INTELIGENCIA ARTIFICIAL Y SEGURIDAD DE LA INFORMACIÓN: HACIA UN MODELO PREDICTIVO DE PROTECCIÓN DIGITAL. UNA REVISIÓN SISTEMÁTICA DE LA LITERATURA

ARTIFICIAL INTELLIGENCE AND INFORMATION SECURITY: TOWARDS A PREDICTIVE MODEL OF DIGITAL PROTECTION. A SYSTEMATIC LITERATURE REVIEW

Betsy Josefina Rodríguez Rivera
Ingeniera de Sistemas, especialista en Educación Superior a Distancia
Universidad Nacional Abierta y a Distancia
https://orcid.org/0009-0005-4012-9575
Betsy.rodriguez@unad.edu.co

Carlos Alberto Rodríguez Sánchez
Ingeniero de Sistemas, especialista en Seguridad Informática, magíster en Gobierno TI
Universidad Nacional Abierta y a Distancia.
https://orcid.org/0009-0001-0750-0679
Carlos.rodriguezs@unad.edu.co

RESUMEN

La inteligencia artificial (IA) ocupa hoy un lugar central en las estrategias de protección digital, pero el crecimiento acelerado de su literatura dificulta la construcción de un panorama integrado de sus componentes técnicos, operativos, organizacionales y éticos. Este trabajo aborda la ausencia de un marco conceptual que explique cómo la convergencia entre el aprendizaje automático, el aprendizaje profundo, la explicabilidad algorítmica y la gobernanza tecnológica permite transitar hacia una ciberseguridad predictiva, mediante una revisión sistemática adaptada al enfoque PRISMA. El estudio examinó 340 registros en Scopus, IEEE Xplore, ScienceDirect, Web of Science, Dialnet y repositorios institucionales (del 3 al 28 de febrero del 2025); 37 estudios se incorporaron a la síntesis final y 16 se citan de forma directa. La metodología combinó análisis de contenido temático, una matriz de extracción de datos y criterios explícitos de calidad, junto con un análisis de coocurrencia de 15 términos clave que reveló tres clústeres: técnico, operativo y ético-normativo. Los resultados muestran que los modelos predictivos e híbridos mejoran sustancialmente la capacidad de anticipación frente a amenazas, aunque su efectividad depende de la calidad de los datos y de la supervisión humana, y ponen en evidencia una brecha entre la sofisticación técnica y la preparación institucional necesaria para la adopción responsable de estas tecnologías. Se propone el Marco Integrado de Ciberseguridad Predictiva Basada en IA (MICPIA), con cuatro dimensiones —tecnológica, operativa, organizacional y ética— y cinco proposiciones teóricas derivadas de la evidencia. Se concluye que el avance de la ciberseguridad basada en IA depende tanto del progreso algorítmico como de la capacidad organizacional para gobernarlo bajo criterios éticos explícitos, y que el aporte original reside en integrar evidencia técnica y exigencias normativas que la literatura suele tratar por separado.

Palabras clave: inteligencia artificial; ciberseguridad predictiva; seguridad de la información; revisión sistemática; gobernanza tecnológica; inteligencia artificial explicable.

ABSTRACT

Artificial intelligence (AI) now occupies a central place in digital protection strategies; however, the rapid growth of the literature makes it difficult to obtain an integrated understanding of its technical, operational, organizational, and ethical components. This study addresses the lack of a conceptual framework explaining how the convergence of machine learning, deep learning, algorithmic explainability, and technology governance enables a transition from reactive to predictive cybersecurity. To this end, a systematic literature review adapted from the PRISMA approach was conducted. The study examined 340 records identified in Scopus, IEEE Xplore, ScienceDirect, Web of Science, Dialnet, and institutional repositories between 3 and 28 February 2025. Of these, 37 studies were included in the final synthesis, and 16 were directly cited. The methodology combined thematic content analysis, a data extraction matrix, and explicit quality criteria with a co-occurrence analysis of 15 key terms, which revealed three conceptual clusters: technical, operational, and ethical-regulatory. The results show that predictive and hybrid models substantially improve the capacity to anticipate threats, although their effectiveness depends on the quality of training data and human oversight. The findings also reveal a widening gap between technical sophistication and institutional readiness. Based on these findings, the study proposes the Integrated Framework of AI-Based Predictive Cybersecurity (MICPIA), structured around four interrelated dimensions—technological, operational, organizational, and ethical—and grounded in five theoretical propositions. The study concludes that advances in AI-driven cybersecurity depend as much on algorithmic progress as on organizations' capacity to govern these technologies according to explicit ethical criteria. The original contribution of this work lies in integrating technical evidence and normative requirements, which the literature usually addresses separately, within a single framework.

Keywords: artificial intelligence; predictive cybersecurity; information security; systematic review; technology governance; explainable artificial intelligence.

