adapten a la evolución de la tecnología y permitan ajustes rápidos para enfrentar nuevos retos éticos y sociales [1][2].

La cooperación internacional se presenta como una necesidad urgente, dado que la IA es una tecnología global. La armonización de las regulaciones entre diferentes países y sectores es esencial para evitar la fragmentación normativa y asegurar que las prácticas de gobernanza ética sean coherentes y efectivas a nivel mundial [3][11]. Los artículos destacan la necesidad de políticas globales que promuevan una gobernanza compartida y que involucren a actores gubernamentales, académicos y del sector privado para desarrollar normas comunes.

En cuanto a los principios éticos, la responsabilidad algorítmica emerge como un tema central. Los desarrolladores de IA y las organizaciones deben ser responsables por las decisiones tomadas por sus sistemas, especialmente cuando estas decisiones tienen un impacto directo en las personas. La necesidad de sistemas de IA auditables y explicables es crucial para que los usuarios comprendan cómo y por qué se toman las decisiones automatizadas. Sin embargo, a pesar de los esfuerzos por mejorar la transparencia y la explicabilidad, los artículos subrayan que aún existen barreras técnicas y sociales que dificultan que los sistemas sean completamente comprensibles, especialmente en sectores críticos como la salud y el derecho [5][12].

La justicia y la equidad también son principios clave que se deben incorporar en el diseño de los sistemas de IA. La IA debe evitar la reproducción de sesgos y desigualdades sociales que ya existen en la sociedad. Es necesario que los datos utilizados para entrenar los modelos de IA sean diversos y representativos de todas las poblaciones para evitar que los sistemas favorezcan a ciertos grupos sociales en detrimento de otros. Los artículos coinciden en que, para garantizar una IA imparcial, los desarrolladores deben ser proactivos en la identificación y eliminación de sesgos algorítmicos durante todo el ciclo de vida del sistema [6][7][13].

El análisis de los artículos también ha resaltado cómo los principios de gobernanza ética de la IA se aplican en sectores como la salud y la gobernanza pública. En el sector de la salud, la IA tiene el potencial de transformar la atención médica, mejorando la precisión de los diagnósticos y facilitando el tratamiento de enfermedades. Sin embargo, si los datos utilizados para entrenar los modelos no son representativos de todas las poblaciones, existe el riesgo de que los sistemas de IA perpetúen sesgos en los diagnósticos y tratamientos. Los estudios revisados subrayan que la explicabilidad y la responsabilidad son fundamentales para garantizar que los sistemas de IA sean utilizados de manera ética y transparente en este sector [8].

En el ámbito de la gobernanza pública, aunque se reconoce el potencial de la IA para mejorar la eficiencia administrativa y la distribución de recursos, también se identifican riesgos de sesgo y falta de supervisión que podrían resultar en decisiones injustas. Los marcos regulatorios deben garantizar que los algoritmos utilizados en la administración pública sean imparciales y transparentes, y que no se utilicen para perpetuar desigualdades ni favorecer intereses particulares. La supervisión humana sigue siendo crucial para evitar que los sistemas de IA perpetúen inequidades sociales en la gestión pública [9][14].

Por último, los artículos revisados también identifican varios desafíos y oportunidades futuras en la implementación ética de la IA. A pesar de los avances realizados, los sistemas de IA siguen enfrentando problemas relacionados con la transparencia y la explicabilidad, especialmente cuando se utilizan en contextos en los que las decisiones automatizadas tienen un impacto significativo en la vida de las personas. Se sugiere que la mejora de la transparencia y la creación de marcos regulatorios globales son pasos fundamentales para superar estos desafíos. Los estudios también destacan la importancia de la colaboración intersectorial, que involucra a gobiernos, industria y academia, para promover un enfoque integral que asegure que la IA se utilice de manera ética y responsable [10].

Conclusiones

Esta revisión, basada en 18 artículos publicados entre 2021 y 2025, ofrece una síntesis actualizada sobre la gobernanza ética de la inteligencia artificial (IA). Se evidencian avances en la formulación de marcos regulatorios y principios éticos, aunque persisten desafíos estructurales para garantizar un uso responsable, justo y transparente de la tecnología.

Aportes principales de la revisión:

- Necesidad de marcos regulatorios adaptativos y coherentes: Se confirma la urgencia de regulaciones flexibles, capaces de ajustarse al ritmo de innovación tecnológica sin desvincularse de principios éticos universales como justicia, responsabilidad y transparencia.
- Centralidad de la responsabilidad algorítmica: La rendición de cuentas de desarrolladores y organizaciones emerge como eje fundamental, especialmente en sistemas que afectan derechos y oportunidades de las personas.
- Importancia de la explicabilidad y la transparencia: Aunque existen limitaciones técnicas —particularmente en modelos complejos y en sectores sensibles como salud, empleo y gobernanza pública, se identifican avances en auditoría y diseño de modelos más interpretables.
- Énfasis en justicia y equidad: La revisión subraya la necesidad de prevenir sesgos y discriminaciones algorítmicas mediante datos diversos y representativos, condición clave para la legitimidad social de la IA.

Líneas futuras de investigación:

- Desarrollo de metodologías avanzadas de explicabilidad y auditoría en sistemas de alta complejidad.
- Diseño de estándares internacionales armonizados que permitan una gobernanza global coherente.
- Evaluación empírica del impacto real de marcos regulatorios existentes.
- Creación de herramientas técnicas para mitigación sistemática de sesgos.

