

Uso de la inteligencia artificial para fortalecer el control interno empresarial: una revisión sistemática de la literatura (2016–2026)

Artificial Intelligence for Strengthening Internal Control and Tax Management in Firms: A Systematic Literature Review (2016–2026)

Author 1: Vanessa Olimpia Rodríguez-Flores; e-mail: vanerf1387@gmail.com; Organization: Quality Assurance Department, Institutional evaluation and accreditation management, State University of Milagro, Milagro, Ecuador; ORCID: <https://orcid.org/0009-0004-0130-0176>

Author 2: Andrea Alejandra Cujilán-Guaman; e-mail: acujilang@unemi.edu.ec; Organization: Quality Assurance Department, Institutional evaluation and accreditation management, State University of Milagro, Milagro, Ecuador; ORCID: <https://orcid.org/0009-0002-7307-584X>

Resumen

El presente estudio analiza el papel de la inteligencia artificial (IA) en el fortalecimiento de los sistemas de control interno empresarial mediante una revisión sistemática de la literatura correspondiente al período 2016–2026. En un contexto caracterizado por la creciente complejidad de los entornos organizacionales, las empresas enfrentan desafíos significativos en materia de supervisión, gestión de riesgos y toma de decisiones. En este escenario, la IA emerge como una herramienta transformadora capaz de optimizar los mecanismos tradicionales de control a través del uso intensivo de datos.

La investigación se desarrolló bajo el protocolo PRISMA, utilizando la base de datos Scopus como fuente principal. Se seleccionaron 11 artículos relevantes tras la aplicación de criterios de inclusión y exclusión. El análisis se realizó mediante un enfoque cualitativo de contenido, lo que permitió identificar patrones, tendencias y vacíos en la literatura.

Los resultados evidencian que la IA contribuye al fortalecimiento del control interno en cuatro áreas principales: (i) mejora de los procesos de auditoría interna, (ii) detección de fraude y anomalías, (iii) fortalecimiento de la gestión de riesgos y (iv) transformación de los sistemas de control hacia modelos predictivos y proactivos. En particular, el uso de machine learning e inteligencia artificial explicable permite mejorar la precisión en la toma de decisiones.

No obstante, también se identifican desafíos relevantes, como la opacidad algorítmica, los riesgos éticos y la necesidad de marcos de gobernanza adecuados. En conjunto, se concluye que la IA no sustituye el control interno, sino que lo transforma en un modelo híbrido.

Palabras clave: Inteligencia artificial; Control interno; Auditoría interna; Gestión de riesgos; Gobernanza corporativa

Abstract

This study analyzes the role of artificial intelligence (AI) in strengthening internal control systems in organizations through a systematic literature review covering the period 2016–2026. As

business environments become increasingly complex, organizations face growing challenges in ensuring effective monitoring, risk management, and decision-making processes. In this context, AI has emerged as a transformative tool capable of enhancing traditional control mechanisms through data-driven approaches.

The study follows the PRISMA methodology and uses Scopus as the primary data source. A total of 11 relevant articles were selected after applying inclusion and exclusion criteria. The analysis was conducted using a qualitative content approach to identify patterns, trends, and research gaps.

The findings reveal that AI significantly contributes to internal control through four main areas: (i) improvement of internal audit processes, (ii) enhanced fraud detection and anomaly identification, (iii) strengthening of risk management capabilities, and (iv) transformation of organizational control systems toward predictive and proactive models. In particular, machine learning and explainable AI techniques enable real-time monitoring and increase decision accuracy.

However, the results also highlight critical challenges, including algorithmic opacity, ethical concerns, and the need for appropriate governance frameworks. The successful implementation of AI depends not only on technological capabilities but also on organizational readiness and regulatory alignment.

Overall, this study concludes that AI does not replace traditional control systems but transforms them into hybrid models combining human judgment and intelligent systems. The research contributes to the literature by providing a structured synthesis and identifying future research directions.

Keywords: Artificial Intelligence; Internal Control; Internal Audit; Risk Management; Corporate Governance

1. Introducción

En las últimas décadas, la creciente complejidad de los entornos empresariales ha intensificado la necesidad de fortalecer los sistemas de control interno como mecanismo fundamental para garantizar la transparencia, la eficiencia operativa y la gestión adecuada de riesgos organizacionales. En este contexto, el control interno ha evolucionado desde enfoques tradicionales centrados en la supervisión manual hacia modelos más dinámicos basados en el análisis de información y la toma de decisiones estratégicas (Schuett, 2025).

