https://doi.org/10.56048/MQR20225.9.1.2025.e128

Security threats associated with the integration of artificial intelligence in information systems: Systematic review

Amenazas de seguridad asociadas con la integración de inteligencia artificial en sistemas de información: Revisión sistemática

Autores:

Calle-Méndez, Jorge Leonardo
UNIVERSIDAD CATÓLICA DE CUENCA
Estudiante de la Maestría en Ciberseguridad
Cuenca – Ecuador
jorge.calle@ucacue.edu.ec
https://orcid.org/0009-0008-5518-1183

Barriga-Andrade, Jhonattan Javier
UNIVERSIDAD CATÓLICA DE CUENCA
Cuenca – Ecuador
jhonattan.barriga@ucacue.edu.ec
https://orcid.org/0000-0001-7334-9113

Yamba-Yugsi, Marco Vinicio
UNIVERSIDAD CATÓLICA DE CUENCA
Estudiante de la maestría en Ciberseguridad. Posgrados
Cuenca – Ecuador
marco.yamba@ucacue.edu.ec
https://orcid.org/0000-0003-4095-1444

Fechas de recepción: 04-ENE-2025- aceptación: 04-FEB-2025 publicación: 15-MAR-2025

https://orcid.org/0000-0002-8695-5005

http://mqrinvestigar.com/

Resumen

La presente revisión sistemática tuvo como objetivo, analizar las amenazas de seguridad asociadas con la integración de la inteligencia artificial en los sistemas de información mediante una revisión sistemática, con el fin de identificar vulnerabilidades, riesgos emergentes y estrategias de mitigación que permitan fortalecer la ciberseguridad en un entorno cada vez más automatizado y dependiente de la tecnología. Se realizó una búsqueda de bibliográfica de artículos científicos en bases de datos como: WOS, SCOPUS, SPRINGER Y SCIELO. Como estrategia de búsqueda se utilizó operadores booleanos "AND" y "OR", con el fin de tener el mayor número de artículos previo a su exclusión. Los resultados identificaron una variedad de amenazas críticas derivadas de la integración de la (IA) en sistemas de información. Entre las principales se encuentran la manipulación de datos, los ataques adversarios, la generación de phishing personalizado, los ataques de denegación de servicio (DDoS) y la intercepción de datos en tránsito. Estas amenazas comprometen la integridad, la confidencialidad y la disponibilidad de los sistemas, exponiendo información sensible y permitiendo que actores maliciosos exploten vulnerabilidades en entornos distribuidos como la nube, sistemas IoT y redes basadas en IA. Se destacó que la automatización y la complejidad tecnológica amplifican significativamente el panorama de riesgos, afectando tanto la protección de datos como la confiabilidad de los modelos implementados. Para mitigar estos riesgos, las estrategias propuestas incluyen la implementación de criptografía avanzada, tecnologías como blockchain y detección de anomalías, así como la necesidad de establecer marcos regulatorios y fomentar la cooperación internacional.

Palabras clave: Sistemas de información; inteligencia artificial; vulnerabilidades; riesgos; ciberseguridad; amenazas

Abstract

The present systematic review aimed to analyze the security threats associated with the integration of artificial intelligence into information systems through a systematic review, in order to identify vulnerabilities, emerging risks, and mitigation strategies that strengthen cybersecurity in an increasingly automated and technology-dependent environment. A bibliographic search of scientific articles was conducted in databases such as WOS, SCOPUS, SPRINGER, and SCIELO. The search strategy utilized Boolean operators "AND" and "OR" to maximize the number of articles before applying exclusion criteria. The results identified a variety of critical threats arising from the integration of AI into information systems. Key threats include data manipulation, adversarial attacks, personalized phishing generation, denial-of-service (DDoS) attacks, and data interception during transmission. These threats compromise the integrity, confidentiality, and availability of systems, exposing sensitive information and enabling malicious actors to exploit vulnerabilities in distributed environments such as the cloud, IoT systems, and AI-based networks. Automation and technological complexity were highlighted as significantly amplifying the risk landscape, affecting both data protection and the reliability of implemented models. To mitigate these risks, proposed strategies include the implementation of advanced cryptography, technologies such as blockchain and anomaly detection, as well as the need to establish regulatory frameworks and promote international cooperation.

Keywords: Information systems; artificial intelligence; vulnerabilities; risks; cybersecurity; threats

Introducción

Según Oluwafemi (2022), en los últimos años, la inteligencia artificial (IA) ha transformado significativamente los sistemas de información, ofreciendo soluciones innovadoras para la automatización, la toma de decisiones y la mejora de la ciberseguridad. Sin embargo, este avance tecnológico también ha traído consigo una nueva serie de desafíos, particularmente en el ámbito de la seguridad. Las amenazas de ciberseguridad han evolucionado en complejidad y sofisticación, aprovechando las capacidades de la IA para lanzar ataques más inteligentes y adaptativos. En este contexto, surge la necesidad de comprender las vulnerabilidades y riesgos asociados con la integración de la IA en los sistemas de información, así como de proponer estrategias que fortalezcan las defensas cibernéticas.

