INTELIGENCIA ARTIFICIAL

EXPLICABLE Y TRANSPARENCIA ALGORÍTMICA: DESAFÍOS Y PROGRESOS EN DECISIONES AUTOMATIZADAS

EXPLAINABLE ARTIFICIAL INTELLIGENCE AND ALGORITHMIC TRANSPARENCY: CHALLENGES AND ADVANCES IN AUTOMATED DECISION-MAKING

REVISTA UGC | Revista científica de la Universidad del Golfo de California | ISSN: 2448-8097
Volumen 4 | Número 2 | abril-junio - 2026

Génesis Karolina Robles-Zambrano[1]
E-mail: uq.genesisrobles@uniandes.edu.ec
ORCID: https://orcid.org/0000-0002-2965-2091
Ingrid Joselyne Diaz-Basurto[1]
E-mail: uq.ingriddiaz@uniandes.edu.ec
ORCID: https://orcid.org/0000-0003-2934-4010
Deinier Ros-Álvarez[1]
E-mail: uq.deinierra09@uniandes.edu.ec
ORCID: https://orcid.org/0000-0002-1531-3355
Verónica Zuleyka Ramírez-Bósquez[1]
E-mail: veronicarb66@uniandes.edu.ec
ORCID: https://orcid.org/0009-0005-1187-893X
[1] Universidad Regional Autónoma de Los Andes. Ecuador.

Cita sugerida (APA, séptima edición)

Robles-Zambrano, G. K., Diaz-Basurto, I. J., Ros-Álvarez, D., & Ramírez-Bósquez, V. Z. (2026). Inteligencia artificial explicable y transparencia algorítmica: desafíos y progresos en decisiones automatizadas. *Revista UGC,* 4(2), 175-180.

Fecha de presentación: 26/12/2025 **Fecha de aceptación:** 19/02/2026 **Fecha de publicación:** 01/04/2026

RESUMEN

La creciente dependencia de sistemas automatizados en ámbitos como la salud, la justicia, las finanzas y la administración pública ha intensificado la necesidad de comprender cómo funcionan los modelos de inteligencia artificial (IA) que influyen en decisiones de alto impacto. La inteligencia artificial explicable (XAI) surge como una respuesta a la opacidad algorítmica, promoviendo mecanismos que permitan interpretar, auditar y justificar las predicciones o recomendaciones generadas por los sistemas inteligentes. Más allá de la eficiencia técnica, la explicabilidad se convierte en un principio ético y jurídico indispensable para garantizar la confianza pública, prevenir sesgos discriminatorios y asegurar la protección de derechos fundamentales. En este contexto, se observa un avance significativo en técnicas de interpretabilidad, desde modelos intrínsecamente transparentes hasta herramientas poshoc que facilitan la visualización de patrones internos. Sin embargo, persisten desafíos relevantes como la estandarización de métricas, la tensión entre rendimiento y explicabilidad, y la necesidad de marcos regulatorios sólidos que definan responsabilidades y límites de uso. La consolidación de la XAI representa un punto de inflexión para la gobernanza algorítmica, permitiendo una interacción más segura, equitativa y comprensible entre humanos y sistemas automatizados.

Palabras clave:

Inteligencia artificial explicable, transparencia algorítmica, decisiones automatizadas, ética digital, gobernanza tecnológica.

ABSTRACT

The growing reliance on automated systems in fields such as healthcare, justice, finance, and public administration has intensified the need to understand how artificial intelligence (AI) models influence high-impact decisions. Explainable AI (XAI) responds to algorithmic opacity by promoting mechanisms that enable the interpretation, auditing, and justification of intelligent systems' predictions or recommendations. Beyond technical performance, explainability becomes an essential ethical and legal principle to foster public trust, mitigate discriminatory biases, and protect fundamental rights. In this context, notable progress has been made in interpretability techniques, ranging from intrinsically transparent models to post-hoc tools that visualize internal patterns. Nevertheless, significant challenges persist, including the lack of standardized metrics, the trade-off between performance and explainability, and the pressing need for robust regulatory frameworks defining responsibilities and usage boundaries. The consolidation of XAI represents a key inflection point for algorithmic governance, enabling safer, fairer,

and more understandable interactions between humans and automated systems.

Keywords:

Explainable artificial intelligence, algorithmic transparency, automated decisions, digital ethics, technological governance.

