RAZÓN CRÍTICA
Núm. 18 (2025)
https://doi.org/10.21789/25007807.2111
ARTÍCULO DE INVESTIGACIÓN

POTENCIAL Y APLICACIONES DE LA INTELIGENCIA ARTIFICIAL EN SEGURIDAD

ARTIFICIAL INTELLIGENCE POTENTIAL AND APPLICATIONS IN SECURITY

POTENCIAL E APLICAÇÕES DA INTELIGÊNCIA ARTIFICIAL EM SEGURANÇA

Freddy Linares-Torres
Maestría en Ingeniería de Sistemas, Universidad del Pacífico, Perú.
linares_f@up.edu.pe | https://orcid.org/0000-0003-3644-0545

Brandon Melchor Salazar-Curichimba
Analista Neurometrics, Perú.
brandon.salazar@neurometrics.la | https://orcid.org/0000-0003-3018-3752

Fecha de recepción: 16 de julio de 2024
Fecha de aceptación: 3 de marzo de 2025
Disponible en línea: 29 de mayo de 2025

Sugerencia de citación: Linares-Torres, F., Salazar-Curichimba, B. M. (2025). Potencial y aplicaciones de la inteligencia artificial en seguridad. *Razón Crítica, 18*, 1-16. https://doi.org/10.21789/25007807.2111

Resumen

El presente artículo explora el impacto de la inteligencia artificial (IA) en las prácticas de seguridad a través de una revisión de sus aplicaciones en diversas áreas, incluyendo la automatización de procesos, protección de datos, ciberseguridad y defensa nacional. El propósito del artículo es proporcionar una visión integral de cómo la IA está aplicándose en estos campos, destacando tanto sus beneficios como los retos asociados. La automatización de procesos mediante la IA permite una mayor eficiencia y precisión, reduciendo errores humanos. En cuanto a la seguridad de los datos, las tecnologías de IA resguardan y responden a potenciales amenazas de manera más proactiva. En ciberseguridad, la IA proporciona capacidades predictivas y analíticas para la detección temprana de ataques y su neutralización. Por último, en el área de defensa nacional, la IA potencia diversas capacidades militares. A pesar de sus numerosos beneficios, el artículo también aborda los desafíos éticos y operativos de la implementación de la IA en seguridad, subrayando la importancia de una integración ética.

Palabras clave: Inteligencia artificial; seguridad; defensa; digitalización; ciencias sociales.

Abstract

This article examines the transformative impact of artificial intelligence (AI) on contemporary security strategies and practices. Through a comprehensive review of AI applications in process automation, data protection, cybersecurity, and national defense, the study aims to shed light on how AI is reshaping these critical domains—highlighting both its advantages and the challenges it presents. In the realm of process automation, AI enhances operational efficiency and precision, minimizes human error, and increases overall productivity. When it comes to data protection, advanced AI technologies play a pivotal role in safeguarding information by proactively detecting and responding to potential threats. In the field of cybersecurity, AI contributes powerful predictive and analytical tools that enable early detection and neutralization of attacks. Meanwhile, in national defense, AI strengthens surveillance systems and improves strategic military capabilities. While the benefits of AI integration are substantial, the article also addresses the ethical and operational concerns surrounding its implementation. It emphasizes the need for a balanced, responsible, and ethical approach to the adoption of AI in security frameworks.

Keywords: Artificial intelligence; security; defense; digitalization; social sciences.

Resumo

Este artigo explora o impacto transformador da inteligência artificial (IA) nas estratégias e práticas de segurança por meio de uma análise dos aplicativos de IA em várias áreas, incluindo automação de processos, proteção de dados, segurança cibernética e defesa nacional. O objetivo é oferecer uma visão geral abrangente de como a IA está revolucionando esses campos, destacando seus benefícios e desafios associados. A automação de processos por meio da IA proporciona maior eficiência e precisão, reduzindo o erro humano e melhorando a produtividade. Na proteção de dados, as tecnologias avançadas de IA protegem os dados e respondem a possíveis ameaças de forma proativa, protegendo informações confidenciais. Na segurança cibernética, a IA oferece recursos preditivos e análises avançadas para a detecção precoce e a neutralização de ataques. Na defesa nacional, a IA otimiza o monitoramento e melhora os recursos militares. Apesar de seus muitos benefícios, o artigo também aborda os desafios éticos e operacionais de sua implementação na segurança, destacando a importância de uma adoção equilibrada e ética.

Palavras-chave: inteligência artificial; segurança; defesa; digitalização; ciências sociais.

