Riesgos para la Infraestructura Crítica Cibernética del Sector Defensa: implementación de inteligencia artificial e identificación con metodología FAIR.

Mayor (My) CARLOS ANDRES REY CASTIBLANCO

Artículo para optar al título profesional:

Magister en Ciberseguridad y Ciberdefensa

Escuela Superior de Guerra "General Rafael Reyes Prieto"
Bogotá D.C., Colombia
2025

Escuela Superior de Guerra "General Rafael Reyes Prieto"
Bogotá D.C., Colombia

DATOS GENERALES		
Nombre del estudiante	**:**	Mayor (EJC) Carlos Andrés Rey Castiblanco
Identificación	**:**	80138196
Programa académico	**:**	Maestría en Ciberseguridad y Ciberdefensa
Tutor metodológico	**:**	DR. Jairo Andrés Becerra Cuervo
Tutor temático	**:**	DR. Lucas Giraldo
Fecha de entrega	**:**	28 de Septiembre
Extensión	**:**	6.410 palabras

DECLARACIÓN DE ORIGINALIDAD Y CESIÓN DE DERECHOS

El autor declara que este artículo fue escrito de acuerdo con la normatividad de la Escuela Superior de Guerra "General Rafael Reyes Prieto" (ESDEG) y no existe ningún potencial conflicto de interés relacionado con este. Las posturas y aseveraciones presentadas son resultado de un ejercicio académico e investigativo que no representan la posición oficial ni institucional de la ESDEG, las Fuerzas Militares de Colombia o el Ministerio de Defensa Nacional.

Este artículo es enteramente mi propio trabajo y no ha sido presentado para la obtención de un título en esta u otra Institución de Educación Superior. Se han referenciado todos los trabajos y puntos de vista de otros autores, así como los datos de otras fuentes utilizadas. No se emplearon herramientas de generación de contenido por Inteligencia Artificial para su elaboración.

El autor acepta ceder los derechos de publicación en favor de la ESDEG y su Sello Editorial de acuerdo con los términos de la licencia Creative Commons: Reconocimiento-NoComercial-SinObrasDerivadas.

AUTORIZACIÓN DE PUBLICACIÓN

El autor autoriza que este artículo sea publicado por el Sello Editorial ESDEG en su repositorio institucional y esté disponible bajo una modalidad de acceso abierto.

Escuela Superior de Guerra "General Rafael Reyes Prieto"
Bogotá D.C., Colombia

Riesgos para la Infraestructura Crítica Cibernética del Sector Defensa: implementación de inteligencia artificial e identificación con metodología FAIR.

Risks to the Cyber Critical Infrastructure of the Defense Sector: Implementation of Artificial Intelligence and Identification with FAIR Methodology.

MY. Carlos Andrés Rey Castiblanco [1]
Escuela Superior de Guerra "General Rafael Reyes Prieto"

Resumen: El aumento de ciberamenazas en Colombia, especialmente aquellas vinculadas a inteligencia artificial (IA), plantea riesgos significativos para la infraestructura crítica cibernética del sector defensa. Este estudio cualitativo y descriptivo analiza cómo la IA incrementa la complejidad y persistencia de los ataques cibernéticos, destacando desafíos en la prevención y mitigación de riesgos. Se emplearon metodologías como FAIR y triangulación conceptual para identificar vulnerabilidades, evaluar impactos y formular estrategias de mitigación basadas en amenazas persistentes avanzadas. Los resultados evidencian que la IA no solo amplifica los riesgos al automatizar ataques, sino que también requiere enfoques estratégicos de ciberseguridad que incluyan actualización tecnológica y análisis anticipado de riesgos. La investigación concluye que es esencial implementar procesos robustos de gestión del riesgo y metodologías preventivas para proteger la infraestructura crítica del sector defensa frente a amenazas cibernéticas impulsadas por IA.

Palabras clave: Ciberseguridad, inteligencia artificial, ciberamenazas, infraestructura crítica, mitigación, defensa

Abstract: The rise of cyber threats in Colombia, particularly those linked to artificial intelligence (AI), poses significant risks to the critical cyber infrastructure of the defense sector. This qualitative and descriptive study examines how AI increases the complexity and persistence of cyberattacks, highlighting challenges in risk prevention and mitigation. Methodologies such as FAIR and conceptual triangulation were employed to identify vulnerabilities, assess impacts, and propose mitigation strategies based on advanced persistent threats. Results show that AI not only amplifies risks by automating attacks but also demands strategic cybersecurity approaches involving technological updates and anticipatory risk analysis. The study concludes that robust risk management processes and preventive methodologies are essential to safeguard the critical infrastructure of the defense sector against AI-driven cyber threats.

Keywords: Cybersecurity, artificial intelligence, cyberthreats, critical infrastructure, mitigation, defense.

[1] Oficial del Ejército Nacional; candidato a magister en ciber seguridad y ciber defensa. https://orcid.org/0009-0003-2103-6918

Introducción

El aumento en el número de detecciones de ciber amenazas sobre contexto colombiano para 2024 fue notable, registrando un incremento del 38.6% en comparación con 2023, y del 162% respecto a 2021 y 2022 (Comando Cibernético de las Fuerzas Militares, 2024).

Dicho crecimiento evidencia que la rápida expansión de ciberamenazas en el ámbito digital colombiano es un fenómeno estructural que expone desafíos tanto conceptuales como epistemológicos. Estos últimos, producto de las brechas de conocimiento técnico – específico que ameritan la actualización constante en los sistema de defensa y/o seguridad cibernética (Vakulyk, Petrenko, Kuzmenko, Pochtovyi, y Orlovskyi, 2020).

Entre los años 2020 y 2025, las investigaciones científicas en el campo de la ciberseguridad han girado en torno a dos categorías predominantes: la infraestructura crítica cibernética y la inteligencia artificial.

Ello, ha constituido un núcleo científico de estudio que en contexto nacional se encuentra reflejado en las políticas de seguridad digital emitidas a través de Consejos de Política Económica y Social.

Ambas áreas – infraestructura cibernética e inteligencia artificial- son consideradas disruptivas por su naturaleza tendencial, y presentan vacíos significativos en términos procedimentales, estratégicos y tecnológicos, lo que las convierte en un marco técnico para la exploración científica centrada en la identificación de amenazas como ransomware, que cifra sistemas esenciales para exigir rescates, ataques de denegación de servicio (DDoS) que saturan redes para interrumpir operaciones, y malware diseñado para sabotear o extraer información (Pătrașcu, 2019).

El debate acerca de la infraestructura crítica cibernética (ICCN) y la inteligencia artificial (IA), constituye un punto clave de atención y desarrollo para la ciberseguridad de sistemas que son parte de la estrategia nacional de desarrollo.

Ante esa perspectiva, y en caso colombiano, la relación de ICCN e IA ha evolucionado progresivamente desde el CONPES 3701[*] de 2011, pasando por el CONPES 3854[†] de 2020, hasta llegar al CONPES 4144 de 2025.

Este último documento destaca la importancia de integrar la inteligencia artificial en un modelo de gobernanza digital que abarque tanto a la ciudadanía como al Estado, consolidando un enfoque estratégico para enfrentar los retos en entornos cibernéticos.

Bajo el argumento previo, es importante conectar la problemática con el aumento exponencial de ciber amenazas que de acuerdo con el boletín nº 24 del Comando Conjunto de Ciberdefensa de las Fuerzas Militares, encuentra en los ransomware fenómenos de afectación cuya configuración transmuta a partir de la creación colectiva de códigos maliciosos.

Esa creación colectiva o conjunta depende de un eje transversal: el conocimiento técnico en campos como la programación compleja y la creación de software.