INTRODUCCIÓN

La Inteligencia Artificial Explicable (XAI, por sus siglas en inglés) constituye hoy una de las áreas más dinámicas y estratégicas en el desarrollo de tecnologías digitales avanzadas. En un contexto socioeconómico marcado por la creciente automatización de procesos y decisiones, la IA ha dejado de ser una herramienta experimental para convertirse en un componente esencial de sectores como la salud, la justicia, las finanzas, la seguridad y la administración pública (Hulsen, 2023; Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura 2024).

Sin embargo, la transición desde modelos simples y fácilmente interpretables, como los árboles de decisión o las regresiones lineales, hacia arquitecturas más potentes y complejas, principalmente en el ámbito del aprendizaje profundo, ha generado un efecto colateral preocupante: la opacidad en el razonamiento interno de los sistemas (Corvalán, 2018). Este fenómeno, ampliamente conocido como el "problema de la caja negra", plantea un desafío fundamental para la confianza y legitimidad de las decisiones automatizadas (Floridi et al., 2018).

En sus primeras etapas, la inteligencia artificial se enfocaba en reglas explícitas y sistemas expertos cuya lógica era transparente y explícitamente programada. Estos modelos permitían rastrear y explicar cada decisión a partir de inferencias lógicas detectables por un humano (Molnar, 2020). Sin embargo, la explosión de datos masivos y el incremento de poder computacional motivaron el desarrollo de modelos estadísticos de alta complejidad, como las redes neuronales profundas, que optimizan su desempeño a partir de patrones no evidentes y distribuciones de parámetros en capas de procesamiento. Esta ganancia en precisión y adaptabilidad vino acompañada de una pérdida significativa en la capacidad de interpretar los procesos internos de decisión (Lundberg & Lee, 2017; Ribeiro et al., 2016).

La comunidad científica comenzó a identificar que esta falta de explicabilidad no es solo un inconveniente técnico, sino un riesgo para derechos fundamentales, ya que un sistema que no puede ser auditado o comprendido dificulta la rendición de cuentas y el control ciudadano (Guidotti et al., 2018). De ahí surge la XAI como una disciplina destinada a desarrollar técnicas, marcos conceptuales y normativas que permitan comprender cómo y por qué una IA produce una decisión determinada.

La necesidad de la XAI se articula en tres planos:

1. Plano técnico: mejora la comprensión, depuración y optimización de modelos complejos.

2. Plano social y ético: garantiza que las decisiones sean justas, no discriminatorias y coherentes con valores humanos (Floridi et al., 2018).

3. Plano normativo: facilita el cumplimiento de regulaciones como el Reglamento General de Protección de Datos (GDPR) en Europa o iniciativas regulatorias latinoamericanas (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, 2024).

Diversos organismos internacionales, han emitido directrices que enfatizan la necesidad de sistemas con explicaciones claras y comprensibles tanto para expertos como para usuarios finales. Estas directrices aumentan la urgencia de la investigación en XAI, ya que el déficit de explicabilidad puede derivar en rechazo social, conflictos legales y sanciones regulatorias (Alayón Miranda, 2024; García-Vera & Juca-Maldonado, 2026).

En los últimos diez años, se han propuesto metodologías post-hoc como LIME (Ribeiro et al., 2016) y SHAP (Lundberg & Lee, 2017), que generan explicaciones sobre modelos ya entrenados mediante aproximaciones locales o asignación de importancia de variables. Aunque estas herramientas marcan un avance significativo, la literatura muestra que presentan limitaciones: las explicaciones pueden ser inestables, variar ante pequeñas perturbaciones en los datos y, en algunos casos, no reflejar fielmente los procesos internos del modelo (Molnar, 2020). Por otro lado, se desarrollan enfoques intrínsecamente interpretables, como redes neuronales explicables o modelos de reglas, que sacrifican parte de la exactitud a cambio de completa legibilidad.

El panorama actual también muestra una fragmentación metodológica: no existe aún un marco de evaluación estandarizado que permita comparar de forma objetiva el nivel de explicabilidad alcanzado por distintos modelos (Guidotti et al., 2018). Asimismo, trabajos como el de Floridi et al. (2018) insisten en que la explicabilidad no debe analizarse solo como un reto técnico, sino como un imperativo ético aplicable desde la definición misma del caso de uso de IA.

El presente trabajo se centra en el análisis de los retos y avances actuales de la XAI para la transparencia en sistemas de toma de decisiones automatizadas. El alcance incluye la revisión de métodos técnicos, evaluación crítica de marcos regulatorios y discusión de dimensiones éticas y sociales implicadas. No se trata únicamente de describir herramientas existentes, sino de examinar cómo estas se integran en escenarios reales de alto impacto social. El objetivo principal es identificar brechas, tensiones y

oportunidades que permitan orientar el desarrollo hacia sistemas de IA confiables, auditables y alineados con valores democráticos y principios de equidad.