INTRODUCCIÓN

La inteligencia artificial (IA) ha crecido de forma exponencial en la última década y se ha consolidado como una de las tecnologías con mayor impacto en la seguridad de la información, con más de 14 500 registros identificados entre el 2014 y el 2024 (Ofusori et al., 2024). La IA permite anticipar vulnerabilidades antes de que sean explotadas (Kaur et al., 2023; Ferrag et al., 2020), lo que habilita el desarrollo de modelos predictivos; sin embargo, la IA generativa se ha convertido también en un vector de ataque (Achuthan et al., 2024), en un contexto de aumento del costo global del cibercrimen (IBM, 2024). La literatura revela tres vacíos que justifican este trabajo: uno de integración conceptual, pues la abundancia de estudios técnicos no se ha traducido en una cartografía integral de sus subcampos (Albahri y AlAmoodi, 2023; Ofusori et al., 2024); uno teórico-normativo, pues la explicabilidad y la gobernanza suelen tratarse como secciones de cierre (Kaur et al., 2023; Papagiannidis et al., 2025), y uno propositivo, pues la evidencia rara vez se traduce en modelos orientados a la práctica (McKinsey & Company, 2024). Estos vacíos configuran el problema de investigación: la literatura sobre IA aplicada a la seguridad de la información se ha desarrollado de forma fragmentada, con una dimensión técnica que avanza más rápido que su marco de gobernanza, sin una síntesis que articule ambos planos en un modelo único orientado a la práctica.

Para efectos de este análisis, conviene precisar que la seguridad de la información protege la confidencialidad, integridad y disponibilidad de los datos (modelo CIA), mientras que la ciberseguridad es el subconjunto orientado a proteger sistemas y datos digitales; que el aprendizaje automático (ML) y el aprendizaje profundo (DL) son subcampos anidados de la IA que permiten detectar comportamientos anómalos y anticipar incidentes —lo que constituye la ciberseguridad predictiva—, y que la inteligencia artificial explicable (XAI) y la gobernanza tecnológica son las dos vías mediante las cuales esa capacidad predictiva se audita y se legitima institucionalmente.

En consecuencia, el objetivo general de este estudio es analizar críticamente la producción científica sobre IA aplicada a la seguridad de la información para construir un marco conceptual integrado de ciberseguridad predictiva. De él se desprenden cuatro objetivos específicos: (a) caracterizar los fundamentos técnicos —modelos de ML, DL e híbridos— que sustentan la transición hacia la ciberseguridad predictiva; (b) identificar tendencias, convergencias, divergencias y vacíos de investigación en la literatura del periodo 2014-2025; (c) determinar la estructura conceptual del campo mediante un análisis de coocurrencia de términos clave, y (d) formular las dimensiones y proposiciones teóricas de un marco integrado. Estos objetivos se traducen en cuatro preguntas de investigación (PI1-PI4) referidas, respectivamente, a los fundamentos técnicos que sustentan la ciberseguridad predictiva, las

tendencias y vacíos de la literatura del periodo 2014-2025, la estructura conceptual del campo según el análisis de coocurrencia, y las dimensiones y proposiciones que debe integrar un marco conceptual de ciberseguridad predictiva basada en IA.