Ambos elementos, programación compleja y creación de software, aceleran la evolución de nuevas amenazas digitales que resultan ser desconocidas para la estrategia de ciberdefensa nacional basada en el análisis de problemas de contexto, y no en la construcción de escenarios de futuro que permitan la identificación de nuevas fenomenologías criminales digitales.

Aunado a lo anterior, el auge de códigos de infección diseñados exclusivamente para infraestructura cibernética de tipología pública, diseñados con inteligencia artificial, dinamiza la materialización de riesgos de afectación y transgresión, pues genera ralentización sobre los procesos de seguridad digital, ya que la estrategia pública no siempre posee procesos de actualización conexos, acordes o alineados con la genealogía delictiva digital de las tendencias derivadas del marco general de terrorismo cibernético.

Si bien no hay reportes oficiales específicos que permitan conocer de forma cuantitativa la cantidad de afectaciones a infraestructura cibernética por amenazas digitales,

[*] Este CONPES trae consigo la creación de una política de seguridad digital conformada por actores de gobierno. Allí se determinan los objetivos estratégicos relacionados con la protección de activos digitales nacionales.

[†] Este CONPES hace alusión al análisis del riesgo digital a partir de la identificación de estructuras vulnerables.

la literatura pública, técnica, académica y de fuentes de información, corroboran que las vulnerabilidades presentes dependen de la ausencia de procesos de gestión para el estudio e identificación anticipada de amenazas duales, es decir, cibernéticas pero con objetivos e intenciones políticas, religiosas o sectarias.

Para la identificación temprana, un proceso de gestión del riesgo que parta con su análisis y exploración facilitaría la construcción de hipótesis de ciber ataque, pero sobre todo, el análisis temprano de sus elementos tecnológicos.

Siendo así, la identificación de posibles amenazas tendría que comenzar con el estudio conceptual y estructural de riesgos cibernéticos, ya que:

- Primero, actualmente, el decreto 338 de 2022 es enfático al establecer en el numeral 2.2.21.1.4.1., que el levantamiento de infraestructura crítica cibernética es una responsabilidad de los actores del Estado, y sobre todo de aquellos que son parte del modelo de gobernanza digital. Entre las condiciones para su análisis está la identificación de riesgos que se aumenta contra este tipo de infraestructura, y que resulta ser producto de la intervención técnica conexa con inteligencia artificial.
- Segundo, la identificación de riesgos representa un proceso estructural necesario para teorizar las contribuciones a las disciplinas del conocimiento ciencias de la computación y ciencias multidisciplinares. No obstante, a 2025 como se demuestra en la construcción conceptual, hay vacíos de conocimiento relacionados con el crecimiento de ciber amenazas complejas que superan en capital intelectual al Estado.
- Tercero, no hay metodologías claras para estudiar la posible afectación de amenazas cibernéticas configuradas con inteligencia artificial a infraestructura digital en el sector defensa.

Son estas tres causas las que llevan el proceso de investigación a desarrollar un análisis de enfoque cualitativo que genere resultados validados para responder a este interrogante de investigación: ¿Cómo el surgimiento de la inteligencia artificial aumenta los riesgos y afectaciones técnicas sobre la infraestructura crítica cibernética perteneciente al sector Defensa?

La respuesta a esta pregunta se planteó en tres partes. La primera, estudiar la influencia de la Inteligencia Artificial en la Ciberseguridad del Sector Defensa a partir de la identificación de Retos y Oportunidades.

La segunda, identificar los riesgos y vulnerabilidades en la ICCN del sector defensa en relación con tecnologías de evasión y persistencia. La tercera, establecer las consecuencias potenciales de la posible generación de amenazas en contra de ICCN utilizando la metodología FAIR (**Factor Analysis of Information Risk)**.

La cuarta y última parte corresponde a la configuración de una propuesta para mitigar el impacto de posibles ataques cibernéticos a ICCN, tomando como base la metodología de amenazas persistentes avanzadas.

Metodología

Este trabajo es de enfoque cualitativo y diseño descriptivo, y para su desarrollo se llevan a cabo tres procesos. En el primero, identificar los riesgos y vulnerabilidades en la infraestructura crítica cibernética del sector defensa con probabilidad de uso mediante inteligencia artificial conexa a tecnologías de evasión y persistencia.

La técnica empleada en este caso es la revisión de fuentes de información con artículos de investigación cuyo enfoque mixto permita establecer tanto tecnologías como procesos empleados para la protección de infraestructura crítica cibernética.

En el segundo, evaluar los impactos potenciales de las amenazas cibernéticas creadas con inteligencia artificial en contra de la infraestructura crítica cibernética del sector defensa utilizando la metodología FAIR. En esta parte la investigación busca establecer esos impactos, a partir de un factor metodológico que permitiría un primer acercamiento al relacionamiento de impactos y afectaciones por sintaxis de código complejas.

En el tercero, formular propuestas de mitigación orientadas a la prevención de ataques tomando como base argumentativa el concepto teórico de las amenazas persistentes avanzadas (APT). Para tal fin, se aplicará una técnica de triangulación teórico – conceptual, a fin de entender cómo los impactos se disipan, expanden y/o evolucionan.

Cabe destacar que el proceso metodológico de la investigación se ciñe a la contribución técnica de Hernández et al (2014).

Impacto de la Inteligencia Artificial en la Ciberseguridad del Sector Defensa: Retos y Oportunidades.

La evolución de la inteligencia artificial (IA) ha generado impactos transformadores en múltiples sectores, incluido el campo de ciberseguridad aplicado a la defensa. En particular, la inclusión de IA en la infraestructura crítica cibernética del sector defensa ha traído elementos favorables en términos de eficiencia operativa, automatización y predicción de ciberataques.

Sin embargo, esta evolución también ha producido una serie de riesgos inherentes que comprometen la seguridad y resiliencia de las infraestructuras críticas cibernéticas; las anteriores, altamente indispensables para el desarrollo multidimensional e intersectorial del sector defensa.

Uno de los problemas principales es el alcance en la superficie de ataque, fenómeno procedente de la creciente complejidad y dependencia de sistemas basados en IA.

La automatización de procesos críticos, si bien mejora la capacidad de respuesta y optimización de recursos, introduce vulnerabilidades complejas, no conocidas en sintaxis, que son explotadas en el margen de actividades cibernéticas en contra de la estructura de defensa cibernética del país.

Entre estas amenazas destacan los ciberataques creados con inteligencia artificial, diseñados específicamente para aprovechar las debilidades algorítmicas de los sistemas de información.

Esas técnicas incluyen la inyección de datos para manipular modelos predictivos, y generar resultados erróneos, así como para explotar vulnerabilidades en algoritmos de aprendizaje automático, que podrían comprometer la integridad de las infraestructuras críticas de la nación. Sobre todo, las del sector defensa.

Bajo este contexto, la infraestructura crítica cibernética del sector defensa enfrenta un desafío dual: por un lado, garantizar la adopción de procesos metodológicos de ciber seguridad para establecer capacidades estratégicas de tipología cibernética.

Por el otro, reducir los riesgos asociados a estas mismas tecnologías. La amenaza deriva de múltiples actores, y se materializa con ciber ataques de naturaleza diferente.

Es así, como, por ejemplo, que, durante el 2024, de acuerdo con el CCOCI (2024), el número de ciberataques aumentó un 22% en contra de la infraestructura crítica cibernética en general. Si bien no es comprobable el factor de crecimiento exponencial de ataque cibernéticos por causa o razón conexa con el empleo de inteligencia artificial, este sí es un vector de análisis que se ha revisado con anterioridad en las investigaciones de Guembe *et al* (2022) y Chakraborty *et al* (2023).

De hecho, este grupo de autores ha constituido una idea clara acerca de la inclusión de IA al desarrollo de ciber ataques en contra de infraestructura crítica. La versión de Guembe *et al* (2022) y Chakraborty *et al* (2023) plantea que el diseño y producción de ciberataques con IA aumenta el flujo numérico de acciones, mientras que al tiempo constituye nuevas formas de impacto y penetración, así como complejidad técnica en la sintaxis.