Esta investigación se desarrolla en tiempo presente, revisando los aportes más influyentes de la literatura y normativas internacionales, con el fin de proponer un marco de comprensión que sirva de base para investigaciones futuras y para la formulación de políticas públicas inclusivas. A lo largo del documento se abordarán los fundamentos conceptuales, los retos técnicos y éticos, y los avances recientes que están configurando la agenda de investigación en XAI.

Uno de los desarrollos más emblemáticos a nivel global es IBM Watson Health, un sistema de IA enfocado en análisis clínico y diagnóstico asistido. Watson fue diseñado para ayudar a oncólogos y médicos en el tratamiento de cáncer, ofreciendo recomendaciones basadas en grandes repositorios médicos, historias clínicas y literatura científica.

Sin embargo, la aceptación real del sistema por parte de los profesionales clínicos dependió de la capacidad de Watson para explicar sus recomendaciones de manera comprensible y auditada. Investigaciones muestran que los médicos demandan justificaciones claras para confiar plenamente en los tratamientos sugeridos, especialmente en casos que involucran vidas humanas (Hulsen, 2023). IBM debió incorporar funciones de XAI que permitieran rastrear las evidencias citadas por el sistema y que los médicos pudieran verificar el razonamiento seguido por la inteligencia artificial. Estas mejoras favorecieron la integración de Watson en hospitales, pero también evidenciaron limitaciones cuando los algoritmos ofrecían recomendaciones basadas en correlaciones poco intuitivas para los especialistas (Floridi et al., 2018).

El caso Watson pone de relieve que, en contextos médicos, la explicabilidad es tanto una demanda ética como operacional, y su ausencia puede limitar la adopción de sistemas automáticos con potencial transformador. Se concluye que la XAI no solo requiere herramientas técnicas, sino también procesos de diseño centrados en el usuario final y marcos normativos que aseguren la validación de resultados clínicos por expertos humanos.

El sistema COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) es una herramienta de IA utilizada en tribunales estadounidenses para evaluar el riesgo de reincidencia criminal. COMPAS analiza variables sociodemográficas y antecedentes penales para recomendar niveles de riesgo en la gestión judicial de individuos.

Desde su implementación, COMPAS ha generado controversias sobre transparencia, equidad y sesgo algorítmico. Investigaciones críticas revelan que el sistema es percibido como una "caja negra" por jueces y abogados, y que puede favorecer resultados discriminatorios contra minorías (Angwin et al., 2016). La ausencia de explicaciones claras sobre cómo se ponderan las variables y por qué se asigna cierto riesgo a los acusados limitó la confianza en el sistema y motivó debates jurídicos y mediáticos. Ante las demandas de explicabilidad y rendición de cuentas, desarrolladores y reguladores debieron incorporar mecanismos de auditoría externa y reportes interpretables que permitieran a las partes afectadas cuestionar la lógica utilizada (Floridi et al., 2018; Guidotti et al., 2018).

El caso COMPAS ilustra que la XAI es un elemento central para garantizar derechos fundamentales en la toma de decisiones automatizadas, especialmente cuando estas afectan la libertad individual. Las críticas internacionales contra la opacidad del sistema han impulsado reformas metodológicas y propuestas legales que exigen explicaciones auditables, justificación del razonamiento, y validación de resultados por órganos independientes.

En la Unión Europea, la aplicación de la XAI en el sector financiero se apoya en el marco normativo del Reglamento General de Protección de Datos (GDPR) y en propuestas de la Regulación de IA en curso (European Commission, 2021; Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, 2024). Los bancos y entidades crediticias utilizan modelos predictivos complejos para determinar la elegibilidad y el perfil de riesgo de los solicitantes de crédito, lo que históricamente ha quedado fuera del alcance de escrutinio público y de los propios clientes.

La normativa europea exige que las decisiones automatizadas sean explicables y que los afectados puedan solicitar información sobre las variables que influyeron en el resultado. Esto ha requerido la adopción de herramientas XAI que permitan mostrar, en lenguaje accesible, por qué un cliente recibe cierto puntaje, facilitando que los solicitantes cuestionen y rectifiquen posibles errores o sesgos algorítmicos. Varios estudios reportan casos en los que la falta de explicabilidad derivó en rechazo de solicitudes legítimas y sanciones regulatorias a entidades financieras (Guidotti et al., 2018; Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, 2024).