De acuerdo con Corchado, (2023), la capacidad de la IA para analizar grandes volúmenes de datos, predecir comportamientos y adaptarse a amenazas emergentes ha revolucionado la forma en que las organizaciones abordan la seguridad de la información. Sin embargo, esta misma tecnología puede ser utilizada de manera maliciosa, aumentando los riesgos de ataques avanzados, como el phishing inteligente, el malware adaptativo y las técnicas de suplantación mediante deepfakes. Además, la IA está sujeta a vulnerabilidades específicas, como los sesgos algorítmicos, los ataques adversarios y la manipulación de datos de entrenamiento, que pueden comprometer la confiabilidad y la integridad de los sistemas.

Por otro lado, García (2024) plantea que uno de los riesgos principales asociados con la inteligencia artificial generativa radica en el uso de obras protegidas por derechos de propiedad intelectual (PI) para entrenar los modelos de IA sin el consentimiento explícito de sus autores. Este problema genera una vulneración potencial de los derechos de los creadores, ya que los resultados generados por los sistemas de IA pueden replicar, modificar o explotar de sistemas de IA no asumen la responsabilidad de garantizar que los resultados generados no infringen derechos de terceros, trasladando la carga de la prueba a los autores. Esto implica que los creadores deben identificar y

probar las infracciones, una tarea compleja y poco práctica, lo que deja a los titulares de PI en una posición de desventaja.

Desde una perspectiva tecnológica y ética, el autor resalta que los sistemas de IA generativa dependen completamente de los datos y estructuras lógicas proporcionadas durante su entrenamiento. Esto implica que cualquier sesgo o error en los datos puede influir negativamente en las decisiones y resultados de la IA, aumentando los riesgos tecnológicos y éticos. Además, la falta de supervisión humana en los sistemas de IA puede amplificar estos problemas, dificultando la detección y corrección de errores que podrían tener un impacto significativo en los derechos de los usuarios.

Cabe mencionar que las tecnologías emergentes están desempeñando un papel crucial en la evolución de los sistemas de seguridad, transformando tanto los procesos como las capacidades de las organizaciones. (Flores et al., 2023). De este modo, la IA y el aprendizaje automático (ML) se han convertido en herramientas clave que permiten automatizar la identificación y respuesta ante amenazas. Desde la perspectiva de Uribe (2024), estas tecnologías no solo aceleran los tiempos de reacción, sino que también mejoran la precisión al analizar grandes volúmenes de datos en tiempo real, lo que minimiza los errores humanos y optimiza los recursos.

De acuerdo con Jara et al. (2023), el avance de la IA en los sistemas de información, incluyendo motores de búsqueda, redes descentralizadas y sistemas cuánticos, ha transformado significativamente la capacidad de predicción, automatización y procesamiento de datos. Sin embargo, esta evolución también ha expuesto a los sistemas de información a nuevas y sofisticadas amenazas de seguridad. Según Erazo et al., 2023), la transición de plataformas reactivas a sistemas proactivos, que buscan anticipar las necesidades del usuario, implica una mayor recopilación y análisis de datos sensibles, lo que amplifica los riesgos de manipulación, acceso no autorizado y corrupción de información crítica.

Shehab et al. (2020), señala en su estudio la creciente interdependencia entre la IA y la ingeniería de software (SE), señalando que esta relación impulsa avances significativos en ambas áreas, esto debido a que la IA está transformando los sistemas informáticos al automatizar procesos, mejorar la toma de decisiones y optimizar el rendimiento del software. No obstante, esta integración también trae consigo riesgos y desafíos importantes, entre ellos, se encuentran la falta de

aplicabilidad en los modelos de IA, lo que dificulta la detección de errores o sesgos, así como la posibilidad de vulnerabilidades en los sistemas que pueden ser explotadas por amenazas externas. Asimismo, Otilia (2021), explica en su estudio cómo la IA se está integrando en los sistemas informáticos, desempeñando un papel fundamental en la ciberseguridad al analizar y replicar el comportamiento humano para mejorar la detección de amenazas y la toma de decisiones automatizada. La capacidad de la IA para predecir y responder a ciberataques refuerza la protección de los datos almacenados en la nube, garantizando su integridad y confidencialidad. Sin embargo, esta dependencia de la IA también introduce riesgos.

Aslan et al. (2023) en su estudio aborda las amenazas cibernéticas, vulnerabilidades y ataques más comunes, además de recomendaciones y precauciones para protegerse. Se describen diferentes tipos de amenazas como virus, troyanos, gusanos, rootkits y ransomware, que afectan tanto a los usuarios como a los sistemas. También se destacan herramientas de escaneo de vulnerabilidades que permiten identificar debilidades en los sistemas, así como una amplia gama de ataques, desde ingeniería social hasta ataques de red y de contraseñas.

Cabe destacar que los sistemas de IA pueden ser manipulados o atacados por hackers, lo que podría comprometer su funcionamiento y crear nuevas vulnerabilidades. Además, el avance de la IA en ciberseguridad plantea preocupaciones éticas sobre la privacidad, el control excesivo y la posibilidad de que los sistemas tomen decisiones erróneas sin intervención humana. (Guaña, 2023). Por tanto, si bien la IA fortalece la defensa de los sistemas informáticos, su implementación debe ir acompañada de estrategias que mitiguen estos riesgos y amenazas emergentes.