Paralelamente, la inteligencia artificial (IA) se ha consolidado como una de las tecnologías más disruptivas en el ámbito organizacional, debido a su capacidad para procesar grandes volúmenes de datos, identificar patrones complejos y generar predicciones con altos niveles de precisión. Estas capacidades han abierto nuevas oportunidades para su integración en sistemas de auditoría, gestión de riesgos y control interno, permitiendo a las organizaciones mejorar sus mecanismos de supervisión y respuesta ante eventos adversos (Liu et al., 2026).

Diversos estudios han evidenciado que la IA puede contribuir significativamente a la mejora de los procesos de control interno. Por ejemplo, Caruso et al. (2023) destacan que la combinación de inteligencia artificial y herramientas de business intelligence permite reducir la sobrecarga de indicadores y mejorar el monitoreo de procesos, facilitando la detección de irregularidades y desviaciones. De manera similar, Lu y Wang (2025) demuestran que el uso de algoritmos de

aprendizaje profundo en auditoría interna mejora la identificación de riesgos y la eficiencia en el análisis de información financiera.

En el ámbito de la auditoría, la incorporación de herramientas basadas en IA, como los modelos de lenguaje avanzados, ha permitido mejorar la calidad del análisis y la interpretación de información compleja. Bouziane (2025) señala que el uso de tecnologías como ChatGPT contribuye a mejorar la calidad de las auditorías internas, facilitando la comprensión de información técnica y fortaleciendo el proceso de toma de decisiones. Asimismo, Dong et al. (2025) evidencian que la aplicación de técnicas de machine learning mejora la detección de fraude contable, lo que representa un avance significativo en la integridad de los sistemas financieros.

No obstante, la adopción de la inteligencia artificial en el control interno también plantea desafíos importantes. Entre ellos, destacan la necesidad de garantizar la transparencia en los modelos algorítmicos, la gestión ética de los datos y la definición de marcos regulatorios adecuados que permitan supervisar el uso de estas tecnologías (Adewumi, 2025; Najm & Amaral, 2025). En este sentido, la literatura sugiere que la implementación de la IA debe ir acompañada de estructuras de gobernanza que aseguren su uso responsable y alineado con los objetivos organizacionales.

A pesar del creciente interés en la temática, la literatura existente presenta un carácter fragmentado, con un predominio de estudios centrados en aplicaciones técnicas de la inteligencia artificial y una menor atención a su impacto en los sistemas de control interno empresarial. Esta brecha limita la comprensión integral del papel de la IA en la gestión organizacional y evidencia la necesidad de sistematizar la evidencia científica disponible.

Por lo tanto, el objetivo de este estudio es analizar la producción científica sobre el uso de la inteligencia artificial en el fortalecimiento del control interno empresarial durante el período 2016–2026, identificando tendencias, aplicaciones, aportes teóricos y vacíos de investigación.

2. Metodología

2.1. Diseño de la investigación

El presente estudio se desarrolló bajo el enfoque de revisión sistemática de la literatura (Systematic Literature Review, SLR), con el propósito de identificar, analizar y sintetizar de manera rigurosa la evidencia científica existente sobre el uso de la inteligencia artificial (IA) en el fortalecimiento del control interno en el ámbito empresarial durante el período 2016–2026.

La investigación se estructuró siguiendo las directrices del protocolo PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), el cual permite garantizar la transparencia, trazabilidad y replicabilidad del proceso de selección y análisis de los estudios. Este enfoque metodológico resulta particularmente pertinente para examinar fenómenos complejos y multidimensionales, como la integración de tecnologías inteligentes en los sistemas de control organizacional y cumplimiento fiscal.

Asimismo, el estudio adoptó un enfoque cualitativo de carácter interpretativo, sustentado en el análisis de contenido, con el fin de identificar patrones, tendencias investigativas y vacíos en la literatura científica.

2.2. Fuente de datos

La recopilación de la información se realizó a través de la base de datos Scopus, seleccionada por su amplia cobertura de revistas científicas indexadas de alto impacto en áreas como contabilidad, auditoría, administración, sistemas de información y ciencias computacionales.