Gracias a la presión normativa y social, los bancos ahora implementan reportes interpretables y métodos post-hoc como SHAP, que permiten a los clientes visualizar el impacto relativo de cada factor en la asignación de crédito. Este caso demuestra que la XAI potencia la transparencia institucional, protege derechos económicos y promueve una mayor equidad en el acceso a servicios financieros.

Los ejemplos de IBM Watson Health, COMPAS y el sistema bancario europeo evidencian que la XAI no es solo un avance tecnológico, sino una necesidad estructural en la gobernanza de sistemas automatizados. La ausencia o presencia de explicabilidad impacta directamente la legitimidad, confianza y aceptación social de estos sistemas, así como la protección efectiva de derechos. Los

casos analizados confirman la problemática descrita en la introducción:

En cada contexto, la interpretabilidad y explicabilidad constituyen exigencias técnicas, éticas y normativas irrenunciables.

Los retos persisten en la estabilidad, fidelidad y aplicabilidad de las explicaciones ofrecidas; existen avances metodológicos importantes, pero todavía se requiere mayor armonización y regulación internacional.

La continuidad entre teoría y práctica refuerza el objetivo de esta investigación: identificar los desafíos y oportunidades de la XAI para consolidar sistemas de decisión automatizada que sean transparentes, auditables y socialmente confiables.

MATERIALES Y MÉTODOS

El enfoque de la presente investigación es de tipo cualitativo y descriptivo, sustentado en una revisión sistemática y exhaustiva de literatura académica, documentos doctrinarios y marcos regulatorios pertinentes al estudio de la Inteligencia Artificial Explicable (XAI). Esta aproximación metodológica permite examinar con profundidad los fundamentos conceptuales, los desafíos tecnológicos y éticos, así como los avances recientes en torno a la transparencia algorítmica y la rendición de cuentas en sistemas de toma de decisiones automatizadas. Además, posibilita identificar tendencias globales, evaluar experiencias comparadas y comprender la evolución normativa que se ha generado en torno a la XAI. Todo este proceso se desarrolla siguiendo los criterios establecidos por Hernández Sampieri et al. (2014), quienes plantean lineamientos claros para investigaciones basadas en fuentes documentales y el análisis crítico de información especializada.

La investigación adopta como técnica principal el análisis de contenido, una herramienta metodológica que facilita la selección sistemática, categorización y evaluación de artículos científicos, estudios de caso y reportes institucionales que aportan evidencia relevante para el estudio. Se incluyen casos emblemáticos como IBM Watson Health, el sistema COMPAS utilizado en el ámbito judicial estadounidense y diversas aplicaciones algorítmicas implementadas en entidades financieras europeas. Estos ejemplos permiten ilustrar tanto las limitaciones prácticas de los sistemas opacos como las innovaciones orientadas a mejorar su explicabilidad. El proceso de análisis se desarrolla de forma rigurosa bajo los parámetros metodológicos propuestos por Hernández Sampieri et al. (2014), garantizando la consistencia en la recolección, la interpretación y la validación de las evidencias.

En cuanto al método empleado, la investigación se sustenta en un enfoque inductivo, dado que parte de la observación detallada y el análisis de casos concretos relacionados con la Inteligencia Artificial Explicable en sectores críticos como salud, justicia y finanzas. A partir de la interpretación de estas experiencias reales, se formulan generalizaciones, propuestas analíticas y conclusiones que permiten comprender de manera más amplia los retos estructurales y los avances logrados en el ámbito de la XAI. Hernández Sampieri et al. (2014) destacan que el método inductivo permite construir conocimiento sólido sustentado en la evidencia empírica obtenida directamente de contextos reales de aplicación, fortaleciendo así la validez de las inferencias y la pertinencia de los hallazgos dentro del campo de la investigación científica.

DESARROLLO

La necesidad de proporcionar a las aplicaciones de inteligencia artificial transparencia, trazabilidad y auditabilidad ha impulsado el surgimiento de la XAI. Si bien se reconoce la relevancia de la XAI, diversos autores difieren en cuanto a su implementación y alcance.

Molnar (2020) señala que, en la fase inicial de la inteligencia artificial se basaba en sistemas especializados que facilitaban la comprensión en cada decisión. Sin embargo, la evolución hacia modelos más sofisticados, como las redes neuronales profundas, introdujo mayores niveles de opacidad. Asimismo, Lundberg & Lee (2017); y Ribeiro et al. (2016) advierten que dicha complejidad reduce la interpretabilidad y limita la precisión en la explicación de los resultados.