Por lo anteriormente expresado, el objetivo general de esta investigación es analizar las amenazas de seguridad asociadas con la integración de la inteligencia artificial en los sistemas de información mediante una revisión sistemática, con el fin de identificar vulnerabilidades, riesgos emergentes y estrategias de mitigación que permitan fortalecer la ciberseguridad en un entorno cada vez más automatizado y dependiente de la tecnología. Esta investigación busca proporcionar un marco comprensivo que sirva tanto para la comprensión de los desafíos actuales como para la anticipación de amenazas futuras en el ámbito de la seguridad informática.

En el ámbito académico, esta investigación es relevante porque aborda una problemática emergente en el cruce entre la inteligencia artificial y la seguridad de la información, un área que

demanda estudios exhaustivos debido a su creciente impacto en múltiples disciplinas. Si bien la IA está transformando la manera en que se detectan y responden las amenazas, los riesgos asociados a su implementación, como el uso malicioso de algoritmos o las vulnerabilidades intrínsecas de los sistemas de aprendizaje automático, han sido poco explorados de manera integral en la literatura científica. Este trabajo contribuirá al desarrollo teórico de este campo al identificar tendencias y lagunas en la investigación actual, sirviendo como referencia para futuros estudios sobre cómo balancear innovación tecnológica y seguridad.

Metodología

En la presente investigación sistemática, se adoptó un enfoque cualitativo, lo que permitió comprender en profundidad las amenazas de seguridad relacionadas con la integración de la inteligencia artificial en los sistemas de información. Asimismo, se empleó un diseño descriptivo con el propósito de identificar las principales vulnerabilidades destacadas en la literatura, proporcionando una visión clara y precisa de los desafíos actuales en este ámbito.

Se realizó una búsqueda de bibliográfica de artículos científicos en bases de datos como: WOS, SCOPUS, SPRINGER Y SCIELO. Como estrategia de búsqueda se utilizó operadores booleanos “AND” y “OR”, con el fin de tener el mayor número de artículos previo a su exclusión. A continuación, se presenta la cadena de búsqueda utilizada: (("Security threats" OR "cybersecurity risks" OR "vulnerabilities" OR "amenazas de seguridad" OR "riesgos de ciberseguridad" OR "vulnerabilidades")) (("artificial intelligence" OR "AI" OR "inteligencia artificial")) AND (("information systems" OR "IT systems" OR "sistemas de información" OR "tecnología de la información")) AND (("systematic review" OR "systematic analysis" OR "revisión sistemática" OR "análisis sistemático"))

El proceso de selección se realizó mediante la declaración PRISMA. El uso de PRISMA permitió estructurar de manera clara cada etapa del proceso, desde la identificación y cribado de estudios relevantes hasta la síntesis y presentación de los resultados. (Ver figura 1).

Criterios de inclusión: CI-1 Estudios cualitativos y revisiones, CI-2 artículos entre los años 2020 y 2024, CI-3 artículos publicados en idiomas: inglés y español, CI-4 investigaciones que aborden amenazas cibernéticas, riesgos de seguridad y vulnerabilidades en la integración de IA en sistemas

de información, CI-5 estudios con acceso completo. Criterios de exclusión: CE-1 estudios cuantitativos, CE-2 estudios publicados antes de 2020, CE-3 estudios publicado en otros idiomas que no sea en inglés y español ce-4 investigaciones que no aborden específicamente amenazas cibernéticas, riesgos de seguridad o vulnerabilidades relacionadas con la integración de IA en sistemas de información, CE-4 artículos sin acceso completo al texto (solo resúmenes o de acceso restringido).

Para la evaluación de la calidad de los estudios en la presente investigación sobre las amenazas de seguridad asociadas con la integración de inteligencia artificial en sistemas de información, se utilizó la herramienta Critical Appraisal Skills Programme (CASP). Esta herramienta permitió analizar de manera estructurada y objetiva la validez, relevancia y aplicabilidad de los estudios seleccionados, asegurando que cumplieran con los criterios predefinidos del método PRISMA. Gracias al CASP, se garantizó que los estudios incluidos fueran metodológicamente sólidos y relevantes para abordar los objetivos de la revisión sistemática.

Para el análisis de los datos en este estudio se empleó el método de síntesis narrativa debido a la diversidad y heterogeneidad de las investigaciones disponibles sobre el tema. Los estudios revisados abarcan diferentes enfoques teóricos, perspectivas técnicas y contextos aplicados, por lo tanto, este método permitió identificar, organizar y contextualizar de manera exhaustiva las principales amenazas de seguridad, resaltando patrones comunes, discrepancias y temas emergentes.