Una interpretación similar, y que también se ajusta a contexto colombiano proviene de Mosteanu (2020), quien expone que la rápida transformación en la modalidad de ataques incluyendo la sintaxis y las formas de operación del código, incrementa la efectividad de ciber ataques desarrollados con inteligencia artificial.

El caso colombiano no es indiferente a la versión de Mosteanu (2020), si se tiene en cuenta que los ciber ataques complejos van en aumento. Según CCOCI (2024), en materia de ciber ataques con IA en Colombia (probables), se han detectado Ransomware (hasta 4,500 detecciones), *exploits* (más de 1,000 detecciones) y amenazas web (superando las 110,000 detecciones). Estos ataques, como el uso de *exploits* avanzados (CVE-2021-44228) o redes saturadas por Flood ICMP (más de 400,000 detecciones), representan un riesgo crítico para infraestructuras esenciales. La inteligencia artificial amplifica estos riesgos al automatizar ataques y optimizar la explotación de vulnerabilidades, exigiendo defensas que se configuren bajo el concepto técnico de prevención y anticipación (Abdullahi *et al*, 2022).

Otra versión ajustada al contexto del sector defensa, proviene de Yamin *et al* (2021), quienes toman como punto de partida la utilización de inteligencia artificial en el marco de ciber ataques configurados a partir de la codificación rápida de sintaxis desconocidas para los modelos tecnológicos y procedimentales utilizados en el marco de la ciber defensa.

La versión de Yamin *et al* (2021) permite comprender el tema de ciber ataques, a través de un planteamiento poco explorado: utilización de inteligencia artificial como un

arma de ciber ataques, cuyo propósito es la disrupción técnica de infraestructuras críticas cibernéticas del sector defensa. De ahí, que en el Plan de Campaña Ayacucho 2.0 se hayan incluido dos objetivos intermedios correlacionados con ciber seguridad y restricción de amenazas cibernéticas en zonas de injerencia militar.

Las versiones expuestas hasta acá se complementan con la contribución de Zhange *et al* (2022), quien expone que n la inteligencia artificial facilita la implementación de ciber ataques tipo como los malware inteligentes y los ataques adversariales con machine learning.

El panorama descrito trae consigo una necesidad: implementar enfoques robustos de ciberseguridad con medidas preventivas y reactivas, que aseguren capacidades de anticipación y respuesta eficaz frente a las posibles amenazas que presenta la inclusión de inteligencia artificial al campo del ciber terrorismo.

Frente a tal necesidad, contribuciones como las de Alhayani *et al* (2021) llevan al estudio de procesos, métodos y procedimientos empleados para analizar la probabilidad de riesgo conectado con la materialización de ciber ataques a infraestructuras críticas, producto de inteligencia artificial.

La versión de Alhayani *et al* (2021) es apropiada en el contexto colombiano porque presenta ante el escenario académico una revisión de literatura en la que existe amplia convergencia acerca de: primero, las técnicas de inteligencia artificial empleadas para denegar ciber ataques y segundo, utilización de IA para restringir impactos a partir de medidas de prevención.

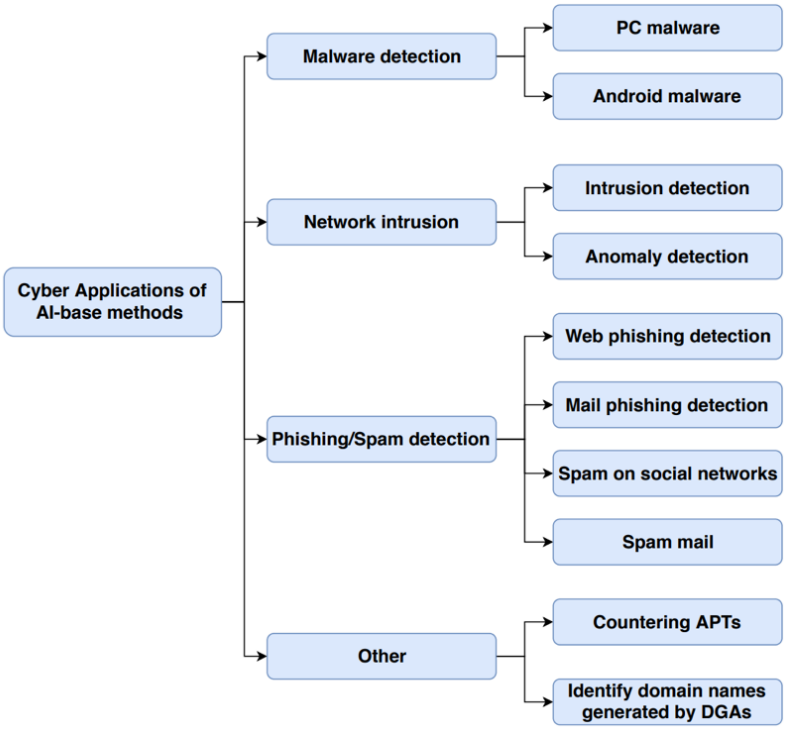
De hecho, como Alhayani *et al* (2021), Zhang *et al* (2022) mencionan de igual forma la construcción conceptual de medidas de prevención de ciber ataques que trae consigo la inteligencia artificial pero que también son un producto defensivo de la misma. Es decir, la inteligencia artificial en materia de ciber ataques o ciber seguridad.

Ambas versiones, la de Alhayani *et al* (2021) y Zhang et al (2022) constituyen un primer acercamiento al tema no explorado en contexto colombiano: la utilización de metodologías de prevención para la mitigación de riesgos, aunado a la implementación de factores tecnológicos que conlleven a la rápida actualización de los sistemas de defensa cibernética.

Al respecto de esa necesidad de actualización, Truong *et al* (2020) explican que en el marco del ciber dominio, la ciber seguridad a constituir debe obedecer al análisis del riesgo, su gestión y diseño estratégico ajustado al escenario cibernético.

Esto significa, tal y como surge en el contexto de la infraestructura cibernética del sector defensa, que un proceso estratégico para la protección anticipada surge del análisis estructural y funcional del riesgo, su naturaleza y sus formas de materialización. Esos riesgos y esos elementos de materialización en contra de la infraestructura cibernética son observables en la figura 1:

Figura 1. Principales brechas de ciber seguridad que surgen por la aplicación e inclusión de inteligencia artificial.



Nota: información recuperada de Truong *et al* (2020)

A la versión de Truong *et al* (2020), que se centra en el análisis de riesgo, se suma la versión de Chehri *et al* (2020), la cual configura un modelo para la prevención del riesgo, tomando como punto clave el diseño de propuestas metodológicas para la reducción de afectación e impactos.

También, Maddireddy *et al* (2020) y Kure *et al* (2020) se centran en la prevención del riesgo cibernético diseñado con IA, pero con mayor inclinación a la construcción estratégica de medidas preventivas y de anticipación. De ahí que ambas versiones conlleven al reconocimiento de una necesidad técnica: implementación de medidas de evaluación para la identificación temprana de ciber ataques posibles a infraestructuras críticas del sector defensa para el caso.

Sin embargo, este tipo de metodologías o acciones estratégicas de prevención temprana son objeto de estudio, y la producción nacional de conocimiento para mejorar la orientación del enfoque de ciber defensa constituye un reto conceptual y técnico en el caso del sector defensa.

Por ello, el diseño de estrategias para restringir o denegar impactos a la infraestructura crítica y cibernética militar amerita, no solo diseñar un protocolo funcional pues esta es una tarea impuesta por la evolución del elemento IA, sino empezar con el reconocimiento, exploración e identificación básica y avanzada de las afectaciones que trae consigo el diseño de ciberataques hacia infraestructura cibernética con vectores asociados con IA.