La elección de Scopus se fundamenta en la calidad de sus procesos de indexación, la confiabilidad de sus registros bibliográficos y la disponibilidad de metadatos estructurados, lo que facilita la reproducibilidad del estudio. No obstante, se reconoce que el uso de una única base de datos puede limitar la inclusión de investigaciones relevantes indexadas en otras fuentes, lo cual constituye una limitación del presente trabajo.

2.3. Estrategia de búsqueda

Se diseñó una estrategia de búsqueda estructurada mediante el uso de operadores booleanos (AND, OR), orientada a capturar estudios que integren tres dimensiones principales: inteligencia artificial, control interno/gobernanza organizacional y gestión tributaria.

La ecuación de búsqueda fue aplicada en los campos de título, resumen y palabras clave (TITLE-ABS-KEY), con el objetivo de garantizar la precisión y pertinencia de los resultados. La expresión utilizada fue la siguiente:

La ecuación utilizada fue la siguiente: TITLE-ABS-KEY ("artificial intelligence" OR "machine learning" OR "deep learning" AND "internal control" OR "internal audit" OR "corporate governance").

Adicionalmente, se establecieron los siguientes filtros:

- Periodo de publicación: 2016–2026
- Idiomas: inglés y español
- Tipo de documento: artículos científicos

Esta estrategia permitió delimitar el corpus de estudio a investigaciones relevantes en el contexto empresarial y con implicaciones en la gestión financiera y tributaria.

2.4 Criterios de inclusión y exclusión

Con el fin de garantizar la calidad metodológica y la pertinencia temática de los estudios seleccionados, se definieron criterios explícitos de inclusión y exclusión.

Criterios de inclusión:

- Artículos científicos publicados en revistas indexadas
- Estudios que aborden la aplicación de la inteligencia artificial en control interno, auditoría o gestión tributaria
- Investigaciones desarrolladas en contextos organizacionales o empresariales
- Publicaciones en idioma inglés o español
- Periodo comprendido entre 2016 y 2026

Criterios de exclusión:

- Documentos no arbitrados (conferencias, libros, capítulos)
- Estudios de carácter exclusivamente técnico sin implicaciones en la gestión organizacional
- Investigaciones fuera del ámbito empresarial
- Artículos duplicados
- Estudios sin acceso a texto completo

2.5 Proceso de selección de estudios (PRISMA)

El proceso de selección de los estudios se desarrolló en cuatro fases, conforme a las directrices del protocolo PRISMA: identificación, cribado, elegibilidad e inclusión (Ver Figura 1).