Guidotti et al. (2018) añaden que la falta de explicabilidad representa un riesgo para los derechos fundamentales, dado que limita la rendición de cuentas y el control ciudadano. Ante este panorama surge la XAI, disciplina que busca desarrollar enfoques y normas para comprender las decisiones de la IA.

Desde una perspectiva práctica, Hulsen (2023) destaca que IBM Waltson Health, un sistema de IA, orientado al diagnóstico clínico asistido, enfrentó demandas de médicos que exigían explicaciones claras para confiar en sus tratamientos. De manera similar, Anwing (2016) señala que Compas, herramienta utilizada en tribunales estadounidenses para evaluar el riesgo de reincidencia criminal, ha sido cuestionada al ser percibido como “caja negra” y por su potencial sesgo discriminatorio hacia minorías.

Esta situación también se refleja en el sector financiero, regulado por el Reglamento General de Protección de Datos, donde, según la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (2024), los bancos y entidades crediticias usan IA para evaluar la elegibilidad de solicitantes mediante modelos predictivos.

En el sector financiero, la importancia de XAI se ejemplifica en la aplicación del Reglamento General de Protección de Datos (GDPR) en Europa, donde bancos y entidades crediticias están obligados a entregar explicaciones comprensibles sobre las decisiones que afectan a los solicitantes, promoviendo así la equidad, la transparencia y el

derecho a la rectificación frente a decisiones automatizadas injustas.

Los casos analizados evidencian de manera consistente que la Inteligencia Artificial Explicable (XAI) constituye un elemento esencial para generar confianza pública, garantizar la protección de los derechos fundamentales y asegurar el cumplimiento de los marcos normativos que regulan el uso de tecnologías automatizadas. En el caso de IBM Watson Health, la ausencia inicial de explicaciones claras, comprensibles y verificables sobre el funcionamiento de sus recomendaciones clínicas limitó significativamente su aceptación por parte del personal médico. Esto demuestra que el alto rendimiento técnico o la capacidad predictiva avanzada no son suficientes para promover la adopción efectiva en entornos críticos como la salud, donde la confianza profesional depende de la transparencia, la trazabilidad y la posibilidad de interpretar el razonamiento del sistema.

De manera similar, el sistema COMPAS, utilizado en procesos judiciales para evaluar riesgos de reincidencia, evidenció profundas complicaciones éticas al operar como una "caja negra", dificultando la comprensión de los criterios que fundamentaban sus puntuaciones. Esta falta de explicabilidad comprometió garantías esenciales como la equidad, la justicia y la no discriminación, demostrando que la opacidad algorítmica puede derivar en decisiones automatizadas que afectan directamente la vida, los derechos y la dignidad de las personas. En este sentido, la XAI no solo busca mejorar la transparencia ética, sino también garantizar que los sistemas automatizados sean comprensibles, auditables, verificables y confiables para todos los actores involucrados (Ibarra-Pincay & Alcívar Cevallos, 2024).

Por otro lado, desde una perspectiva ética más amplia, la "caja negra" inherente a muchos modelos complejos puede perpetuar sesgos estructurales y discriminación algorítmica cuando no existen mecanismos de revisión independiente, auditorías continuas o explicaciones accesibles que permitan justificar el razonamiento automatizado. Esto representa un riesgo significativo, especialmente en áreas como justicia, salud y finanzas, donde las decisiones algorítmicas pueden tener consecuencias sociales profundas e irreversibles (McGrath & Jonker, 2024).

En este mismo orden de ideas, en el ámbito regulatorio se evidencia que la presión normativa —particularmente en Europa con el desarrollo del AI Act y otros estándares internacionales— ha impulsado la adopción de marcos legales que exigen explicabilidad, trazabilidad y responsabilidad en el diseño y uso de sistemas de IA. Estas exigencias no solo fomentan prácticas responsables, sino que también contribuyen a la armonización global de lineamientos éticos y técnicos, promoviendo un avance sostenido hacia una Inteligencia Artificial más justa, segura y alineada con los valores democráticos.