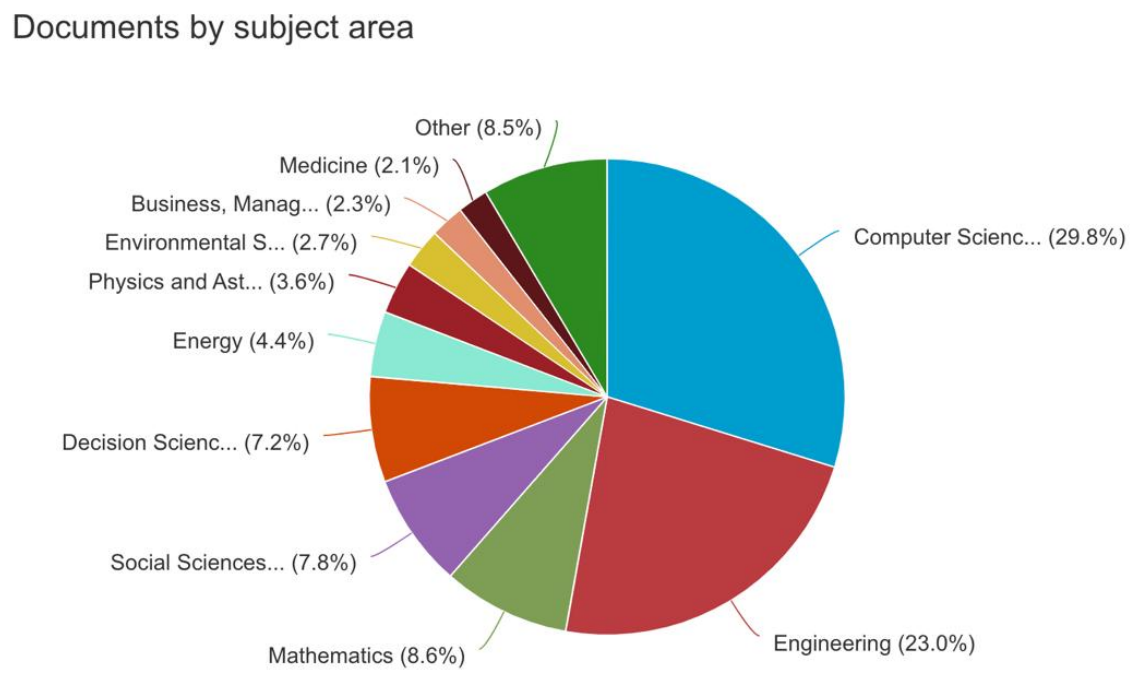
Riesgos y vulnerabilidades en la ICCN del sector defensa; relación con tecnologías de evasión y persistencia.

Analizar el impacto de la inteligencia artificial en el campo de los ataques cibernéticos hacia infraestructura crítica cibernética, permite establecer parámetros exploratorios orientados a la identificación formal de riesgos y vulnerabilidades.

Para tal fin, se presenta en esta parte de la investigación un proceso analítico y exploratorio que tiene por objetivo establecer riesgos y vulnerabilidades a través de una técnica de identificación de contribuciones allegadas a las categorías del problema.

El análisis se desarrolló con las bases de datos SCOPUS y Web of Science, y la descripción de sus resultados se expone a continuación:

Figura 2. Áreas de publicación a 2025

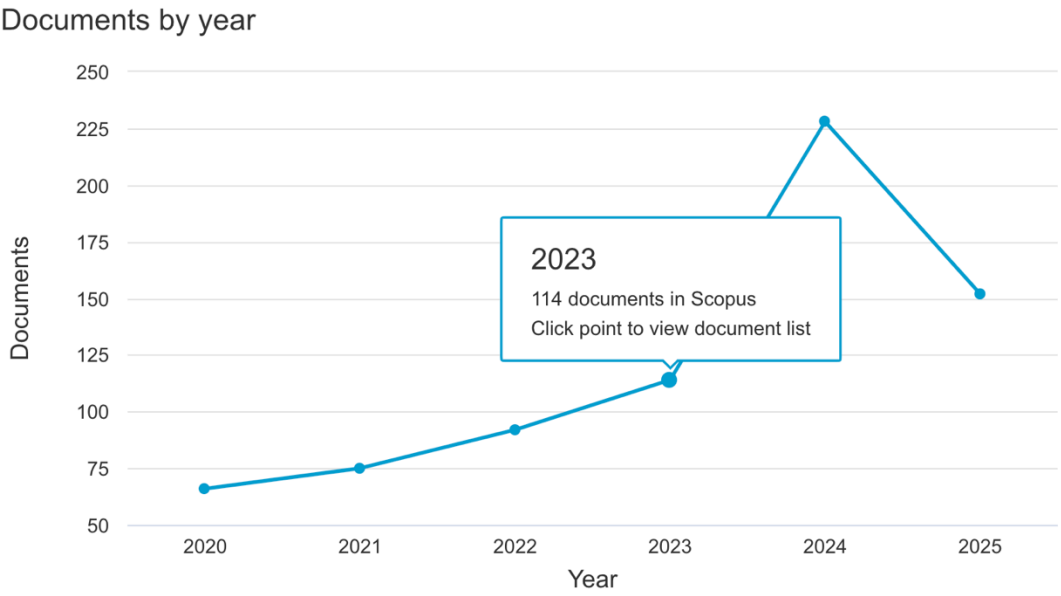


Nota: información recuperada de SCOPUS (2025)

De acuerdo con el ejercicio de revisión de áreas de conocimiento, las ciencias de la computación (29,8%) y las ingenierías (23%) ocupan el segmento con mayor publicación científica.

El patrón particular entre esas áreas para gestionar conocimientos se ubica en el lapso temporal de 2020 a 2025, siendo 2023 y 2024 los años con más exposición científica indexada. (Ver figura 3):