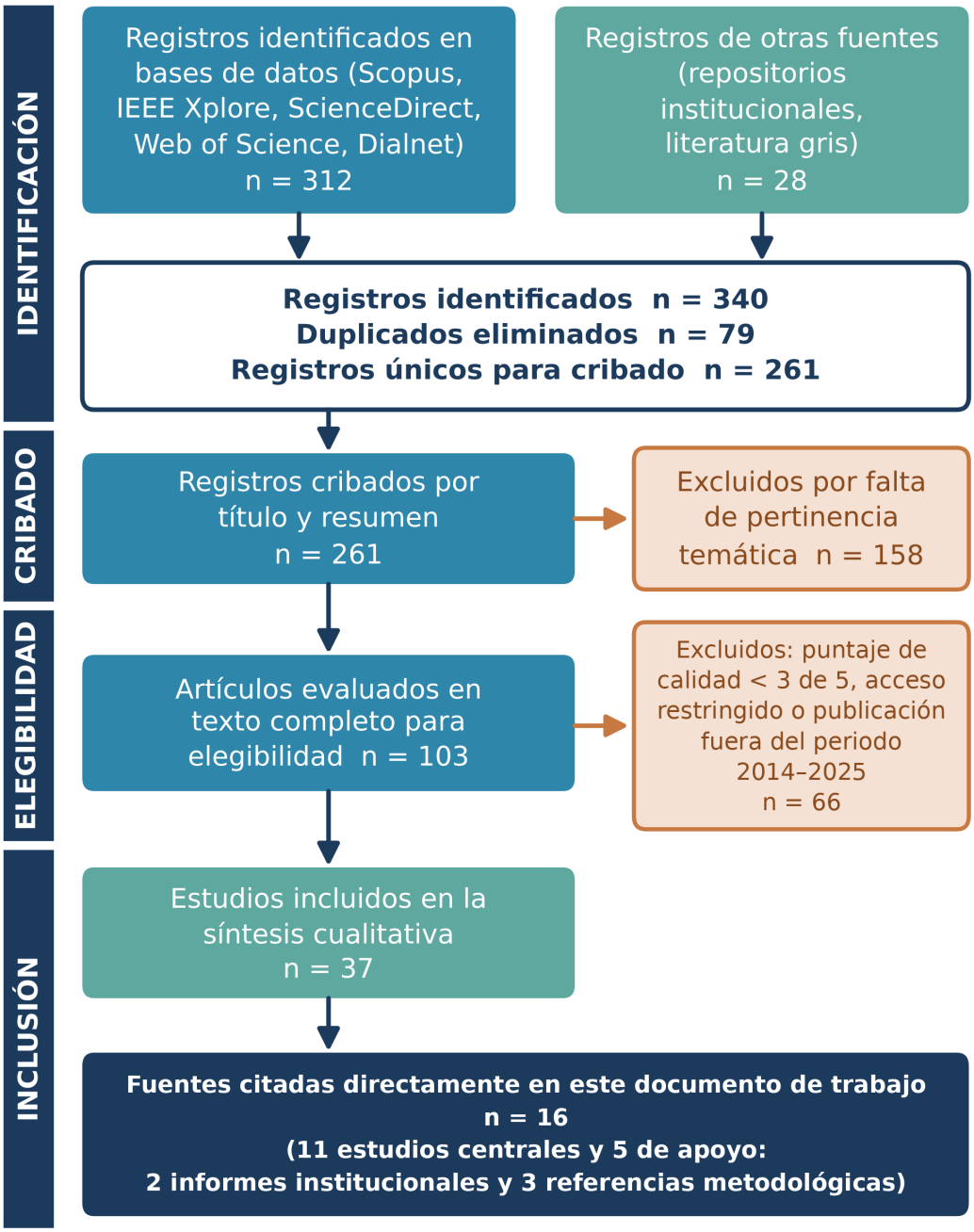
METODOLOGÍA

Se optó por una metodología cualitativa de revisión documental (Hernández Sampieri et al., 2014), con diseño no experimental descriptivo-analítico, siguiendo las cuatro fases del enfoque PRISMA (Page et al., 2021) y complementándola con un análisis de coocurrencia de términos clave. La estrategia de búsqueda se derivó de las preguntas PI1-PI4 mediante tres bloques conceptuales: tecnología (inteligencia artificial, aprendizaje automático o *machine learning* y aprendizaje profundo o *deep learning*), dominio (ciberseguridad, seguridad de la información) y resultado esperado (modelos predictivos, protección digital) —, traducidos en ecuaciones booleanas comparables entre bases de datos. La ecuación base en inglés fue: ("artificial intelligence" OR "machine learning" OR "deep learning") AND ("cybersecurity" OR "information security") AND ("predictive model*" OR "digital protection"), y su equivalente en español, aplicado en Dialnet y en los repositorios institucionales: ("inteligencia artificial" OR "aprendizaje automático" OR "aprendizaje profundo") AND ("ciberseguridad" OR "seguridad de la información") AND ("modelo* predictivo*" OR "protección digital"). Las ecuaciones se aplicaron a los campos de título, resumen y palabras clave, adaptando la sintaxis de cada base (por ejemplo, TITLE-ABS-KEY en Scopus y TS= en Web of Science) y filtrando los registros correspondientes al periodo 2014-2025.

Se consultaron seis fuentes: Scopus e IEEE Xplore, por su producción técnica en ingeniería; ScienceDirect, por sus revistas de alto impacto; Web of Science, para reducir el sesgo de una sola base; Dialnet, por la producción iberoamericana, y repositorios institucionales, por su literatura gris con datos empíricos. Las cinco bases de datos aportaron 312 registros y los repositorios institucionales, 28. Se incluyeron estudios que abordaran la aplicación de la IA —mediante ML, DL o modelos híbridos— a la ciberseguridad o la seguridad de la información; que se hubieran publicado entre el 2014 y el 2025; que correspondieran a artículos revisados por pares, revisiones sistemáticas o informes institucionales con datos empíricos, y que estuvieran disponibles en texto completo en inglés o español. Se excluyeron los registros duplicados, los que no trataban la IA aplicada a la protección digital, los de acceso restringido, los publicados fuera del periodo y los que no alcanzaron el umbral mínimo de calidad descrito más adelante. El proceso de selección, adaptado del diagrama PRISMA 2020 (véase la figura 1), se desarrolló en cuatro fases: identificación (340 registros), cribado (79 duplicados eliminados y 158 descartados por pertinencia, con

lo que quedaron 103 registros en evaluación), elegibilidad (66 excluidos por calidad insuficiente, acceso restringido o fecha de publicación) e inclusión (37 estudios en la síntesis final, de los cuales 16 se citan directamente: 11 estudios centrales y 5 de apoyo, correspondientes a 2 informes institucionales y 3 referencias metodológicas).

Figura 1. Diagrama de flujo del proceso de selección documental



Fuente: adaptado del enfoque PRISMA 2020 (Page *et al.*, 2021).