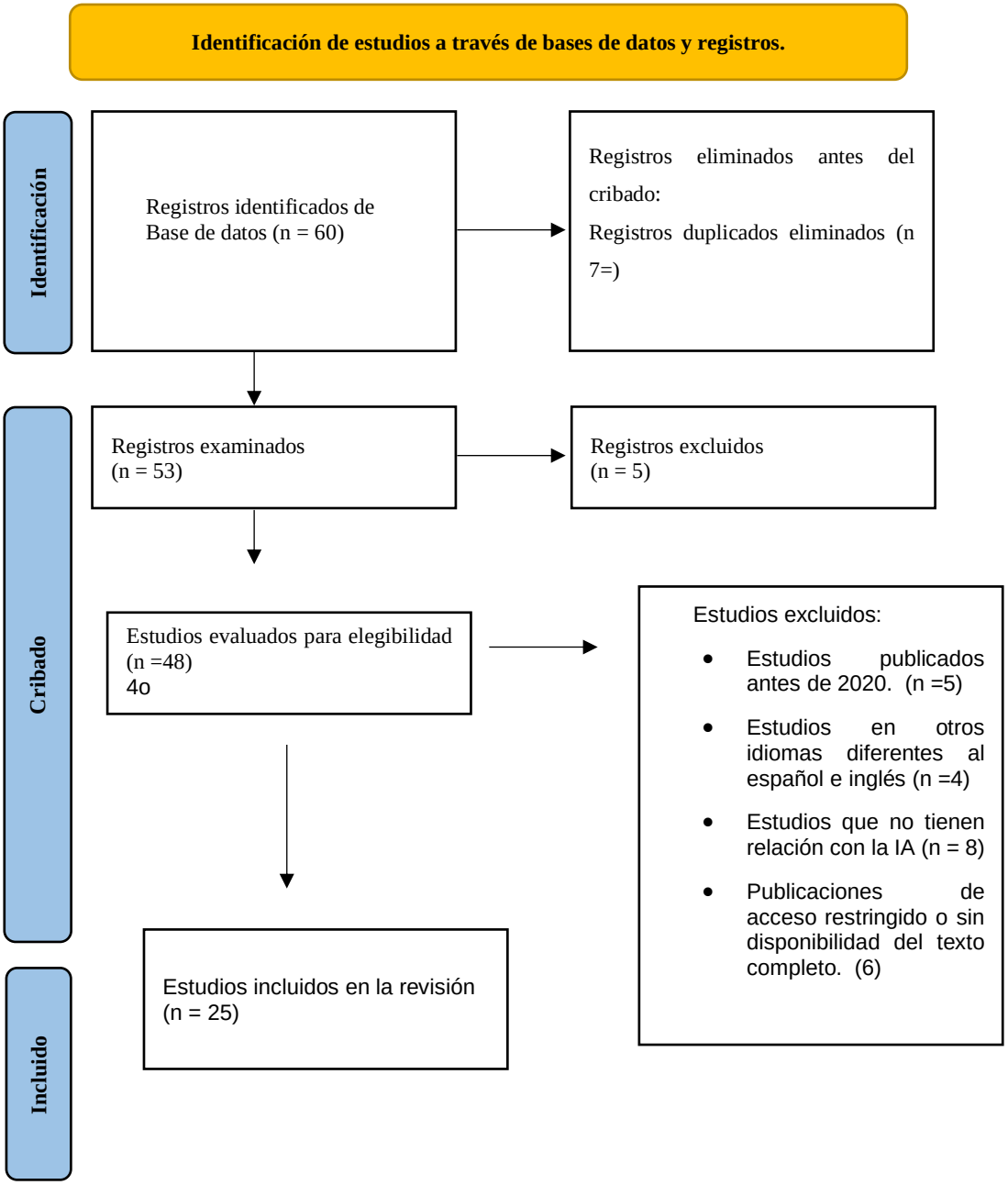
Resultados

En la figura 1, inicialmente se identificaron 60 registros en bases de datos. Después de eliminar duplicados, quedaron 53 registros que fueron examinados. De estos, 5 fueron excluidos, lo que dejó 48 estudios para ser evaluados en términos de elegibilidad. Durante este proceso, se descartaron 23 estudios adicionales debido a que no cumplían con ciertos criterios: 5 eran publicaciones anteriores a 2020, 4 estaban en otros idiomas distintos al español e inglés, 8 no estaban relacionados con IA y 6 tenían acceso restringido o no contaban con el texto completo. Finalmente, 25 estudios fueron incluidos en la revisión. De estos, 10 fueron utilizados para la

introducción y los 15 restantes se emplearon para los resultados, proporcionando una base sólida para el análisis y las conclusiones del trabajo.

Figura 1.

Diagrama de prisma



Fuente: Elaboración Propia.

Los resultados de la presente revisión sistemática se han dividido según el objetivo de la investigación. A continuación, se presentan los hallazgos.

Amenazas de seguridad en la integración de IA en sistemas de información

La integración de la inteligencia artificial (IA) en los sistemas de información introduce una serie de amenazas críticas, destacadas por diversos autores que subrayan los riesgos que estas tecnologías representan para la seguridad digital. Estas amenazas abarcan desde la manipulación de datos y ataques adversarios hasta el aprovechamiento malicioso de modelos avanzados de lenguaje y la generación automatizada de ciberataques.

Según Hongsong et al. (2024) una de las amenazas más preocupantes es la manipulación de datos de entrada, que puede comprometer la integridad de los modelos de IA al introducir información maliciosa en el sistema. Los ataques adversarios representan otra vulnerabilidad significativa, ya que los atacantes pueden realizar alteraciones mínimas en los datos para engañar a los modelos y generar decisiones erróneas. Asimismo, los sistemas de IA que operan en entornos distribuidos, como el aprendizaje automático en la nube (Drone-Cloud), son especialmente susceptibles a la intercepción de datos durante su transferencia, exponiendo información confidencial a actores maliciosos. Estas amenazas destacan la complejidad de mantener la seguridad en entornos altamente interconectados y automatizados.