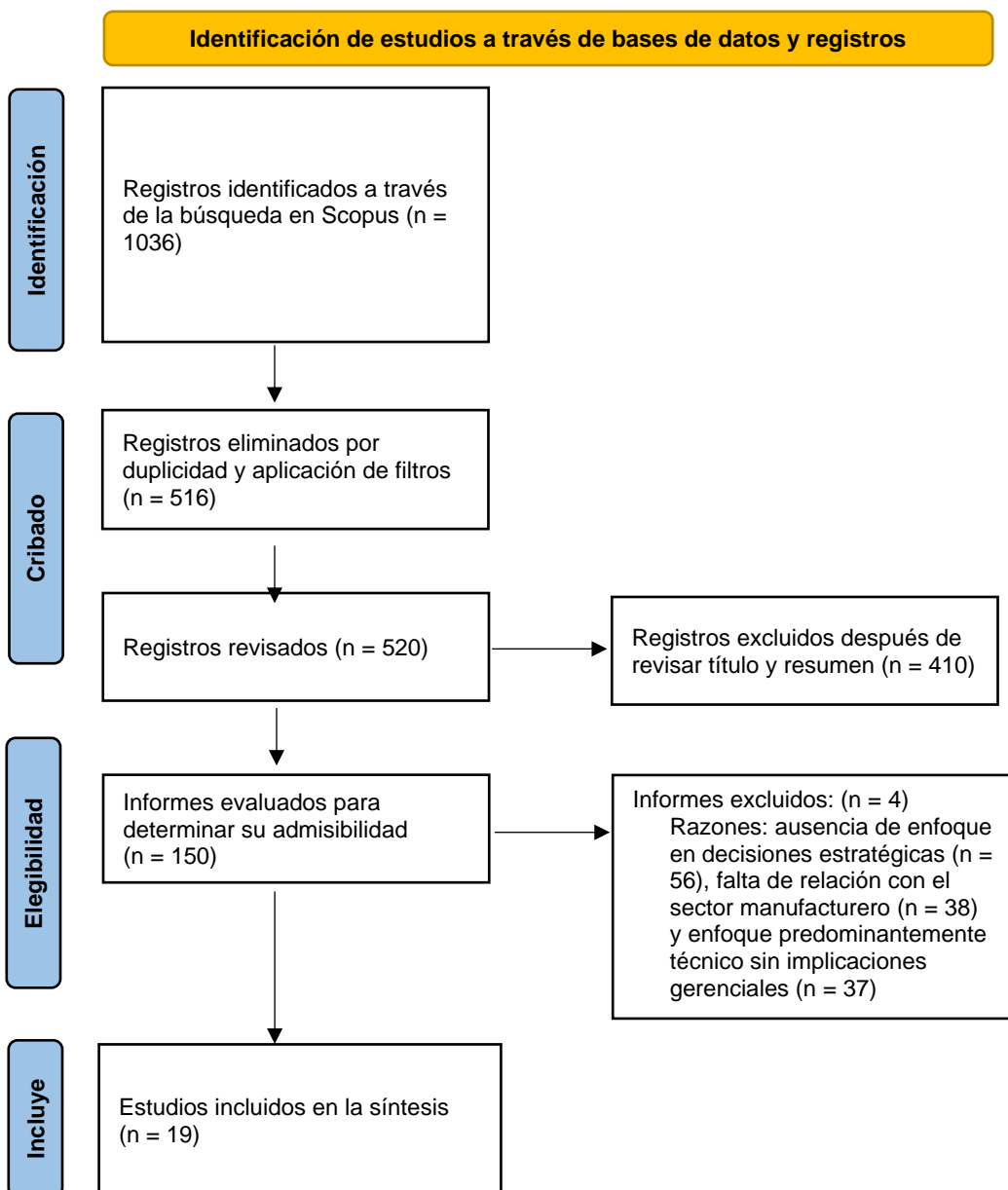


Figura 1 Diagrama de flujo PRISMA del proceso de selección de estudios

2.6 Extracción de datos

Para cada uno de los estudios incluidos, se elaboró una matriz de extracción de datos que permitió sistematizar la información relevante de manera estructurada. Las variables consideradas fueron:

- Autor(es) y año de publicación
- País o contexto de estudio
- Tipo de tecnología de inteligencia artificial utilizada
- Área de aplicación (control interno o gestión tributaria)
- Tipo de decisión (estratégica, táctica u operativa)
- Enfoque metodológico
- Principales hallazgos

Esta sistematización facilitó el análisis comparativo de la evidencia científica y la identificación de patrones recurrentes.

2.7 Análisis de la información

El análisis de la información se llevó a cabo mediante un enfoque de análisis cualitativo de contenido, el cual permitió clasificar los estudios en categorías temáticas emergentes. Este proceso se desarrolló de manera iterativa, asegurando la coherencia y consistencia en la interpretación de los datos.

Las principales categorías identificadas fueron:

- Uso de la inteligencia artificial en sistemas de control interno
- Aplicaciones de la inteligencia artificial en la gestión tributaria
- Transformación de los roles financieros y contables
- Desafíos de gobernanza, ética y regulación

Este enfoque permitió no solo describir la literatura existente, sino también generar una comprensión integrada del fenómeno estudiado.

2.8 Consideraciones de rigor metodológico

Con el objetivo de garantizar la calidad y validez del estudio, se adoptaron diversas estrategias metodológicas, entre las que destacan:

- Definición explícita de criterios de inclusión y exclusión
- Uso de una estrategia de búsqueda estructurada y replicable
- Aplicación del protocolo PRISMA
- Documentación detallada del proceso de selección
- Sistematización de la información mediante una matriz de extracción

Estas acciones contribuyen a fortalecer la transparencia, confiabilidad y reproducibilidad del estudio, alineándolo con los estándares internacionales de investigación científica.