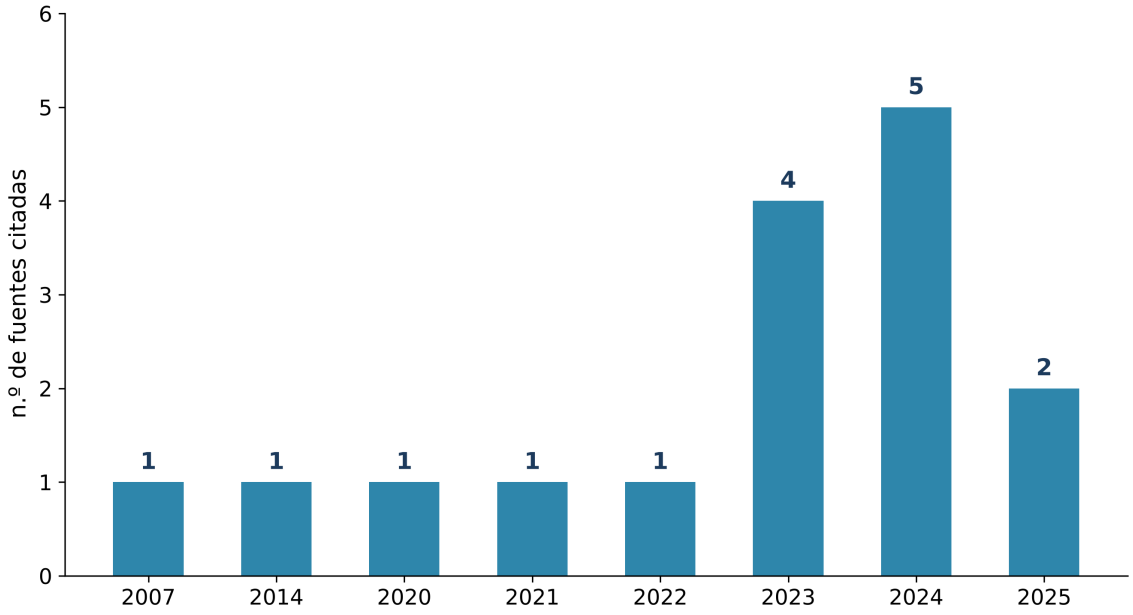
Para cada uno de los 37 estudios se diligenció una matriz de extracción estandarizada que garantizó la comparabilidad de la información y redujo el sesgo de selección. Los 103 estudios en evaluación se calificaron con una escala adaptada de Kitchenham y Charters (2007) que contempló cinco criterios: claridad del objetivo, idoneidad del corpus, validez del diseño, transparencia de métricas y tratamiento ético. Cada criterio se puntuó con 0 (no cumple), 0,5 (cumple parcialmente) o 1 (cumple), y se fijó un umbral mínimo de aceptación de 3 puntos sobre 5; los estudios por debajo de ese puntaje se excluyeron en la fase de elegibilidad. El análisis combinó el análisis de contenido temático, organizado en cinco ejes a partir de la matriz de extracción, con un análisis de coocurrencia de 15 términos clave sobre el contenido efectivamente desarrollado por los autores —y no sobre metadatos bibliográficos—, tomando el párrafo como unidad de análisis y representando los resultados en una matriz de 15 × 15, agrupada por clústeres de densidad relativa.

El estudio se basa en fuentes secundarias y no involucra sujetos humanos. Para mitigar el sesgo de selección, se mantuvo un registro sistemático del proceso de búsqueda, se aplicaron criterios de calidad explícitos y se emplearon dos revisores independientes. Su naturaleza documental impone límites a la validación empírica de los hallazgos, aspecto que se retoma en las conclusiones.

RESULTADOS

La incorporación de la IA en la seguridad de la información se ha convertido en un factor decisivo para la detección de amenazas y la gestión de riesgos, aunque persisten brechas relevantes: numerosas vulnerabilidades en modelos desplegados permanecen sin mitigación adecuada (IBM, 2024), en un contexto de infraestructuras heredadas y vacíos regulatorios. Las 16 fuentes citadas directamente —un subconjunto del corpus de 37 estudios— se concentran mayoritariamente entre el 2022 y el 2025 (12 de 16; véase las figuras 2 y 3), lo que coincide con la aceleración de la IA generativa y con el aumento de estudios sobre explicabilidad y gobernanza. Este patrón sugiere que el campo atraviesa una fase de consolidación normativa posterior a una primera ola de estudios predominantemente técnicos (2014-2019).

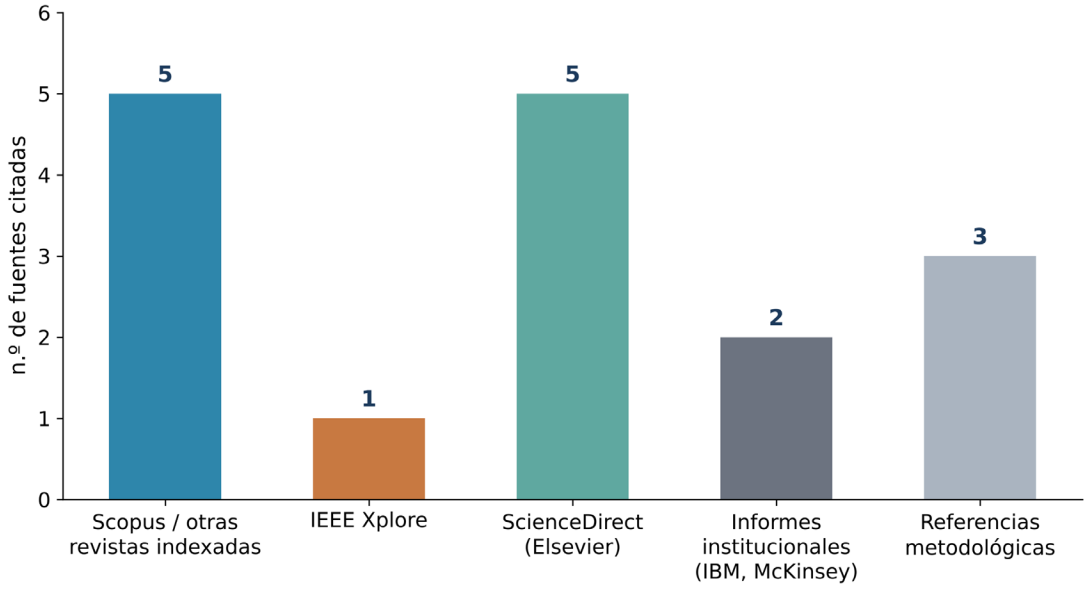
Figura 2. Distribución temporal de las fuentes citadas directamente



Nota: n = 16; subconjunto del corpus de 37 estudios. Incluye una referencia metodológica de 2007 (Kitchenham y Charters, 2007).

Fuente: elaboración propia.