Por otro lado, Barik et al. (2022), analizan cómo los ataques adversarios explotan la dependencia de los modelos de IA en grandes volúmenes de datos. Alterar estos datos de entrenamiento puede distorsionar el aprendizaje del modelo, comprometiendo su rendimiento y generando riesgos para la seguridad de los sistemas de información. Esta vulnerabilidad es especialmente relevante en aplicaciones críticas donde decisiones erróneas podrían tener consecuencias severas, como en la detección de fraudes o la seguridad biométrica.

En este mismo orden de ideas, Shahid & Imteaj, (2024) señalan que los modelos avanzados como los LLM (Large Language Models) presentan nuevas formas de amenazas debido a sus capacidades lingüísticas sin precedentes. Por ejemplo, los LLM pueden ser utilizados para generar correos electrónicos de phishing personalizados y altamente convincentes que imitan estilos lingüísticos específicos y eluden los filtros de seguridad tradicionales. Este tipo de ataques exacerba el riesgo de ingeniería social, explotando la confianza humana como un vector de ataque

eficaz. Además, su potencial para automatizar tareas complejas amplía el panorama de amenazas, permitiendo que incluso actores sin experiencia técnica puedan llevar a cabo ataques sofisticados. Por su parte, Boonkrong, (2023) destaca la amenaza de ataques de denegación de servicio (DDoS), que pueden interrumpir el funcionamiento normal de los sistemas de IA mediante el acceso no autorizado. Este tipo de ataque no solo compromete la disponibilidad del sistema, sino que también puede exponer datos sensibles durante el entrenamiento o uso del modelo. La interceptación de información privada o confidencial representa una vulnerabilidad crítica en la implementación de la IA, especialmente en sectores donde la protección de datos es esencial.

Guembe et al. (2022) proporcionan un análisis detallado de las fases de un ciberataque impulsado por IA, destacando que el 56% de las técnicas identificadas se concentran en la fase de acceso y penetración, mientras que el resto se distribuye en las fases de explotación, comando y control, reconocimiento y entrega. Este desglose resalta cómo la IA puede ser utilizada para orquestar ataques sofisticados en múltiples etapas, aprovechando su capacidad para automatizar procesos y evadir las defensas tradicionales.

Estos autores identifican un panorama alarmante de amenazas derivadas de la integración de la IA en los sistemas de información. Estas incluyen la manipulación de datos, la explotación de vulnerabilidades en algoritmos y la generación de ataques avanzados que amplían significativamente los riesgos de ciberseguridad. Estas amenazas subrayan la necesidad de desarrollar defensas más robustas y enfoques proactivos para mitigar los riesgos asociados con el creciente uso de la IA en entornos críticos.

Riesgos emergentes y vulnerabilidades críticas

La integración de la (IA) en sistemas de información ha generado riesgos emergentes y expuesto vulnerabilidades críticas que, si no son gestionadas adecuadamente, pueden comprometer la seguridad de las infraestructuras digitales. Estos riesgos se derivan de la complejidad de los entornos tecnológicos, la interconexión de sistemas y el creciente nivel de automatización que caracteriza a la IA.

Al respecto, Boretti, (2024) explora los riesgos específicos asociados con la integración de IA en sistemas cuánticos, señalando que la manipulación y gestión de cúbits representa un punto crítico de vulnerabilidad. Dado que los algoritmos de IA desempeñan un papel crucial en la corrección

de errores y la estabilidad de los cúbits, cualquier explotación de estas vulnerabilidades por actores malintencionados podría comprometer los cálculos y generar resultados erróneos. Este tipo de ataques puede tener implicaciones graves en áreas como la criptografía cuántica, donde la precisión y confiabilidad de los sistemas son esenciales.

En un contexto descentralizado, Almalawi et al. (2024), analiza los riesgos en la integración de IA de borde en sistemas de información distribuidos. La descentralización introduce amenazas adicionales, como los ataques de datos maliciosos que manipulan los patrones aprendidos por los modelos de codificación automática. Aunque mecanismos como la criptografía híbrida basada en tablas hash distribuidas (DHT) y algoritmos de optimización avanzados refuerzan la protección, estas medidas no son infalibles. Los atacantes pueden explotar vulnerabilidades en los modelos de IA para comprometer la integridad y privacidad de los datos.

Desde otra pespectiva, Ramos & Ellul, (2024) destacan la interconexión entre IA y blockchain como un área de oportunidad y riesgo. Si bien la tecnología blockchain se propone como una solución para garantizar la integridad de los datos y mejorar la trazabilidad, la combinación de estas tecnologías también introduce vulnerabilidades críticas. Los atacantes pueden explotar brechas en la implementación para manipular datos o comprometer la privacidad. Este hallazgo subraya la importancia de un diseño cuidadoso y una supervisión rigurosa en la integración de estas tecnologías.

A su vez, Hoang et al. (2024) presentan la automatización como un arma de doble filo en la ciberseguridad. Si bien la IA permite a los administradores de red identificar vulnerabilidades antes indetectables, también crea nuevas oportunidades para que los atacantes manipulen sistemas automatizados. Inserciones maliciosas de código o desactivaciones de defensas clave son algunos de los riesgos señalados, lo que resalta la importancia de implementar medidas de protección específicas para salvaguardar los sistemas automatizados.