3. Resultados

Los resultados de la revisión sistemática evidencian que la literatura sobre inteligencia artificial aplicada al control interno empresarial, aunque emergente, presenta un crecimiento sostenido en el período reciente, con un énfasis particular en aplicaciones orientadas a la auditoría interna, la detección de riesgos y la gobernanza organizacional.

Tabla 1. Artículos seleccionados

Autor (Año)	Contexto	IA utilizada	Área de aplicación	Tipo de decisión	Método	Hallazgo principal
Caruso et al. (2023)	Sector público	AI + Big Data	Control interno / monitoreo	Operativa	Design Science	IA reduce sobrecarga de indicadores y mejora control
Lu & Wang (2025)	Empresarial	Deep Learning (BPNN)	Auditoría interna / riesgos	Operativa	Empírico	IA mejora identificación de riesgos
Bouziane (2025)	Empresarial	ChatGPT	Auditoría interna	Estratégica	Empírico	IA mejora calidad del proceso de auditoría
Liu et al. (2026)	Bancario	XAI + BI	Control interno	Estratégica	Empírico	IA mejora predicción y transparencia del control
Schuett (2025)	Organizacional	AI governance	Gestión de riesgos	Estratégica	Conceptual	Modelo "3 líneas de defensa" aplicable a IA
Chen et al. (2025)	Empresarial	Machine Learning	Gobernanza / control interno	Estratégica	Empírico	IA mejora gobernanza mediante control interno
Dong et al. (2025)	Empresarial	ML (RUSBoost)	Fraude contable	Operativa	Empírico	IA mejora detección de fraude
Ma et al. (2024)	Empresarial	Machine Learning	Calidad de control interno	Estratégica	Empírico	Transformación digital mejora control interno
Castillo-Oliva et al. (2025)	Organizacional	Redes neuronales	Gestión financiera	Estratégica	Revisión	IA mejora control presupuestario
Adewumi (2025)	Global	AI systems	Gobernanza	Estratégica	Conceptual	Necesidad de regulación global de IA
Najm & Amaral (2025)	Organizacional	AI ethics frameworks	Gobernanza	Estratégica	Conceptual	IA requiere marcos éticos y de control

A partir del análisis de los estudios seleccionados, se identificaron cuatro líneas temáticas principales: (i) inteligencia artificial aplicada a la auditoría interna, (ii) detección de fraude y gestión de riesgos, (iii) evaluación y mejora del control interno, y (iv) gobernanza y regulación de la inteligencia artificial.

En primer lugar, la aplicación de la inteligencia artificial en auditoría interna constituye una de las áreas más desarrolladas. Estudios como los de Lu y Wang (2025) y Bouziane (2025) evidencian que el uso de algoritmos de aprendizaje profundo y herramientas como ChatGPT permite mejorar significativamente la calidad de los procesos de auditoría, especialmente en la identificación de riesgos y en el análisis de información compleja. Estos avances permiten pasar de auditorías tradicionales a enfoques más automatizados y analíticos.

En segundo lugar, la detección de fraude y gestión de riesgos emerge como un eje central del uso de la IA en el control interno. Investigaciones como la de Caruso et al. (2023) muestran que la integración de inteligencia artificial con herramientas de business intelligence permite detectar anomalías, fraudes y desviaciones en procesos organizacionales. De manera complementaria, Dong et al. (2025) demuestran que el uso de modelos de machine learning mejora la precisión en la detección de fraude contable, lo que fortalece la integridad de los sistemas financieros.

En tercer lugar, se identifica un conjunto de estudios centrados en la evaluación y mejora del control interno, donde la IA se utiliza para transformar la forma en que se mide y supervisa el desempeño organizacional. Liu et al. (2026) proponen sistemas basados en inteligencia artificial explicable (XAI) que permiten no solo predecir riesgos, sino también mejorar la transparencia en la toma de decisiones. Asimismo, Ma et al. (2024) evidencian que la transformación digital, apoyada en tecnologías de inteligencia artificial, contribuye a mejorar la calidad del control interno en las organizaciones.