Introducción

El panorama de la seguridad y la defensa se encuentra en constante transformación impulsado por la aparición de nuevas amenazas, el permanente desarrollo de tecnologías y la creciente relevancia de los espacios digitales. Muchas de estas condiciones derivan del complejo proceso de digitalización global, que ha traído cambios importantes en el modo en que se ejecutan distintos procesos productivos y sociales. El Internet es el principal motor de estos cambios debido a su expansión como el medio de comunicación más relevante en la actualidad, en el que no solamente se comparte constantemente información de diversos tipos, sino porque cada vez es más accesible para la población. Para 2024, más de 5.44 mil millones de personas a nivel mundial usan Internet, es decir, aproximadamente el 67.1 % de la población mundial, cifra que en 2010 equivalía a 27.5 % (Kemp, 2024; Linares-Torres, 2024). Este mayor acceso al Internet permite a más personas acceder al contenido, servicios y espacios digitales ofrecidos en la red para diversos tipos de actividades (como educación, trabajo, salud, entretenimiento, búsqueda de información, entre otros). La complejidad y variedad de la

actividad digital de la población, debido a diversos factores, se ha transformado de forma significativa, superando tanto limitaciones vinculadas al *hardware* como al *software*. Desde un teléfono inteligente moderno se pueden hacer muchas más actividades en Internet que las posibles con una computadora de escritorio hace décadas. Ello ha implicado un aumento en el volumen de la actividad en Internet, especialmente en los años más recientes, dado que el volumen de datos producidos y gestionados en Internet experimenta un crecimiento exponencial cada año.

En 2015, la cantidad de datos generados, capturados, copiados y consumidos en Internet a nivel mundial fue de 15.5 *zettabytes* (medida que equivale a mil millones de *terabytes*) y se proyecta que el volumen de datos generados y procesados alcance niveles que excederán los 180 *zettabytes* en 2025 (Taylor, 2023). Este drástico aumento en el volumen de datos producidos y procesados refleja el rol crítico que cumple el Internet como medio para la realización de millones de interacciones cada vez más complejas. De esta forma, Internet ha cobrado un papel preponderante en el funcionamiento de la sociedad contemporánea, no solo como un medio complementario a los espacios físicos, sino como un componente esencial para la participación activa de los individuos en la vida social moderna (Laskar, 2023). Reconocer el valor y la complejidad detrás de la realización y dominio de estas actividades *online* es relevante para poder diagnosticar tanto fortalezas como oportunidades de mejora respecto al desarrollo de las capacidades digitales de la población, las cuales son cruciales para facilitar el avance de la sociedad. Evaluaciones de este tipo no pueden limitarse a cuantificaciones simples, como enlistar las actividades que se hacen o no en Internet, sino que requieren instrumentos más detallados como el Índice de Actividad Digital, el cual evalúa la intensidad e incidencia de esta actividad en Internet de las personas para evaluar el avance hacia una verdadera ciudadanía digital (Linares-Torres et al., 2023).

Como resultado de esta expansión del Internet, la seguridad de los espacios y elementos digitales ha cobrado especial importancia para agentes del sector privado y público, especialmente considerando la proliferación de nuevas y sofisticadas amenazas digitales. Los activos digitales de actores que integran, o integrarán, de forma significativa la dimensión digital en sus procesos ha aumentado su valor notablemente, lo que implica diversos desafíos en términos de seguridad (Vargas, 2023). De no reconocer adecuadamente la relevancia de la protección de sus activos digitales, se estará manteniendo vulnerabilidades que pueden ser aprovechadas por ciberdelincuentes, comprometiendo tanto la integridad de los procesos internos como generando daños a terceros. La tabla 1 resume diversos casos de ciberataques que demuestran el alcance que pueden tener diversas amenazas actuales en el ciberespacio.

Tabla 1. Casos de ciberataques a nivel internacional

Ataque	Fechas	Descripción
Stuxnet infectó centrifugadoras en central nuclear de Irán	2010	El programa malicioso Stuxnet infectó y controló por varios meses centrifugadoras mientras camuflaba el malfuncionamiento de las máquinas con datos falsos (Zetter, 2014).
"Octubre rojo": robo de información en instituciones de Europa, Asia central y Norteamérica	Octubre, 2012	Se reveló que el virus Rocra explotaba vulnerabilidades en Windows Office para infiltrarse en correos electrónicos, permitiéndole sustraer documentos sensibles y encriptados de distintas entidades públicas a nivel internacional durante más de cinco años (Wünsch y Machhaus, 2013).
Hackeo a la Oficina de Gestión de Personal de EE. UU.	Junio, 2015	Robo de información personal de cerca de 4 millones de empleados federales, incluyendo información bancaria del personal hasta información sobre capacitaciones recibidas (El Mundo, 2015).
Ciberataque masivo del virus WannaCry	Mayo, 2017	El ciberataque a gran escala mediante el software malicioso Ransom:Win32.WannaCrypt, conocido como WannaCry, afectó los sistemas informáticos de instituciones gubernamentales y empresas a nivel global, infectando más de 200 mil computadoras en 150 países y bloqueando su acceso (BBC, 2017).
Hackeo de China en redes de defensa de Japón	2020	Hackers chinos tuvieron acceso profundo y persistente a planes, capacidades y evaluaciones de las deficiencias militares del sector militar de Japón, el mayor aliado estratégico de EE. UU. en el este de Asia, con el fin de conseguir información de ese país (Paiva, 2023).
Ciberataque al Gobierno de Costa Rica	Abril, 2022	Los ataques de ransomware perpetrados por el grupo ruso Conti comprometieron múltiples instituciones gubernamentales, generando la interrupción de servicios esenciales y afectando significativamente la administración pública. Como consecuencia, el país afectado se convirtió en el primero en declarar un estado de emergencia debido a un ciberataque (Tornaghi, 2023).