Hashmi et al. (2024) asegura que una de las principales preocupaciones identificadas es la vulnerabilidad de los sistemas de IA a los ataques adversarios, que pueden comprometer la integridad y la confiabilidad de la protección de datos confidenciales. Reforzando lo anterior, (Salmon et al., 2024), evidencia que la introducción de la IA Incluye riesgos relacionados con la privacidad y seguridad de datos personales, vigilancia masiva y problemas de ciberseguridad como

los ciberataques a las redes eléctricas. En conjunto, estos autores destacan cómo la evolución de la IA, junto con su integración en sistemas distribuidos, cuánticos y descentralizados, ha generado un panorama de riesgos emergentes.

Estrategias de mitigación y marcos regulatorios

Para abordar las amenazas de seguridad y vulnerabilidades asociadas con la integración de la inteligencia artificial (IA) en los sistemas de información, diversos autores proponen estrategias de mitigación y enfoques regulatorios que buscan fortalecer la ciberseguridad y garantizar la confiabilidad de las tecnologías emergentes.

De acuerdo con Oluwafemi (2022), existe una necesidad de desarrollar criptografía post-cuántica como una medida proactiva para contrarrestar las amenazas futuras que plantea la tecnología cuántica. Este enfoque asegura que los sistemas de IA mantengan una línea de defensa sólida frente a la posibilidad de que los estándares actuales de cifrado sean superados. La criptografía post-cuántica se presenta como un componente esencial para preservar la integridad y confidencialidad de los datos en un entorno digital en rápida evolución.

En un contexto más amplio, Ramos & Ellul, (2024) proponen el uso de la tecnología blockchain como una herramienta para mitigar riesgos en sistemas de IA de alto riesgo. Blockchain ofrece soluciones como la mejora de la trazabilidad de los procesos, la creación de registros inmutables y la garantía de la integridad de los datos. Estas características permiten establecer una capa adicional de seguridad que protege tanto los datos como los modelos de IA frente a ataques maliciosos, especialmente en entornos altamente interconectados.

De igual forma, Mohamed, (2023), analiza el uso de la detección de anomalías basada en IA como una estrategia efectiva para identificar amenazas cibernéticas. Aunque el aprendizaje no supervisado ofrece un método poderoso para detectar comportamientos atípicos, también enfrenta desafíos como la generación de falsos positivos y la susceptibilidad al envenenamiento de datos. Mohamed subraya la necesidad de complementar estas herramientas con supervisión humana y mecanismos de validación para garantizar su eficacia y minimizar interrupciones innecesarias.

Por su parte, Radanliev, (2024), enfatiza la importancia de establecer normas globales y promover la cooperación internacional para garantizar el uso responsable de la IA en la seguridad de la información. Este enfoque regulatorio busca cerrar la brecha entre la innovación tecnológica y las

políticas de ciberseguridad, facilitando la creación de marcos legales que aborden cuestiones éticas, técnicas y sociales asociadas con la implementación de IA.

En el ámbito de los dispositivos IoT, Aziz et al. (2023) destacan varias estrategias de mitigación, como la implementación de mecanismos de control de acceso, protocolos de comunicación seguros y la aplicación regular de parches y actualizaciones. Estas medidas son esenciales para proteger los dispositivos conectados de vulnerabilidades que puedan ser explotadas por atacantes y garantizar la seguridad en entornos altamente automatizados.

Estas estrategias reflejan un enfoque integral para mitigar los riesgos asociados con la integración de la IA en sistemas de información. Desde soluciones técnicas, como la criptografía post-cuántica y el blockchain, hasta medidas operativas y regulatorias, como la detección de anomalías y la cooperación internacional, los autores destacan la necesidad de combinar innovación y supervisión para fortalecer la ciberseguridad en un entorno cada vez más automatizado y complejo.

Discusión

La revisión sistemática realizada ha puesto en evidencia una serie de desafíos y vulnerabilidades en la ciberseguridad asociada con la integración de la inteligencia artificial en los sistemas de información. A partir del análisis de múltiples estudios y artículos, emergen temas clave relacionados con las amenazas de seguridad, los riesgos emergentes y las estrategias de mitigación que deben ser abordados para fortalecer la protección de los datos y garantizar la confiabilidad de los sistemas. En esta sección, se discuten en profundidad estos hallazgos, comparándolos con investigaciones previas.

Un estudio realizado por Alanazi et al. (2023), resalta la relevancia de los protocolos de comunicación en las redes de IoT, destacando su impacto en aspectos como el consumo de energía, el ancho de banda, la latencia y la calidad del servicio, con un enfoque específico en la seguridad. Estos hallazgos se relacionan directamente con los resultados de la revisión sistemática, especialmente en lo referente a las amenazas de seguridad asociadas con la integración de la inteligencia artificial (IA) en sistemas de información, incluidos los dispositivos IoT.

Una conexión destacada entre ambos estudios es la vulnerabilidad que representan los protocolos de comunicación en los dispositivos IoT. Tanto Aziz et al. (2023) como los resultados de la

revisión recalcan que los protocolos de comunicación seguros son esenciales para mitigar riesgos como la manipulación de datos y la intercepción en tránsito, amenazas que también son discutidas por Hongsong et al. (2024) y Boonkrong (2023) en el contexto de la IA. Una selección inadecuada o la implementación deficiente de estos protocolos incrementa la exposición de los sistemas a ataques que comprometen su integridad y confidencialidad.