Finalmente, la literatura resalta la importancia de la gobernanza y regulación de la inteligencia artificial. Estudios como los de Schuett (2025), Adewumi (2025) y Najm y Amaral (2025) enfatizan que la implementación de la IA en sistemas de control requiere marcos regulatorios y éticos sólidos. En particular, el modelo de las “tres líneas de defensa” se presenta como un enfoque relevante para integrar la gestión de riesgos asociados a la inteligencia artificial en el control organizacional.

En conjunto, los resultados evidencian que la inteligencia artificial se posiciona como un componente clave en el fortalecimiento del control interno empresarial, no solo desde una perspectiva operativa, sino también estratégica y de gobernanza.

4. Discusión

Los hallazgos de esta revisión sistemática permiten evidenciar que la inteligencia artificial está transformando de manera significativa los sistemas de control interno empresarial, aunque su adopción aún se encuentra en una fase de consolidación conceptual y empírica. A diferencia de otros campos donde la IA ha alcanzado una madurez más avanzada, la literatura analizada muestra que su integración en el control interno sigue siendo parcial y altamente dependiente del contexto organizacional.

Desde una perspectiva instrumental, los estudios revisados coinciden en que la IA mejora la eficiencia de los procesos de auditoría y control mediante la automatización del análisis de datos

y la detección de anomalías. En este sentido, herramientas basadas en machine learning y deep learning permiten procesar grandes volúmenes de información en tiempo real, reduciendo la dependencia de métodos manuales y aumentando la precisión en la identificación de riesgos. Este enfoque resulta particularmente evidente en los trabajos de Lu y Wang (2025) y Dong et al. (2025), donde la IA se utiliza como un mecanismo para fortalecer la detección de fraude y la evaluación de riesgos.

No obstante, una lectura más profunda de la evidencia sugiere que el impacto de la inteligencia artificial va más allá de la eficiencia operativa. La integración de sistemas avanzados como los propuestos por Liu et al. (2026) indica una transición hacia modelos de control interno basados en inteligencia aumentada, donde la toma de decisiones se apoya en sistemas predictivos y explicables. Este cambio implica una transformación estructural del control interno, que pasa de ser un sistema reactivo a uno proactivo y orientado a la anticipación de riesgos.

En este contexto, los resultados pueden interpretarse a la luz de la teoría de capacidades dinámicas, en la medida en que las organizaciones que logran integrar la inteligencia artificial en sus sistemas de control desarrollan habilidades superiores para adaptarse a entornos inciertos. La capacidad de procesar información, identificar patrones y generar respuestas rápidas se convierte en un recurso estratégico que fortalece la resiliencia organizacional.

Sin embargo, la literatura también evidencia tensiones importantes en torno a la gobernanza de la inteligencia artificial. Estudios como los de Schuett (2025) y Adewumi (2025) advierten sobre los riesgos asociados a la opacidad algorítmica, la falta de transparencia y la ausencia de marcos regulatorios adecuados. Estos desafíos son particularmente críticos en el ámbito del control interno, donde la confiabilidad y la trazabilidad de la información son fundamentales para garantizar la integridad organizacional.

Asimismo, la evidencia sugiere que la implementación de la IA requiere una redefinición de los roles profesionales dentro de las organizaciones. Los auditores y responsables de control interno deben desarrollar competencias analíticas y tecnológicas que les permitan interactuar con sistemas inteligentes, interpretar sus resultados y supervisar su funcionamiento. En este sentido, la IA no reemplaza el juicio humano, sino que lo complementa, dando lugar a un modelo híbrido de control organizacional.

En síntesis, los resultados indican que la inteligencia artificial tiene el potencial de fortalecer significativamente el control interno empresarial, siempre que su implementación esté acompañada de capacidades organizacionales adecuadas y de marcos de gobernanza sólidos. La literatura analizada sugiere que el futuro del control interno estará marcado por la integración de sistemas inteligentes, pero también por la necesidad de garantizar su uso ético, transparente y responsable.

5. Conclusiones

El presente estudio permitió analizar de manera sistemática el papel de la inteligencia artificial en el fortalecimiento del control interno empresarial durante el período 2016–2026, evidenciando una evolución significativa en la forma en que las organizaciones diseñan y ejecutan sus mecanismos de control.