Figura 3. Distribución de las fuentes citadas directamente según base de datos u origen



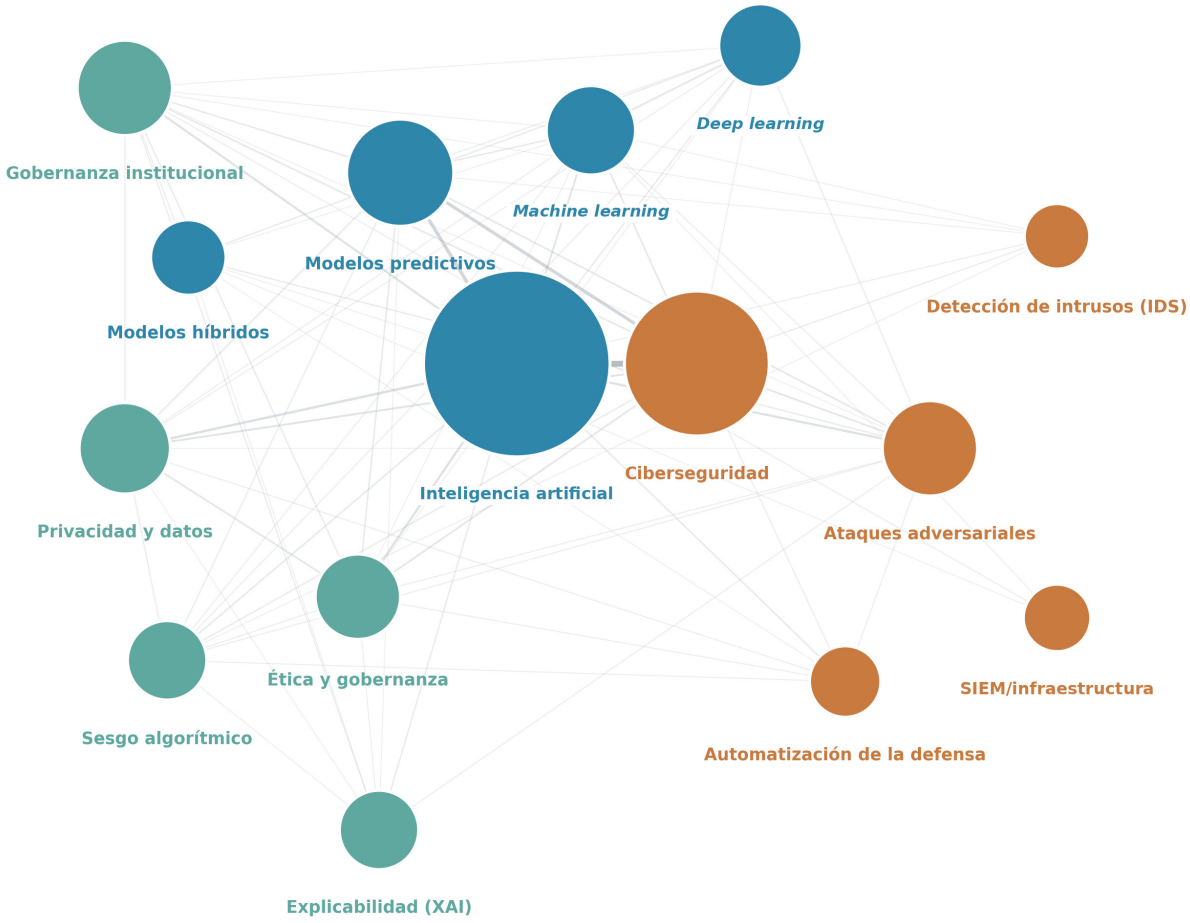
Nota: n = 16; subconjunto del corpus de 37 estudios.

Fuente: elaboración propia.

El análisis comparativo-crítico de los estudios centrales, que incorpora su contribución diferencial y su principal limitación, permite extraer tres lecturas. En primer lugar, se observa una convergencia metodológica: los estudios técnicos validan sus modelos sobre un número reducido de *datasets* de referencia —NSL-KDD, CICIDS2017 y UNSW-NB15—, lo que facilita la comparabilidad pero concentra el riesgo de sobreajuste. En segundo lugar, se identifica una divergencia de énfasis: los estudios del 2020 al 2022 priorizan el desempeño predictivo, mientras que los del 2023 al 2025 priorizan la explicabilidad. En tercer lugar, se pone en evidencia un vacío compartido: ninguno de los estudios revisados integra simultáneamente desempeño, explicabilidad y gobernanza, lo que justifica el marco integrado propuesto en este trabajo.

El análisis de coocurrencia de 15 términos clave en los 37 estudios del corpus —que examina el contenido efectivamente desarrollado por los autores y no los metadatos bibliográficos— reveló tres clústeres con densidades desiguales (véase la figura 4). El primero, técnico, agrupa IA, ML, DL y modelos predictivos, con "inteligencia artificial" (64 menciones) como nodo puente; el segundo, operativo, reúne ciberseguridad (43), detección de intrusos y ataques adversariales; el tercero, ético-normativo —privacidad, gobernanza y explicabilidad—, es el de menor densidad interna. El par "inteligencia artificial-ciberseguridad" concentra 38 coapariciones, seguido de "inteligencia artificial-modelos predictivos" (19) y "ciberseguridad-modelos predictivos" (18). También destaca que el vínculo entre "inteligencia artificial" y los términos éticos —privacidad (12), ética y gobernanza (11) — presenta una frecuencia de coaparición cercana a la de su relación con "ataques adversariales" (11), lo que sugiere que la literatura ya no discute la técnica de forma aislada respecto de sus implicaciones éticas (véase la figura 5). Estos hallazgos son consistentes con las lecturas bibliométricas de Albahri y AlAmoodi (2023) y Ofusori et al. (2024).