CONCLUSIONES

En respuesta al objetivo general planteado, la investigación identifica que los principales retos y oportunidades de la XAI orientados a fortalecer la transparencia en la toma de decisiones automatizadas se concentran en la necesidad de armonizar metodologías interpretables, desarrollar marcos regulatorios claros y promover la integración efectiva de técnicas explicativas en contextos reales de aplicación. Estos elementos permiten avanzar hacia la consolidación de modelos de IA que no solo sean técnicamente eficientes, sino también confiables, auditables y plenamente alineados con valores democráticos, principios de equidad y estándares éticos internacionales. Asimismo, se identifica que la consolidación de prácticas de XAI requiere de una articulación multidisciplinaria entre ingeniería, derecho, ética y ciencias sociales, a fin de asegurar soluciones integrales para la gobernanza algorítmica.

Por otro lado, se concluye que, a partir de los casos analizados, la ausencia de mecanismos de explicabilidad en los sistemas de IA genera importantes riesgos éticos y sociales, tales como la pérdida de confianza ciudadana, la reproducción de sesgos discriminatorios, la afectación de la autonomía humana y la vulneración de derechos fundamentales. Esta situación demuestra que la XAI es indispensable para cumplir con las exigencias normativas internacionales, aumentar la confiabilidad institucional y promover condiciones que favorezcan la legitimidad, la aceptabilidad pública y el uso responsable de tecnologías automatizadas en sectores de alta sensibilidad social.

Por último, cabe destacar que la implementación de la Inteligencia Artificial Explicable (XAI) es esencial para fortalecer la transparencia, la trazabilidad y la auditabilidad de los sistemas de toma de decisiones automatizadas. Su incorporación permite que usuarios, especialistas y entidades reguladoras comprendan, supervisen y cuestionen los resultados generados por los modelos, incluso en áreas críticas como la salud, la justicia y las finanzas. De esta manera, la XAI se posiciona como un componente indispensable para garantizar la rendición de cuentas, promover entornos tecnológicos confiables y consolidar una inteligencia artificial ética y socialmente responsable.

REFERENCIAS

Alayón Miranda, S. (2024). El problema de la interpretabilidad de la inteligencia artificial y su impacto en la administración pública. Revista Canaria de Administración Pública, 3, 175–202. https://revistacanarias.tirant.com/index.php/revista-canaria/article/view/42

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica, Online Edition.

Corvalán, J. G. (2018). Inteligencia artificial: retos, desafíos y oportunidades – Prometea: la primera inteligencia artificial de Latinoamérica al servicio de la Justicia. Revista de Investigações Constitucionais, 5(1), 295-316. https://doi.org/10.5380/rinc.v5i1.55334

European Commission. (2021). Proposal for a regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. European Commission. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52021PC0206

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People-an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

García-Vera, Y. S., & Juca-Maldonado, F. X. (2026). ***Inteligencia Artificial Aplicada a la Contabilidad: Flujos, buenas prácticas y control*.** Sophia Editions.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. ACM Computing Surveys, 51(5), 1-42. https://doi.org/10.1145/3236009

Hernández Sampieri, R., Fernández, C., & Baptista, P. (2014). Metodología de la investigación. 6ta. Ed. Mc. Graw Hill Eduaction, México.

Hulsen, T. (2023). Explainable Artificial Intelligence (XAI): Concepts and challenges in healthcare. AI (Basel, Switzerland), 4(3), 652–666. https://doi.org/10.3390/ai4030034

Ibarra-Pincay, M., & Alcívar-Cevallos, R. (2024). Tendencias de la Inteligencia Artificial Explicable en el Área de Psicología. Revista Científica INGENIAR: Ingeniería, Tecnología E Investigación, 7(13), 80-101. https://journalingeniar.org/index.php/ingeniar/article/view/162

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. 31st Conference on Neural Information Processing Systems (NIPS 2017). Long Beach, USA.

McGrath, A., & Jonker, A. (2024). What is AI interpretability? IBM Knowledge Center. https://www.ibm.com/es-es/think/topics/interpretability

Molnar, C. (2020). Interpretable Machine Learning. Lulu. com.

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (2024). Ética de la inteligencia artificial. UNESCO. https://www.unesco.org/es/artificial-intelligence/recommendation-ethics

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, United States.

Conflictos de interés:

Los autores declaran no tener conflictos de interés.

Contribución de los autores:

Génesis Karolina Robles-Zambrano, Ingrid Joselyne Diaz-Basurto, Deinier Ros-Álvarez, Verónica Zuleyka Ramírez-Bósquez: Concepción y diseño del estudio, adquisición de datos, análisis e interpretación, redacción del manuscrito, revisión crítica del contenido, análisis estadístico, supervisión general del estudio.

Declaración ética:

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