En primer lugar, los resultados confirman que la inteligencia artificial se ha consolidado como un recurso estratégico que permite mejorar la eficiencia y la precisión de los procesos de control interno. Su capacidad para analizar grandes volúmenes de datos, identificar patrones complejos y generar predicciones ha permitido optimizar funciones clave como la auditoría interna, la detección de fraude y la gestión de riesgos.

En segundo lugar, se observa una transformación estructural en los sistemas de control interno, los cuales han evolucionado desde enfoques reactivos hacia modelos proactivos y predictivos. Este cambio implica que las organizaciones ya no se limitan a detectar errores o irregularidades una vez ocurridos, sino que pueden anticiparse a ellos mediante el uso de herramientas basadas en inteligencia artificial.

Asimismo, el estudio evidencia que la implementación efectiva de la IA depende de factores organizacionales, tales como la capacidad tecnológica, la cultura organizacional y la existencia de marcos de gobernanza adecuados. En este sentido, la integración de la inteligencia artificial no debe considerarse únicamente como un proceso tecnológico, sino como una transformación organizacional más amplia.

Por otro lado, se identifican desafíos relevantes asociados a la adopción de estas tecnologías, entre los que destacan la opacidad de los algoritmos, los riesgos éticos y la necesidad de regulación. Estos aspectos subrayan la importancia de desarrollar políticas y estructuras de control que garanticen un uso responsable de la inteligencia artificial.

En conclusión, la inteligencia artificial no sustituye los sistemas tradicionales de control interno, sino que los transforma hacia modelos híbridos en los que la interacción entre capacidades humanas y sistemas inteligentes se convierte en el eje central de la gestión organizacional moderna.

6. Referencias

Adewumi, T. A. (2025). Rule of law, corporate governance and AI humanoid robots: Charting the course for a global regulatory framework. *Mizan Law Review*, 19(1), 93–116. <https://doi.org/10.4314/mlr.v19i1.4>

Bouziane, A. (2025). The impact of using ChatGPT on internal audit quality: A case study of a subsidiary of an oil group in Morocco. *EDPACS*, 70(3), 42–54. <https://doi.org/10.1080/07366981.2025.2450876>

Caruso, S., Bruccoleri, M., Pietrosi, A., & Scaccianoce, A. (2023). Artificial intelligence to counteract “KPI overload” in business process monitoring: The case of anti-corruption in public organizations. *Business Process Management Journal*, 29(4), 1227–1248. <https://doi.org/10.1108/BPMJ-11-2022-0578>

Castillo-Oliva, J., Montañez-Díaz, B. A., & Mendoza-Rivera, R. D. (2025). Application of artificial intelligence in budget management: A systematic review. *Aibi Revista de Investigación Administración e Ingenierías*, 13(3). <https://doi.org/10.15649/2346030X.4199>

Chen, L., Yang, K., & Yu, F. (2025). Intelligent governance? Evidence from the adoption of AI in Chinese A-share firms. *Finance Research Letters*, 86, 108652. <https://doi.org/10.1016/j.frl.2025.108652>

Dong, Y., Chen, S. X., & Kim, M. (2025). Can corporate governance information facilitate accounting fraud detection? Machine learning evidence for Chinese listed firms. *International Journal of Accounting*. <https://doi.org/10.1142/S109440602550009X>

Liu, Y., Li, X., & Su, C. (2026). From unstructured text to automated insights: An explainable AI approach to internal control in banking systems. *Systems*, 14(3), 234. <https://doi.org/10.3390/systems14030234>

Lu, X., & Wang, T. (2025). Research on risk assessment of internal audit information in enterprises through data mining. *Journal of Information Processing Systems*, 21(6), 575–584. <https://doi.org/10.3745/JIPS.04.0361>

Ma, Y., Li, S., & Huang, S. (2024). Digital transformation and enterprise external pay gap: Internal mechanism and empirical evidence. *Finance Research Letters*, 69, 106296. <https://doi.org/10.1016/j.frl.2024.106296>

Najm, M., & Do Amaral Giovani, J. F. (2025). ICAIMT — The digital oath for corporate governance: A first model. *Journal of Information and Knowledge Management*, 24(6), 2550071. <https://doi.org/10.1142/S0219649225500716>

Schuett, J. (2025). Three lines of defense against risks from AI. *AI and Society*, 40(2), 493–507. <https://doi.org/10.1007/s00146-023-01811-0>