El estudio de Kumar, (2024), analiza los desafíos de privacidad y seguridad asociados con los servicios en la nube, destacando cómo la escalabilidad y el modelo de pago por uso en entornos multiusuario atraen a una amplia gama de sectores, incluidos la industria, la academia y la defensa. Sin embargo, este paradigma también introduce vulnerabilidades que los atacantes pueden explotar mediante métodos avanzados para acceder a información confidencial. Esta evidencia está en línea con los resultados obtenidos en la revisión sistemática, que también identifican las amenazas asociadas con la integración de la inteligencia artificial (IA) en entornos distribuidos como la nube. Una de las coincidencias más relevantes entre ambos estudios radica en el énfasis en los riesgos relacionados con la manipulación de datos y la explotación de vulnerabilidades del sistema. Kumar et al. (2024) muestran que los atacantes emplean técnicas modernas para comprometer sistemas en la nube, una preocupación similar a la planteada en la revisión, donde autores como Hongsong et al. (2024) y Boonkrong (2023) destacan la intercepción de datos en tránsito y los ataques adversarios en sistemas basados en IA. Ambos análisis coinciden en que la naturaleza multiusuario y la dependencia de grandes volúmenes de datos hacen que estos entornos sean objetivos prioritarios para los atacantes.

Otra conexión importante entre ambos estudios es la necesidad de proteger los datos confidenciales en sistemas basados en IA. Kumar et al. (2024) señalan que los atacantes pueden explotar brechas en la seguridad para obtener acceso no autorizado a información sensible, un problema que también es resaltado por Shahid e Imteaj (2024) en el contexto de los modelos avanzados de lenguaje, donde la generación de phishing y la manipulación de datos son amenazas frecuentes.

En definitiva, tanto el estudio de Kumar et al. (2024) como los resultados de la revisión sistemática identifican riesgos críticos en entornos distribuidos como la nube y sistemas de IA. Ambos trabajos coinciden en la urgencia de desarrollar soluciones específicas para proteger la privacidad y la

seguridad en estos entornos, destacando la importancia de medidas como la implementación de criptografía robusta, la detección de anomalías y el monitoreo en tiempo real.

Conclusiones

La revisión sistemática permitió identificar y analizar las principales amenazas de seguridad asociadas con la integración de la inteligencia artificial en los sistemas de información. Entre las vulnerabilidades más destacadas se encuentran la manipulación de datos, los ataques adversarios, la generación automatizada de ciberataques y las brechas en sistemas distribuidos como la nube y dispositivos IoT. Estas amenazas exponen la fragilidad de los sistemas en un entorno tecnológico altamente automatizado e interconectado.

Además, se identificaron riesgos emergentes derivados de la complejidad tecnológica y la dependencia de la IA, que amplifican el panorama de vulnerabilidades. Para mitigar estos riesgos, las estrategias propuestas incluyen la implementación de criptografía avanzada, tecnologías como blockchain y detección de anomalías, así como la necesidad de establecer marcos regulatorios y fomentar la cooperación internacional. Estos hallazgos subrayan la urgencia de adoptar medidas integrales y proactivas para fortalecer la ciberseguridad y garantizar la confiabilidad de los sistemas en un entorno cada vez más dependiente de la IA.

Referencias

Alanazi, M., Mahmood, A., & Chowdhury, M. J. M. (2023). SCADA vulnerabilities and attacks: A review of the state-of-the-art and open issues. In *Computers and Security* (Vol. 125). Elsevier Ltd. https://doi.org/10.1016/j.cose.2022.103028

Almalawi, A., Hassan, S., Fahad, A., & Khan, A. I. (2024). A Hybrid Cryptographic Mechanism for Secure Data Transmission in Edge AI Networks. *International Journal of Computational Intelligence Systems*, *17*(1). https://doi.org/10.1007/s44196-024-00417-8

Aslan, Ö., Aktuğ, S. S., Ozkan-Okay, M., Yilmaz, A. A., & Akin, E. (2023). A Comprehensive Review of Cyber Security Vulnerabilities, Threats, Attacks, and Solutions. In *Electronics (Switzerland)* (Vol. 12, Issue 6). MDPI. https://doi.org/10.3390/electronics12061333

Aziz Al Kabir, M., Elmedany, W., & Sharif, M. S. (2023). Securing IoT Devices Against Emerging Security Threats: Challenges and Mitigation Techniques. *Journal of Cyber Security Technology*, *7*(4), 199–223. https://doi.org/10.1080/23742917.2023.2228053

Barik, K., Misra, S., Konar, K., Fernandez-Sanz, L., & Koyuncu, M. (2022). Cybersecurity Deep: Approaches, Attacks Dataset, and Comparative Study. In *Applied Artificial Intelligence* (Vol. 36, Issue 1). Taylor and Francis Ltd. https://doi.org/10.1080/08839514.2022.2055399

Boonkrong, S. (2023). Attack Model for Generic Intelligent Systems. *Journal of Applied Security Research*. https://doi.org/10.1080/19361610.2023.2283666

Boretti, A. (2024). Technical, economic, and societal risks in the progress of artificial intelligence driven quantum technologies. *Discover Artificial Intelligence*, *4*(1). https://doi.org/10.1007/s44163-024-00171-y

Corchado, J. M. (2023). *Gestión de crisis mediante la utilización de IA*. https://produccioncientifica.usal.es/documentos/66c62b7ab946137e28efc52b