Figura 3. Años con mayor publicación

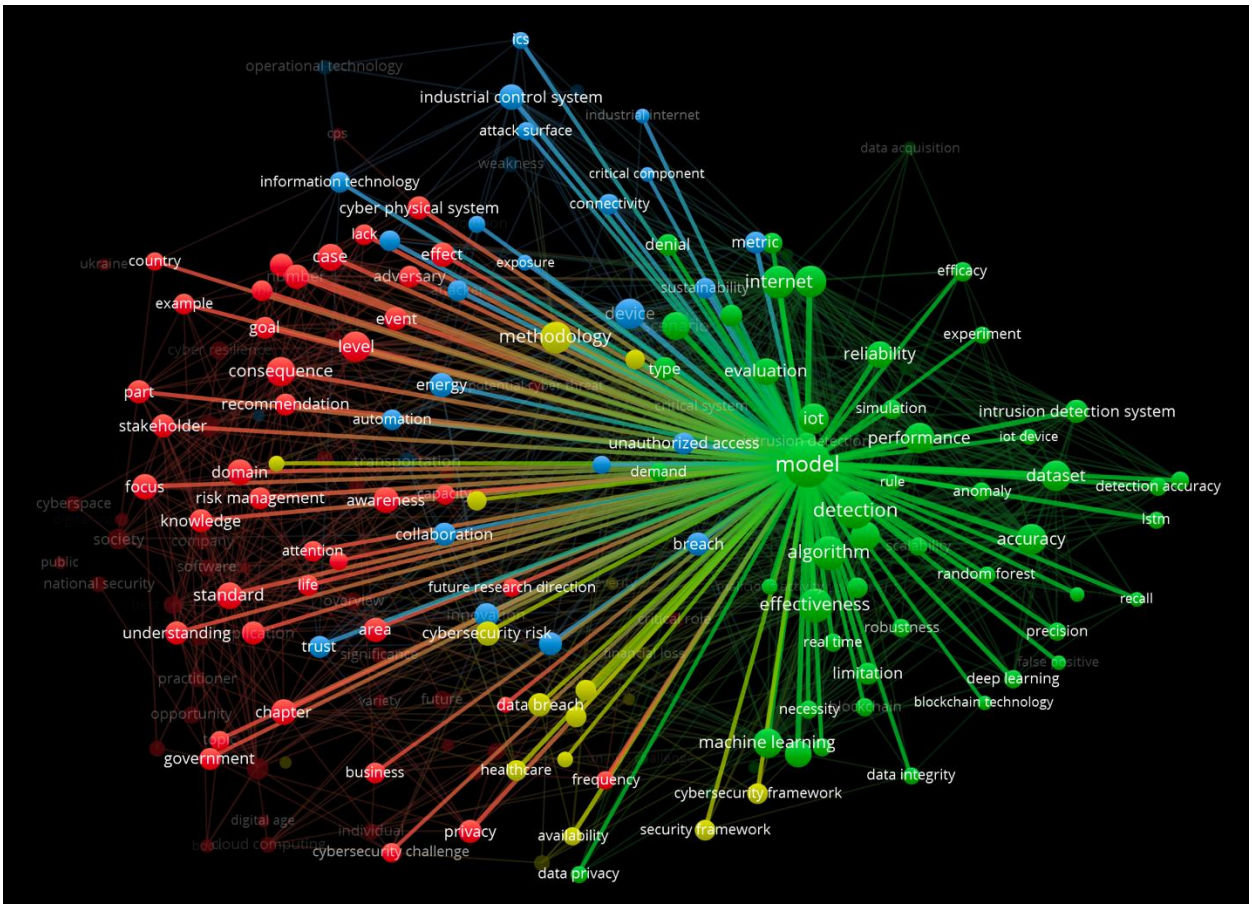


Nota: información recuperada de SCOPUS (2025)

Ahora bien, las investigaciones registradas entre 2023 y 2024 poseen como factor común la indagación y estudio de riesgos hacia infraestructura crítica cibernética. Del núcleo de autores encontrados (142), diez presentan resultados de investigación conexos a la construcción de acciones estratégicas o medidas de prevención. Todo lo anterior, ligado al sector defensa nacional.

El debate para los autores encontrados comienza con la exposición de nuevos riesgos para infraestructura crítica cibernética nacional, que tienen aproximación temática con el factor inteligencia artificial, tal y como se registra en el siguiente mapa de concurrencias:

Figura 4. Concurrencia de términos en relación a ciber seguridad, riesgos digitales y factores conexos.



Nota: elaboración propia con información recuperada de SCOPUS (2025)

Según el ejercicio de concurrencia terminológica desarrollado, los riesgos cibernéticos vigentes, es decir, a 2025, están intrínsecamente correlacionados con: machine learning, intrusión de sistemas, ataque a infraestructuras cibernéticas y metodologías de detección.

Con base en esta descripción, se da inicio al análisis temático, trayendo a colación la investigación de (Volk, 2024), quien explica que los riesgos y vulnerabilidades en la infraestructura crítica cibernética nacional (ICCN) del sector defensa, podrían estar directamente relacionados con el avance de las tecnologías de evasión y persistencia, potenciadas por inteligencia artificial (IA).

Esta es una de las amenazas más relevantes es el uso de algoritmos de aprendizaje automático para realizar ataques de reconocimiento avanzado (Akbarian, Ramezani, Hamidi, Y Haghighat, 2020).

Ese tipo de algoritmos analiza grandes volúmenes de datos sobre redes y sistemas objetivo, identificando puntos vulnerables con precisión. Este tipo de reconocimiento no solo acelera la etapa de preparación del ataque, sino que también permite personalizar vectores de intrusión para maximizar su efectividad.

Así es, por tanto, como los atacantes pueden automatizar el proceso de ciber ataque, lo que reduce significativamente el tiempo necesario para identificar oportunidades de explotación en sistemas críticos del sector defensa.

Otra amenaza significativa, desde la perspectiva de (Sarker, 2024) es la creación de malware polimórfico impulsado por IA, capaz de modificar su estructura constantemente para evadir herramientas de detección basadas en firmas.

Este tipo de malware utiliza algoritmos generativos que alteran su código sin afectar su funcionalidad maliciosa, dificultando que los sistemas tradicionales de seguridad lo identifiquen.

En la ICCN, un malware polimórfico puede infiltrarse en sistemas de control industrial o redes de comunicación militar, comprometiendo su integridad (Chauhan, Sabeel, Izaddoost, & Shah Heydari, 2021). Además, al operar de manera dinámica, este tipo de amenaza se adapta a cambios en las configuraciones de seguridad, prolongando su permanencia dentro del sistema comprometido y aumentando el riesgo de daño a largo plazo.

Otro riesgo prominente es la manipulación de datos mediante técnicas de IA. En este contexto, los atacantes diseñan y reproducen modelos de aprendizaje profundo para generar datos falsificados que imiten patrones legítimos, lo que puede engañar a los sistemas de análisis y control (Choudhary, Choudhary, y Salve, 2018).

En el contexto del sector defensa, esto incluye alteración de datos de sensores en sistemas de vigilancia o el envío de información falsa a plataformas de comando y control. Tales acciones no solo comprometen la toma de decisiones estratégica, puesto que también provocan respuestas estratégicas basadas en información inexacta, afectando la seguridad

nacional de manera directa y con consecuencias potencialmente irreversibles (Goffer, y otros, 2025).

De la misma forma, el análisis de autores trae a colación la construcción científica derivada de dos investigaciones relevantes, ambas centradas en el uso de redes generativas. De acuerdo con Agrawal, Kaur, y Myneni (2024) y Ankalaki, Rajesh, Pallavi, Hukkeri, y Naik (2025), las redes generativas adversariales (GANs) son una amenaza emergente, estructuradas en sintaxis, y complejidad abracada aspectos como: diversificación de variables y algoritmos de ataque.

Estas redes permiten a los atacantes crear contenido sintético altamente realista, como imágenes, audio o video falsificados, los cuales son utilizados en campañas de desinformación o ingeniería social avanzada. En el ámbito de la ICCN, los ataques incluyen la creación de órdenes falsas aparentemente emitidas por altos mandos militares o la simulación de fallos en sistemas críticos.

La capacidad de las GANs para generar contenido indistinguible del real dificulta su detección y amplifica su impacto, especialmente en escenarios donde la velocidad de respuesta depende de protocolos con problemas como: desactualización, desconocimiento técnico y evolución constante de los códigos de intervención e impacto.

Junto con los GAN´s, otro tipo de factores disruptivos como los ataques de denegación de servicio distribuido (DDoS) evolucionan a la par de nuevos sistemas de inteligencia artificial.

En este entorno, los atacantes utilizan algoritmos de aprendizaje para identificar patrones de tráfico que maximicen la efectividad de impactos cibernéticos, saturando redes críticas de manera más eficiente.

Con los algoritmos, los atacantes coordinan redes de bots para distribuir las cargas de ataque de manera óptima, evitando la emisión de códigos o variables compuestas que producen detección.

En la infraestructura crítica cibernética, un ataque DDoS dirigido podría inutilizar sistemas esenciales, como redes de comunicación militar o plataformas de gestión de crisis, buscando como fin la interrupción funcional de redes conectadas a la infraestructura, co-

creando formas de gestión e impacto allegadas a la inyección de código malicioso (Roopesh, Nishat, Rasetti, y Rahaman, 2024).

Otra vulnerabilidad por traer a colación en este análisis son los rootkits avanzados. Para este contexto, los atacantes diseñan rootkits potenciados por IA que se integran profundamente en los sistemas operativos, permitiendo el acceso continuo a los sistemas comprometidos.