Figura 4. Red de coocurrencia de términos clave en el corpus analizado

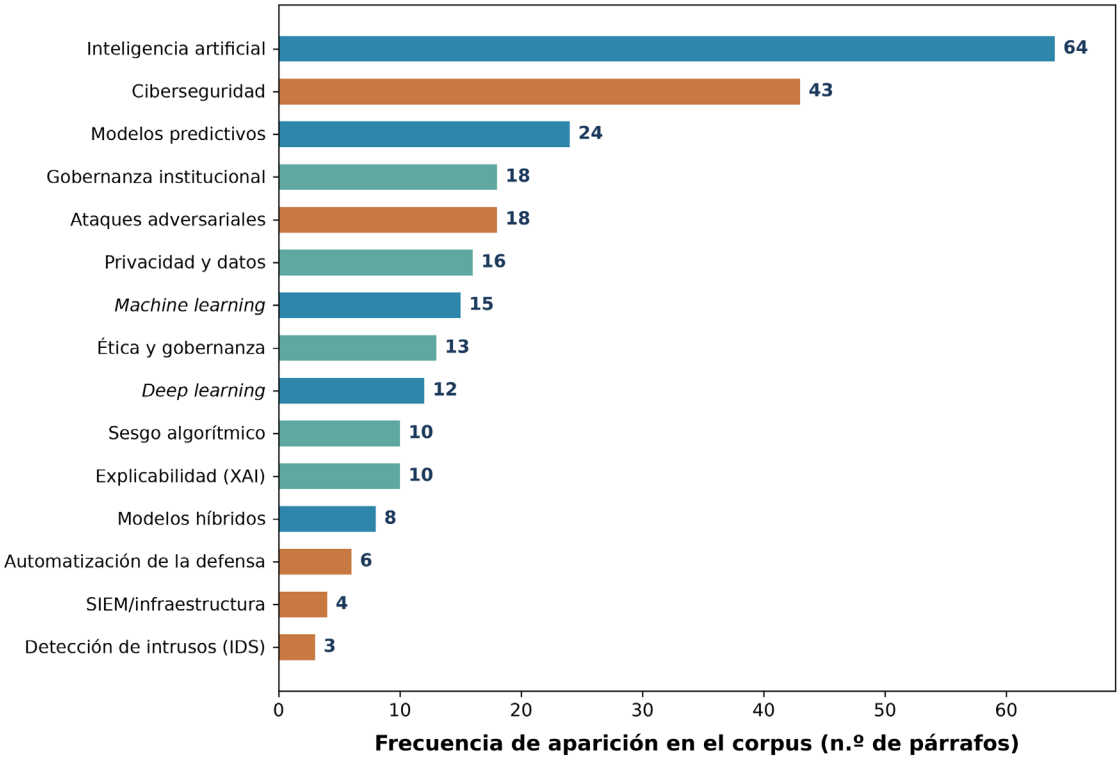


Nota: corpus completo, $n = 37$.

Fuente: elaboración propia.

En términos técnicos, los modelos de ML emplean algoritmos supervisados y no supervisados para detectar comportamientos anómalos en plataformas SIEM (Pinto et al., 2023; Ferrag et al., 2020); los modelos de DL amplían esa capacidad mediante redes CNN, RNN/LSTM e híbridas (Macas et al., 2022; Azam et al., 2023), y el aprendizaje automático automatizado (AutoML) permite automatizar su diseño (Brown et al., 2024). Esta síntesis confirma que el campo avanza de forma desigual en sus tres dimensiones —un núcleo técnico maduro, uno operativo consolidado y uno ético-normativo disperso—, una asimetría que el marco integrado presentado a continuación busca resolver.