Erazo-Luzuriaga, A. F., Ramos-Secaira, F. M., Galarza-Sánchez, P. C., & Boné-Andrade, M. F. (2023). La inteligencia artificial aplicada a la optimización de programas informáticos. *Journal of Economic and Social Science Research*, *3*(1), 48–63. https://doi.org/10.55813/gaea/jessr/v3/n1/61

Flores-Cedeño P, Zambrano-Pilay, E., & Chiriboga-Mendoza, F. (2023). *SEGURIDAD INFORMÁTICA E INTELIGENCIA ARTIFICIAL EN LA INVESTIGACIÓN CIENTÍFICA*. *7*(13). https://www.journalingeniar.org/index.php/ingeniar/article/view/177/250

Garcia-Pomareda, J. D. (2024). Inteligencia artificial generativa: un arma de doble filo para el metaverso. *Revista E-Mercatoria*, *23*(2), 295–323. https://doi.org/10.18601/16923960.v23n2.09

Guaña-Moya, J. (2023). La importancia de la seguridad informática en la educación digital: retos y soluciones. *RECIMUNDO*, *7*(1), 609–616. https://doi.org/10.26820/recimundo/7.(1).enero.2023.609-616

Guembe, B., Azeta, A., Misra, S., Osamor, V. C., Fernandez-Sanz, L., & Pospelova, V. (2022). The Emerging Threat of Ai-driven Cyber Attacks: A Review. In *Applied Artificial Intelligence* (Vol. 36, Issue 1). Taylor and Francis Ltd. https://doi.org/10.1080/08839514.2022.2037254

Hashmi, E., Yamin, M. M., & Yayilgan, S. Y. (2024). Securing tomorrow: a comprehensive survey on the synergy of Artificial Intelligence and information security. *AI and Ethics*. https://doi.org/10.1007/s43681-024-00529-z

Hoang, V. Q., La, V. P., Nguyen, H. S., & Nguyen, M. H. (2024). Some discussions on critical information security issues in the artificial intelligence era. *AI and Society*. https://doi.org/10.1007/s00146-024-02023-w

Hongsong, C., Yongpeng, Z., Yongrui, C., & Bhargava, B. (2021). Security Threats and Defensive Approaches in Machine Learning System Under Big Data Environment. *Wireless Personal Communications*, *117*(4), 3505–3525. https://doi.org/10.1007/s11277-021-08284-8

Jara-Moya, S., Torres-Valverde, L., & Torres-Abril, P. (2023). Evolución de los motores de búsqueda y la revolución con la inteligencia artificial. *Pol. Con*, *8*(9). https://polodelconocimiento.com/ojs/index.php/es

Kumar, S. M. D. y M. K. (2024). Una revisión integral de las vulnerabilidades y la defensa habilitada por IA contra ataques DDoS para proteger los servicios en la nube. *Revisión de Ciencias de La Computación*, *53*(1).

Mohamed, N. (2023). Current trends in AI and ML for cybersecurity: A state-of-the-art survey. *Cogent Engineering*, *10*(2). https://doi.org/10.1080/23311916.2023.2272358

Oluwafemi Kunle-Lawanson. (2022). The role of AI in information security risk management. *World Journal of Advanced Engineering Technology and Sciences*, *7*(2), 308–319. https://doi.org/10.30574/wjaets.2022.7.2.0128

Otilia Mosquera-Chere, S. I. (2021). *La vinculación entre la inteligencia artificial y la seguridad cibernética en el Ecuador*. *6*(2), 1154–1173. https://doi.org/10.23857/pc.v6i2.2430

Radanliev, P. (2024). Cyber diplomacy: defining the opportunities for cybersecurity and risks from Artificial Intelligence, IoT, Blockchains, and Quantum Computing. *Journal of Cyber Security Technology*, 1–51. https://doi.org/10.1080/23742917.2024.2312671

Ramos, S., & Ellul, J. (2024). Blockchain for Artificial Intelligence (AI): enhancing compliance with the EU AI Act through distributed ledger technology. A cybersecurity perspective. *International Cybersecurity Law Review*, *5*(1), 1–20. https://doi.org/10.1365/s43439-023-00107-9

Salmon, P. M., King, B. J., Elstak, I., McLean, S., & Read, G. J. M. (2024). Tomorrow's demons: a scoping review of the risks associated with emerging technologies. In *Ergonomics*. Taylor and Francis Ltd. https://doi.org/10.1080/00140139.2024.2416554

Shahid, A. R. Bin, & Imteaj, A. (2024). Sticks and stones may break my bones, but words will never hurt me!—Navigating the cybersecurity risks of generative AI. *AI and Society*. https://doi.org/10.1007/s00146-024-01934-y

Shehab, M., Abualigah, L., Jarrah, M. I., Alomari, O. A., & Daoud, M. S. (2020). Artificial Intelligence in Software Engineering and inverse: Review. *International Journal of Computer Integrated Manufacturing*, *33*(10–11), 1129–1144. https://doi.org/10.1080/0951192X.2020.1780320

Uribe, L. (2024). *Integración de Inteligencia Artificial en la gestión de tecnologías de la información: un enfoque aplicado en el desarrollo empresarial*. 1–10. https://doi.org/10.26507/paper.3761

Conflicto de intereses:

Los autores declaran que no existe conflicto de interés posible.

Financiamiento:

No existió asistencia financiera de partes externas al presente artículo.

Agradecimiento:

N/A

Nota:

El artículo no es producto de una publicación anterior.