Al respecto, Landauer (2025) explica que los rootkits se adaptan de forma dinámica a cambios en el entorno del sistema, como actualizaciones de software o modificaciones en las configuraciones de seguridad.

En el sector defensa, esta capacidad de persistencia permite a los atacantes mantener acceso a sistemas clasificados durante largos periodos, extrayendo información sensible o manipulando operaciones críticas sin ser detectados (Shaik y Shaik, 2024).

Hasta esta parte del análisis, las amenazas identificadas en la infraestructura crítica cibernética nacional (ICCN) del sector defensa están clasificadas en varias categorías técnicas.

En evasión, se incluyen el malware polimórfico, los rootkits avanzados y las redes generativas adversariales (GANs), todos diseñados para evitar la detección y prolongar su presencia en los sistemas. En persistencia, destacan los ataques de reconocimiento avanzado y la explotación de vulnerabilidades de día cero, que permiten acceso prolongado a sistemas críticos. En manipulación y disrupción, se encuentran la generación de datos falsificados, los ataques de denegación de servicio distribuido (DDoS) y las campañas de spear-phishing avanzadas.

Continuar con el análisis una vez identificadas las características principales amerita incluir otra categoría: injerencia cibernética y automatización de ciber ataques.

Bajo esta perspectiva, Wang, Guo, Li, y Zheng (2024) debaten que la automatización de ataques mediante IA también facilitan el desarrollo de campañas de spear-phishing más sofisticadas.

Los algoritmos de procesamiento de lenguaje natural utilizan grandes volúmenes de datos sobre objetivos específicos para generar correos electrónicos personalizados que imiten comunicaciones legítimas.

En el contexto de la ICCN, un ataque de spear-phishing exitoso podría comprometer credenciales de acceso a sistemas críticos, facilitando la intrusión inicial y el despliegue de otras amenazas más complejas (Birthriya, Ahlawat, y Jain, 2025). Este tipo de ataques se ha vuelto particularmente efectivo debido a la capacidad de la IA para replicar patrones de comunicación humana con alta precisión.

Otra amenaza relevante es la explotación de vulnerabilidades de día cero mediante algoritmos de IA. Estos algoritmos exploran grandes cantidades de código en busca de errores o fallos que aún no han sido descubiertos por los desarrolladores (Jimmy, 2024).

Una vez identificadas, estas vulnerabilidades pueden ser explotadas para comprometer sistemas críticos antes de que se implementen parches de seguridad (Jimmy, 2024). En el sector defensa, donde los sistemas están interconectados y personalizados, la explotación de una vulnerabilidad de este tipo podría tener un impacto desproporcionado, afectando la operatividad de múltiples plataformas de manera simultánea.

Ahora, de la explotación de vulnerabilidad deriva el uso de IA para eludir sistemas de autenticación multifactor.

De acuerdo con Malik (2024), la utilización de algoritmos de aprendizaje para analizar patrones biométricos o de comportamiento, produce códigos de intervención adecuados para engañar sistemas de defensa digital.

En la ICCN, esto afecta el acceso no autorizado a sistemas clasificados o instalaciones críticas, comprometiendo la seguridad física como la cibernética. La capacidad de la IA para replicar patrones complejos aumenta significativamente el riesgo de vulnerabilidad en los sistemas, produciendo formas de ciber ataque disruptivas, ajenas al conocimiento clásico contenidos en los protocolos de ciber seguridad.

Las amenazas cibernéticas hacia la ICCN del sector defensa, han evolucionado de forma paralela con el uso de IA, abarcando desde el reconocimiento avanzado y el malware polimórfico hasta la manipulación de datos y los ataques híbridos. Con un 29,8% de las investigaciones recientes centradas en ciencias de la computación y un 23% en ingenierías, es evidente que estas disciplinas son imperativas para diseñar estrategias de restricción de impactos o denegación de daños en contra de infraestructura cibernética estratégica.

La velocidad de evolución de las amenazas requiere una respuesta igualmente dinámica, que combine investigación avanzada con implementación de soluciones prácticas y colaborativas; de ahí, que el análisis de la probabilidad o identificación de riesgos con capacidad de impacto sea el siguiente paso en la investigación, y para tal fin, la base conceptual a utilizar debe darse, iniciarse y basarse en la matriz de riesgos que se relaciona a continuación:

Tabla 1. Identificación de riesgos asociados a inteligencia artificial e impactos sobre ICCN del sector defensa.

Riesgo Identificado	Descripción Breve	Impacto en la ICCN del Sector Defensa	Probabilidad	Severidad	Nivel de Riesgo
Reconocimiento Avanzado con IA	Uso de algoritmos de aprendizaje automático para analizar redes críticas y personalizar vectores de ataque.	Automatiza la preparación de ciberataques, acelerando la identificación de vulnerabilidades en sistemas de defensa.	Alta	Alta	Crítico
Malware Polimórfico	Malware impulsado por IA que modifica su estructura constantemente para evadir detección.	Infiltra sistemas de control industrial y redes militares, comprometiendo su integridad y prolongando su permanencia en los sistemas comprometidos.	Alta	Alta	Crítico
Manipulación de Datos mediante IA	Generación de datos falsificados que imitan patrones legítimos para engañar sistemas de análisis y control.	Compromete decisiones estratégicas al enviar información inexacta a plataformas de comando y vigilancia, afectando la seguridad nacional.	Media	Alta	Alto
Redes Generativas Adversariales (GANs)	Creación de contenido sintético realista, como órdenes falsas o simulación de fallos en sistemas críticos.	Facilita campañas de desinformación e ingeniería social avanzada, afectando la operatividad y confianza en sistemas de mando y control.	Media	Alta	Alto
Ataques de Denegación de Servicio Distribuido (DDoS)	Saturación de redes críticas mediante bots coordinados por IA para maximizar la disrupción.	Inutiliza redes de comunicación militar y plataformas de gestión de crisis, afectando la capacidad de respuesta en situaciones críticas.	Alta	Alta	Crítico
Rootkits Avanzados Potenciados por IA	Software malicioso que se integra profundamente en sistemas operativos y se adapta dinámicamente a cambios en el entorno.	Permite acceso continuo a sistemas clasificados, facilitando la extracción de información sensible y manipulación de operaciones críticas.	Alta	Alta	Crítico
Spear-Phishing Automatizado	Uso de IA para generar correos electrónicos personalizados que imiten comunicaciones legítimas.	Compromete credenciales de acceso a sistemas críticos, facilitando intrusiones iniciales y el despliegue de amenazas complejas.	Alta	Alta	Crítico
Explotación de Vulnerabilidades de Día Cero	Identificación automatizada de errores en código aún no parcheados para comprometer sistemas críticos.	Afecta múltiples plataformas interconectadas, impactando la operatividad simultánea de sistemas críticos del sector defensa.	Alta	Alta	Crítico
Elusión de Autenticación Multifactor	Análisis de patrones biométricos y de comportamiento mediante IA para engañar sistemas de autenticación.	Permite acceso no autorizado a sistemas clasificados e instalaciones críticas, comprometiendo tanto la seguridad física como cibernética.	Media	Alta	Alto

Nota: elaboración propia. La matriz de riesgo fue desarrollada utilizando la metodología técnica establecida en el marco del NIST SP 800-30, que proporciona un enfoque estructurado para la identificación, análisis y priorización de riesgos en sistemas críticos. Los riesgos se identificaron a partir del análisis de fuentes científicas y técnicas, evaluando su probabilidad y severidad con base en su capacidad de impacto en la Infraestructura Crítica Cibernética Nacional (ICCN) del sector defensa. Cada riesgo fue categorizado según su naturaleza y potencial disruptivo, integrando criterios cualitativos y cuantitativos para determinar el nivel de riesgo asociado.