Figura 5. Frecuencia de términos clave identificados en la revisión

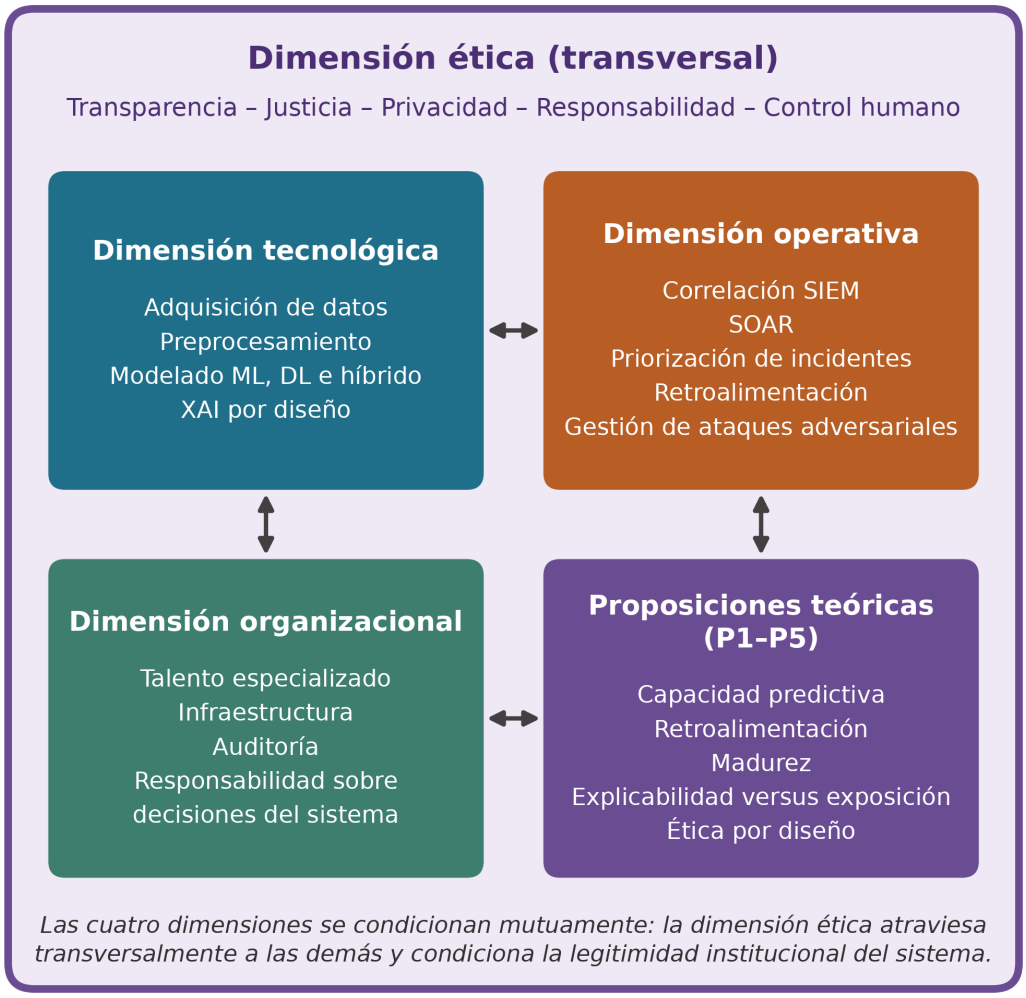


Nota: corpus completo, $n = 37$.

Fuente: elaboración propia.

A partir de estos hallazgos, se propone el Marco Integrado de Ciberseguridad Predictiva Basada en IA (MICPIA), no como un producto de *software*, sino como una estructura de cuatro dimensiones interrelacionadas que una organización puede usar para diagnosticar su madurez y orientar decisiones de diseño, adopción y gobierno (figura 6). La dimensión tecnológica provee la capacidad predictiva (adquisición y preprocesamiento de datos, modelado con ML, DL y modelos híbridos, explicabilidad incorporada por diseño); la operativa traduce esa capacidad en procesos de detección, correlación de alertas y respuesta, e incorpora la gestión de ataques adversariales como riesgo continuo; la organizacional recoge las condiciones institucionales —talento especializado, infraestructura, auditoría periódica y asignación de responsabilidades— que determinan una adopción sostenible, y la ética atraviesa las otras tres dimensiones transversalmente bajo los principios de transparencia, justicia, privacidad, responsabilidad y control humano.

Figura 6. Arquitectura conceptual del Marco Integrado de Ciberseguridad Predictiva Basada en Inteligencia Artificial (MICPIA)



Fuente: elaboración propia.

De la evidencia se derivan cinco proposiciones teóricas, planteadas como líneas de validación empírica futura: (P1) la capacidad predictiva depende más de la calidad de los datos de entrenamiento que de la arquitectura algorítmica; (P2) la efectividad operativa depende de la retroalimentación continua y no solo del desempeño inicial sobre *datasets* de referencia; (P3) la sostenibilidad de la adopción depende más de la madurez organizacional —talento, infraestructura y auditoría— que de la sofisticación técnica del modelo; (P4) la explicabilidad algorítmica mejora la confianza institucional, pero puede exponer información explotable por un atacante, por lo que su diseño debe equilibrar transparencia y exposición, y (P5) la gobernanza ética no puede tratarse como una dimensión de cierre, sino que debe condicionar, desde el diseño técnico inicial, las decisiones tomadas en las demás dimensiones.

CONCLUSIONES

Este estudio se propuso construir, mediante una revisión sistemática de literatura, un marco conceptual integrado de ciberseguridad predictiva basada en IA, en respuesta a cuatro preguntas de investigación. Frente a PI1, la evidencia muestra una trayectoria técnica clara: del aprendizaje automático clásico (SVM, árboles de decisión, bosques aleatorios) hacia arquitecturas de aprendizaje profundo (CNN, RNN/LSTM) y, más recientemente, hacia modelos híbridos con explicabilidad y automatización del diseño (AutoML); esta progresión no es lineal, pues cada ganancia en capacidad predictiva tiende a reducir la interpretabilidad del modelo, lo que exige incorporar la explicabilidad como componente de diseño y no como un añadido posterior.

Frente a PI2, el análisis identificó una convergencia metodológica en torno a un número reducido de *datasets* de referencia, una divergencia temporal de énfasis —de la precisión predictiva hacia la explicabilidad— y, sobre todo, un vacío compartido por los estudios centrales: ninguno integra simultáneamente desempeño técnico, explicabilidad y gobernanza, lo que constituye un síntoma de una asimetría estructural del campo. Este hallazgo es consistente con revisiones bibliométricas recientes (Albahri y AlAmoodi, 2023; Ofusori et al., 2024) y con Achuthan et al. (2024), aunque estos trabajos no abordan la tensión entre explicabilidad y exposición al atacante que aquí se formula como proposición teórica (P4); Papagiannidis et al. (2025) confirman, desde los sistemas de información, que los principios de gobernanza de alto nivel no han estado acompañados de una operacionalización clara, y Díaz et al. (2025) reporta para América Latina un vacío de equidad y transparencia consistente con el identificado en este estudio.