Referencias

- [1] J. Mökander y L. Floridi, "From algorithmic accountability to digital governance," Springer Nature, 2022. [En línea]. Disponible en: <https://doi.org/10.1038/s42256-022-00504-5>
- [2] D. S. Watson, J. Mökander y L. Floridi, "Competing narratives in AI ethics: a defense of sociotechnical pragmatism," Springer Nature, 2024. [En línea]. Disponible en: <https://doi.org/10.1007/s00146-024-02128-2>
- [3] C. Thomas, H. Roberts, J. Mökander, A. Tsamados, M. Taddeo y L. Floridi, "The case for a broader approach to AI assurance: addressing 'hidden' harms in the development of artificial intelligence," Springer Nature, 2024. [En línea]. Disponible en: <https://doi.org/10.1007/s00146-024-01950-y>
- [4] J. Mökander and R. Schroeder, "Artificial Intelligence, Rationalization, and the Limits of Control in the Public Sector: The Case of Tax Policy Optimization," Social Science Computer Review, vol. 42, no. 6, pp. 1359–1378, Mar. 2024, doi: <https://doi.org/10.1177/08944393241235175>.
- [5] J. Amann et al., "To explain or not to explain?—Artificial intelligence explainability in clinical decision support systems," PLOS Digital Health, vol. 1, no. 2, p. e0000016, Feb. 2022, doi: <https://doi.org/10.1371/journal.pdig.0000016>.
- [6] T. Hagendorff, "Blind spots in AI ethics," Springer Nature, 2021. [En línea]. Disponible en: <https://doi.org/10.1007/s43681-021-00122-8>
- [7] T. Hagendorff, "A virtue-based framework to support putting AI ethics into practice," Springer Nature, 2022. [En línea]. Disponible en: <https://doi.org/10.1007/s43681-022-00162-8>
- [8] R. V. Zicari et al., "Co-Design of a Trustworthy AI System in Healthcare: Deep Learning Based Skin Lesion Classifier," Frontiers in Human Dynamics, vol. 3, Jul. 2021, doi: <https://doi.org/10.3389/fhumd.2021.688152>.
- [9] R. V. Zicari et al., "On Assessing Trustworthy AI in Healthcare. Machine Learning as a Supportive Tool to Recognize Cardiac Arrest in Emergency Calls," Frontiers in Human Dynamics, vol. 3, Jul. 2021, doi: <https://doi.org/10.3389/fhumd.2021.673104>.
- [10] T. Hagendorff et al., "Speciesist bias in AI: how AI applications perpetuate discrimination and unfair outcomes against animals," Springer Nature, 2022. [En línea]. Disponible en: <https://doi.org/10.1007/s43681-022-00199-9>
- [11] R. V. Zicari et al., "Co-design of trustworthy AI in healthcare: deep learning as a tool," Springer Nature, 2022. [En línea]. Disponible en: <https://doi.org/10.3389/fhumd.2021.688152>
- [12] "Why we need biased AI: How including cognitive biases can enhance AI systems," Journal of Experimental & Theoretical Artificial Intelligence, 2024, doi: <https://doi.org/10.1080/0952813X.2023.2178517>.
- [13] Maricielo Estefany Caciano-Arroyo, A. F. Vasquez-Cabrera, and A. C. Mendoza-de-los-Santos, "Integración de inteligencia artificial en la gobernanza de TI: Una revisión sistemática," Aibi revista de investigación administración e ingeniería, vol. 13, no. 2, pp. 1–12, May 2025, doi: <https://doi.org/10.15649/2346030X.4532>.
- [14] J. Mökander, "Auditing large language models: a three-layered approach," Springer Nature, 2023. [En línea]. Disponible en: <https://doi.org/10.1007/s43681-023-00289-2>
- [15] S. Taeihagh, "Governance of Generative AI (Introductory Article)," Oxford Academic, 2023. [En línea]. Disponible en: <https://doi.org/10.1093/polsoc/puaf001>
- [16] "Ethical Considerations and Responsible Governance of Generative AI: A Systematic Review - Premier Science," Premier Science, Apr. 26, 2025. <https://premier-science.com/pjai-25-800/> (accessed Feb. 28, 2026).
- [17] B. C. Cheong, "Transparency and accountability in AI systems: safeguarding wellbeing in the age of algorithmic decision-making," Frontiers in Human Dynamics, vol. 6, Jul. 2024, doi: <https://doi.org/10.3389/fhumd.2024.1421273>.

- [18] A. F. Winfield, K. Michael, J. Pitt, and V. Evers, "Machine Ethics: The Design and Governance of Ethical AI and Autonomous Systems [Scanning the Issue]," *Proceedings of the IEEE*, vol. 107, no. 3, pp. 509–517, Mar. 2019, doi: <https://doi.org/10.1109/jproc.2019.2900622>.

Declaración sobre uso de Inteligencia Artificial (IA)

En el presente trabajo, se ha empleado la herramienta de inteligencia artificial (IA) *ChatGPT* con el objetivo de llevar a cabo una revisión exhaustiva del texto, identificando inconsistencias, ambigüedades en la redacción y errores generales, asegurando así una redacción clara y coherente en todo el documento.