Análisis geopolítico y de cooperación internacional sobre riesgos de IA en la infraestructura crítica cibernética nacional del sector defensa.

La integración acelerada de inteligencia artificial (IA) en los ecosistemas de defensa ha desplazado la discusión desde el plano puramente técnico hacia dimensiones geopolíticas, normativas y de gobernanza internacional. Los hallazgos de su investigación muestran una curva ascendente en complejidad y volumen de ataques asistidos por IA, consistente con la ampliación de superficie de exposición, la automatización ofensiva y la degradación de la confianza en datos y sistemas de mando y control.

Este escenario reconfigura el balance estratégico: convierte la infraestructura crítica cibernética nacional (ICCN) en un objetivo de alto valor dentro de una competencia tecnológica donde los Estados, actores no estatales y coaliciones transnacionales explotan asimetrías de conocimiento y cadena de suministro digital.

En el plano geopolítico, la IA aplicada a ciber operaciones se inserta en una dinámica de rivalidad entre potencias tecnológicas que compiten por supremacía en semiconductores avanzados, modelos fundacionales y capacidades de automatización ofensiva. La literatura muestra que la IA acelera el ciclo de reconocimiento, explotación y persistencia, optimizando la selección de vectores y eludir controles defensivos (Huang *et al*., 2023, Communications of the ACM). Ese ajuste de velocidad y escala altera la disuasión: reduce costos marginales de ataque y eleva la incertidumbre del defensor.

En paralelo, la creciente dependencia de software y hardware globalizados introduce vectores de riesgo por terceros países, desde puertas traseras y manipulación de modelos hasta sabotaje de datos de entrenamiento (Brundage *et al*., 2020, Journal of Cyber Policy). Bajo estas condiciones, la protección de la ICCN exige políticas industriales y de seguridad de cadena de suministro coordinadas con aliados, así como marcos de evaluación de confianza técnica para modelos y componentes.

Los resultados que señalan el incremento de ransomware, explotación de vulnerabilidades y DDoS coordinados por IA son coherentes con investigaciones que documentan el uso de aprendizaje automático para ajuste dinámico de cargas, generación de variantes polimórficas y evasión de firmas (Sarker, 2021; Apruzzese y Colajanni, 2018).

En particular, los ataques adversariales contra modelos de defensa y la manipulación de datos operativos comprometen la integridad informacional, un pilar de la conducción estratégica.

Athalye *et al*. (2018) y Carlini y Wagner (2017) evidencian que perturbaciones mínimas pueden inducir decisiones erróneas en sistemas de clasificación y detección, desplazando la batalla desde el perímetro de red hacia la capa estadística de los modelos. En contextos de defensa, esa vulnerabilidad afecta sensores, fusión de datos y cuadros de situación comunes, abriendo margen para la desinformación técnica y la degradación operativa.

La convergencia entre IA generativa y operaciones de influencia amplifica riesgos sistémicos. Además de deepfakes y órdenes simuladas, el uso de modelos de lenguaje para spear-phishing personalizado incrementa tasas de éxito y reduce los indicios de detección humana (Kirsch *et al*., 2023, arXiv; Pearce *et al*., 2023). La combinación con explotación de día cero acelerada por técnicas de minería de repositorios y fuzzing guiado por aprendizaje genera campañas híbridas de alta cadencia (Takanen *et al*., 2020). La consecuencia estratégica es una presión constante sobre los tiempos de parcheo y una mayor dependencia de capacidades de anticipación, caza proactiva y segmentación adaptativa.

Desde la perspectiva de gobernanza y cooperación internacional, existe un incipiente consenso sobre principios de IA segura y responsable que puede trasladarse al dominio de ciberdefensa. El G7 y la OCDE han delineado estándares de gestión de riesgo y transparencia de modelos; la Iniciativa de Seguridad de Modelos de la Cumbre de Bletchley (2023) y guías como NIST AI RMF 1.0 (2023) ofrecen marcos para evaluación de peligros, pruebas de alineamiento y monitoreo pos-despliegue.

Sin embargo, la literatura destaca brechas entre principios y verificación técnica en contextos militares, particularmente en pruebas de robustez adversarial, supply chain security y gobernanza de datos de entrenamiento (Cummings, 2021; Goldstein, 2023). Cerrarlas requiere ejercicios combinados de red-teaming de IA, compartición de inteligencia técnica sobre fallas de modelos y acuerdos de notificación coordinada de vulnerabilidades con componentes clasificados.

En esa línea, los marcos NIST SP 800-30 y 800-53 siguen siendo referencia para cuantificar probabilidad y severidad, pero deben complementarse con pruebas de caja roja para modelos (Uesato *et al*., 2018) y metodologías de evaluación sociotécnica centradas en degradación de misión. Estudios recientes sugieren integrar MLOps de seguridad (SecMLops) que asegure trazabilidad de *datasets*, validación de integridad de pesos y control de deriva de concepto bajo adversario (Breck *et al*., 2017; Habib *et al*., 2023).

Ello se alinea con la necesidad, indicada en sus resultados, de actualizar capacidades con prevención y anticipación; operacionalmente, implica telemetría de alta granularidad, detección basada en comportamiento y verificación criptográfica de software y modelos en el borde táctico.

La cooperación internacional es un multiplicador de resiliencia cuando se orienta a interoperabilidad técnica y respuesta coordinada.

Programas de intercambio de indicadores de compromiso enriquecidos con artefactos de IA (por ejemplo, pesos maliciosos, *prompts* de explotación, firmas de adversarial patches), ejercicios conjuntos de ciberdefensa con escenarios de degradación de modelos y mecanismos de pre-compromiso para asistencia rápida en incidentes de ICCN fortalecen la postura colectiva (Rid y McBurney, 2012; NATO CCDCOE, 2022). Asimismo, las alianzas para la seguridad de semiconductores y la verificación de origen de aceleradores críticos reducen riesgos de insertos maliciosos y dependencia estratégica (Miller y Moss, 2023).

En paralelo, la dimensión legal y de normas de conducta en el ciberespacio exige clarificar umbrales de uso de la fuerza y atribución cuando la IA introduce ambigüedad operacional (Schmitt, 2017). El uso de IA para eludir autenticación biométrica o manipular sistemas SCADA militares podría, según contexto y efectos, cruzar líneas de soberanía y activar cláusulas de defensa colectiva. La investigación sugiere que robustecer la atribución técnica mediante cooperación de múltiples partes confiables y preservación de cadena de custodia de artefactos de IA es clave para sostener costos reputacionales y disuasión (Reeves *et al*., 2021).

Por lo anterior, la estrategia nacional debería integrar tres vectores: 1) resiliencia tecnológica con seguridad por diseño en modelos y datos, segmentación de privilegios algorítmicos y pruebas adversariales continuas; 2) gobernanza multinivel que enlace doctrinas de ciberdefensa con marcos internacionales de IA segura y acuerdos de intercambio de inteligencia técnica; y 3) poder de coalición para asegurar cadenas de suministro, verificación de modelos y respuesta coordinada a incidentes en ICCN.

La evidencia científica converge en que la IA es simultáneamente superficie y herramienta de defensa; su gestión estratégica decide si la aceleración algorítmica inclina la balanza hacia la disrupción o hacia la superioridad resiliente (Brundage *et al*., 2020; NIST, 2023; Apruzzese y Colajanni, 2018).

Referencia

Abdullahi, M., Baashar, Y., Alhussian, H., Alwadain, A., Aziz, N., Capretz, L. F., & Abdulkadir, S. J. (2022). Detecting cybersecurity attacks in internet of things using artificial intelligence methods: A systematic literature review. *Electronics*, *11*(2), 198.

Agrawal, G., Kaur, A., y Myneni, S. (2024). A review of generative models in generating synthetic attack data for cybersecurity. *Electronics*, *13*(2), 322 - 325.