Frente a PI3, el análisis de coocurrencia de contenido documental —distinto de los mapas bibliométricos de metadatos, aunque convergente con ellos— reveló tres clústeres con densidades desiguales: un núcleo técnico denso que actúa como puente, un núcleo operativo consolidado con conectividad ética limitada, y un núcleo ético-normativo disperso. Frente a PI4, el MICPIA articula cuatro dimensiones interdependientes —tecnológica, operativa, organizacional y ética— y las traduce en cinco proposiciones teóricas (P1-P5); la inclusión explícita de una dimensión organizacional, ausente en la mayoría de los estudios técnicos revisados, responde directamente al vacío propositivo identificado desde la introducción.

La contribución original de este trabajo reside en tres aportes: una metodología de coocurrencia de contenido documental reproducible y distinta de la bibliometría de metadatos convencional; una síntesis comparativo-crítica que identifica explícitamente convergencias, divergencias y vacíos entre los estudios centrales, y un marco propositivo, el MICPIA, que traduce esa síntesis en dimensiones y proposiciones orientadas a la práctica y a la investigación futura.

Como principal limitación, la naturaleza documental del estudio no permite validar empíricamente el marco propuesto, y la búsqueda se cerró en febrero del 2025 sin capturar la producción más reciente sobre IA generativa aplicada a la ciberseguridad; validar el MICPIA en un caso organizacional real y replicar el análisis de coocurrencia con herramientas bibliométricas estandarizadas son líneas prioritarias de investigación futura. En particular, una publicación subsecuente ampliará la ventana de revisión hasta la segunda mitad del 2026, con el fin de evaluar empíricamente si la estructura y las proposiciones del MICPIA se mantienen frente a los acelerados avances tecnológicos o si requieren ajustes.

Mientras la sofisticación algorítmica siga creciendo más rápido que la capacidad institucional para auditarla y gobernarla, la adopción de IA predictiva corre el riesgo de amplificar el riesgo que pretende reducir. El futuro de la ciberseguridad basada en IA dependerá tanto del avance algorítmico como de la capacidad de las organizaciones para gobernar esa tecnología bajo criterios éticos explícitos.

REFERENCIAS

Achuthan, K., Ramanathan, S., Srinivas, S., y Raman, R. (2024). Advancing cybersecurity and privacy with artificial intelligence: Current trends and future research directions. *Frontiers in Big Data*, *7*, 1497535. https://doi.org/10.3389/fdata.2024.1497535

Albahri, O. S., y AlAmoodi, A. H. (2023). Cybersecurity and artificial intelligence applications: A bibliometric analysis based on scopus database. *Mesopotamian Journal of CyberSecurity*, *2023*, 158-169. https://doi.org/10.58496/MJCSC/2023/018

Azam, Z., Islam, M. M., y Huda, M. N. (2023). Comparative analysis of intrusion detection systems and machine learning-based model analysis through decision tree. *IEEE Access*, *11*, 80348-80391. https://doi.org/10.1109/ACCESS.2023.3296444

Brown, A., Gupta, M., y Abdelsalam, M. (2024). Automated machine learning for deep learning based malware detection. *Computers & Security*, *137*, 103582. https://doi.org/10.1016/j.cose.2023.103582

Díaz, F., Cerna, N., y Liza, R. (2025). Artificial intelligence and crime in Latin America: A multilingual bibliometric review (2010-2025). *Information*, *16*(11), 1001. https://doi.org/10.3390/info16111001

Ferrag, M. A., Maglaras, L., Moschoyiannis, S., y Janicke, H. (2020). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. *Journal of Information Security and Applications*, *50*, 102419. https://doi.org/10.1016/j.jisa.2019.102419

Hernández Sampieri, R., Fernández Collado, C., y Baptista Lucio, M. P. (2014). *Metodología de la investigación* (6.ª ed.). McGraw-Hill Interamericana.

IBM. (2026). *Cost of a data breach report 2026*. https://www.ibm.com/reports/data-breach

Kaur, R., Gabrijelčič, D., y Klobučar, T. (2023). Artificial intelligence for cybersecurity: Literature review and future research directions. *Information Fusion*, *97*, 101804. https://doi.org/10.1016/j.inffus.2023.101804

Kitchenham, B., y Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering* (EBSE Technical Report EBSE-2007-01). Keele University and University of Durham.

Macas, M., Wu, C., y Fuertes, W. (2022). A survey on deep learning for cybersecurity: Progress, challenges, and opportunities. *Computer Networks*, *212*, 109032. https://doi.org/10.1016/j.comnet.2022.109032

McKinsey & Company. (2024, 14 de noviembre). *The cybersecurity provider's next opportunity: Making AI safer*. https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/the-cybersecurity-providers-next-opportunity-making-ai-safer

Ofusori, L., Bokaba, T., y Mhlongo, S. (2024). Artificial intelligence in cybersecurity: A comprehensive review and future direction. *Applied Artificial Intelligence*, *38*(1), 2439609. https://doi.org/10.1080/08839514.2024.2439609

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, n71. https://doi.org/10.1136/bmj.n71

Papagiannidis, E., Mikalef, P., y Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *The Journal of Strategic Information Systems*, *34*(2), 101885. https://doi.org/10.1016/j.jsis.2024.101885

Pinto, A., Herrera, L.-C., Donoso, Y., y Gutierrez, J. A. (2023). Survey on intrusion detection systems based on machine learning techniques for the protection of critical infrastructure. *Sensors*, *23*(5), 2415. https://doi.org/10.3390/s23052415