Akbarian, F., Ramezani, A., Hamidi, M., y Haghighat, V. (2020). Advanced algorithm to detect stealthy cyber attacks on automatic generation control in smart grid. *ET Cyber-Physical Systems: Theory & Applications*, *54*, 351-358.

Alhayani, B., Mohammed, H. J., Chaloob, I. Z., y Ahmed, J. S. (2021). Effectiveness of artificial Intelligence techniques against cyber security risks apply of IT industry. *Materials Today: Proceedings*, *531*(10.1016).

Ankalaki, S., Rajesh, A., Pallavi, M., Hukkeri, G., & Naik, G. R. (2025). Cyber attack prediction: From traditional machine learning to generative artificial intelligence. *IEEE Access.*, 1-14.

Ansari, M. F., Dash, B., Sharma, P., y Yathiraju, N. (2022). The impact and limitations of artificial intelligence in cybersecurity: a literature review. *International Journal of Advanced Research in Computer and Communication Engineering*.

Birthriya, S. K., Ahlawat, P., & Jain, A. K. (2025). Detection and Prevention of Spear Phishing Attacks: A Comprehensive Survey. . *Computers & Security*, 1-14.

Chakraborty, A., Biswas, A., & Khan, A. K. (2023). Artificial intelligence for cybersecurity: Threats, attacks and mitigation. In *Artificial Intelligence for Societal Issues* (pp. 3-25). Cham: Springer International Publishing.

Chauhan, R., Sabeel, U., Izaddoost, A., & Shah Heydari, S. (2021). Polymorphic adversarial cyberattacks using WGAN. *Journal of Cybersecurity and Privacy*, *1*(4), 767-792.

Chehri, A., Fofana, I., & Yang, X. (2021). Security risk modeling in smart grid critical infrastructures in the era of big data and artificial intelligence. *Sustainability*, *13*(6), 3196.

Choudhary, A., Choudhary, P., & Salve, S. (2018). A study on various cyber attacks and a proposed intelligent system for monitoring such attacks. *2018 3rd International Conference on Inventive Computation Technologies (ICICT)*, *10*(27), 612-617.

Comando Cibernético de las Fuerzas Militares. (2024). Boletín informativo nº 24. *Publicación mensual de CCOCI*.

Consejo de Política Económica y Social . (14 de febrero de 2025). *POLÍTICA NACIONAL DE INTELIGENCIA ARTIFICIAL* . Obtenido de https://www.dnp.gov.co/LaEntidad_/subdireccion-general-prospectiva-desarrollo-nacional/direccion-desarrollo-digital/Paginas/documentos-conpes-confianza-y-seguridad-digital.aspx

Cosano, A. R. (2024). *EL LIBRO BLANCO DE LA INTELIGENCIA ARTIFICIAL GENERATIVA* . ESPAÑA : DigitalES.

Goffer, M., Uddin, M., Has an, S., Barikdar, C., Hassan, J., Das, N., & Hasan, R. (2025). AI-Enhanced Cyber Threat Detection and Response Advancing National Security in Critical Infrastructure. *Journal of Posthumanism, 5*(3), 1667-1689.

Guembe, B., Azeta, A., Misra, S., Osamor, V. C., Fernández-Sanz, L., y Pospelova, V. (2022). The emerging threat of AI-deriven cyber-attacks: A review. Applied Artificial Intelligence, *36*(1), 2037254.

Hernández, R., Fernández, C., & Baptista, P. (2014). *Metodología de la investigación* (Vol. 6, pp. 102-256). México: mcGraw-Hill.

Jimmy, F. (2024). Cyber security vulnerabilities and remediation through cloud security tools. *Journal of Artificial Intelligence General science (JAIGS), 2*(1), 129-171.

Kure, H. I., Islam, S., y Mouratidis, H. (2022). An integrated cyber security risk management framework and risk predication for the critical infrastructure protection. *Neural Computing and Applications*, *34*(18), 15241-15271.

Landauer, M., Alton, L., Lindorfer, M., Skopik, F., Wurzenberger, M., & Hotwagner, W. (2025). Trace of the Times: Rootkit Detection through Temporal Anomalies in Kernel Activity. *Landauer, M., Alton, L., Lindorfer, M., Skopik, F., Wurzenberger, M., & Hotwagner, W. (2025). Trace of the Times: Rootkit Detection through Temporal Anomalies in Kernel Activity.* , 1-12.

Maddireddy, B. R., & Maddireddy, B. R. (2020). Proactive Cyber Defense: Utilizing AI for Early Threat Detection and Risk Assessment. *International Journal of Advanced Engineering Technologies and Innovations*, *1*(2), 64-83.

Malik, S. (2024). The Future of AI in Biometric Security: Enhancing Authentication and Privacy Protection.

Margalef, L., & Arenas, A. (2006). ¿Qué entendemos por innovación Educativa? A proposito del desarrollo curricular. *Perpectiva Educacional, 1*(47), 13-31.

Mosteanu, N. R. (2020). Artificial intelligence and cyber security "face to face with cyber-attack maltese case of risk management approach. *Ecoforum Journal*, *9*(2).

Pătrașcu, P. (2019). CYBERNETIC ACTIONS ON CRITICAL INFRASTRUCTURES IN THE MILITARY FIEL. . *Bulletin of" Carol I" National Defence University* , 40-45.

Roopesh, M., Nishat, N., Rasetti, S., & Rahaman, M. A. (2024). A Review Of Machine Learning And Feature Selection Techniques For Cybersecurity Attack Detection With A Focus On DDoS Attacks. *Academic Journal on Science, Technology, Engineering & Mathematic Education*, *4*(03), 178-194.

Roselli, N. (2011). Teoria del aprendizaje colaborativo y la teoria de la repesentación social: convergencias y posibles articulaciones. *Revista colombiana de Ciencias Sociales, 2*(2), 173-191.

Sarker, I. H. (2024). AI for critical infrastructure protection and resilience. . *AI-driven cybersecurity and threat intelligence: Cyber automation, intelligent decision-making and explainability*, 153-172.

Shaik, A., & Shaik, A. (2024). Integrated AI Cyber Secured Approach for Detection of Rootkits in Malware. In International Conference on Frontiers of Intelligent Computing: Theory and Applications. *Singapore: Springer Nature Singapore.*, *12*, 567-576.

Thomas, T. (2016). Russia's information warfare strategy: can the nation cope in future conflicts?. . *The Transformation of Russia's Armed Forces* , 148-177.

Truong, T. C., Diep, Q. B., & Zelinka, I. (2020). Artificial intelligence in the cyber domain: Offense and defense. *Symmetry*, *12*(3), 410.

Vakulyk, O., Petrenko, P., Kuzmenko, I., Pochtovyi, M., & Orlovskyi, R. (2020). CYBERSECURITY AS A COMPONENT OF THE NATIONAL SECURITY OF THE STATE. *Journal of Security & Sustainability Issues*, *9*(3), 1-10.

Volk, M. (2024). A safer future: Leveraging the AI power to improve the cybersecurity in critical infrastructures. *Electrotechnical Review/Elektrotehniski Vestnik*, *91*(3), 1-10.

Wang, J., Guo, J., Li, K., & Zheng, H. (2024). Distributed adaptive event-triggered control of connected automated vehicle platoon systems with spoofing cyber attacks. *IEEE Transactions on Vehicular Technology*., 1-10.

Yamin, M. M., Ullah, M., Ullah, H., & Katt, B. (2021). Weaponized AI for cyber-attacks. *Journal of Information Security and Applications*, *57*, 102722.

Zhang, Z., Al Hamadi, H., Damiani, E., Yeun, C. Y., & Taher, F. (2022). Explainable artificial Intelligence applications in cyber security: State-of-the-art in Research. *IEEE Access*, *10*, 93104-93139.