Nota. Tomado de Linares-Torres, 2024, pp. 143-144.

Tanto para el sector privado como público, se ha vuelto necesario fortalecer la ciberseguridad; para lo cual se requiere una completa evaluación de las capacidades tecnológicas para proteger las redes, los datos y las infraestructuras digitales que se poseen. En ese sentido, el avance de la IA, como ChatGPT , ha permitido por ejemplo, masificar el acceso a código fuente potencialmente empleado en fines no lícitos.

En esta investigación se explora el panorama actual de la IA en seguridad y defensa desde una perspectiva global, basándose para ello en una revisión de la literatura de diversas aplicaciones en el sector durante la última década considerando tanto la literatura académica como casos de empresas tecnológicas y gobiernos.

La IA y la seguridad

En una época en la cual las computadoras son cada vez más veloces y poderosas (mejores tarjetas de video y procesadores), es mucho más fácil acceder a diversos recursos digitales (material de estudio, foros de discusión, cursos *online*) y herramientas útiles para la programación (como Anaconda, una distribución de los lenguajes de programación Python y R para computación científica), por lo que el desarrollo de *software* se ve significativamente impulsado.

Gracias a estos avances, dos ramas de la IA han experimentado un gran desarrollo: el aprendizaje automático o *machine learning*, y el aprendizaje profundo o *deep learning*. El aprendizaje automático permite a los sistemas mejorar su eficiencia y rendimiento a través del aprendizaje de la experiencia, así con solo examinar los datos históricos, estos modelos pueden identificar patrones y realizar predicciones (Chio y Freeman, 2018). Este campo abarca varios métodos como los algoritmos de regresión, árboles de decisión, vectores de apoyo y redes neuronales, los cuales se han empleado en la ciencia de datos para aproximación e inferencia (Goodfellow et al., 2016).

El aprendizaje profundo, una subdisciplina del aprendizaje automático, emplea redes neuronales artificiales con varios estratos para discernir y modelar patrones intrincados en los datos. Esta tecnología es esencial para la progresión de sistemas sofisticados de interpretación de imágenes, denominada visión por computadora (*computer vision*), y en el procesamiento del lenguaje natural (*natural language processing*), ramas de la inteligencia artificial centrados en la comprensión y el análisis de los datos visuales y escritos, respectivamente (LeCun et al., 2015).

Tras el anuncio y popularización de servicios como ChatGPT, la IA generativa ha demostrado su versatilidad para desenvolverse en la creación de distintos tipos de contenidos, popularizando la idea de que la IA se enfocará en reemplazar y desplazar a las personas dedicadas a áreas como el arte, cuando esta tecnología puede actuar como una importante herramienta de apoyo para muchos roles en distintos sectores (Chui et al., 2023), incluyendo las actividades relacionadas con la seguridad. La integración de la IA en la seguridad y la defensa representa un riesgo y una tarea pendiente, ofreciendo nuevas capacidades para proteger a organizaciones, empresas y usuarios individuales.

Este proceso de digitalización en sectores públicos y privados varía considerablemente entre países, pero la tendencia global hacia la adopción de tecnologías digitales es clara. No solo se trata de una forma para innovar los procesos, sino de reconocer un cambio estructural en la sociedad, en la cual la adopción de la tecnología se va volviendo una necesidad alineada con una transformación de las instituciones más tradicionales, como por ejemplo el Estado, el cual para optimizar su rol como ofertante de servicios públicos requerirá transicionar de una entidad física basada en oficinas e instituciones a una más parecida a una plataforma digital unificada que integre adecuadamente los servicios públicos clave (Linares y Contreras, 2023). De esa manera, es natural que en las infraestructuras establecidas de los países altamente digitalizados la implementación de IA en ciberseguridad ya sea una realidad (Brynjolfsson y McAfee, 2014). No obstante, en regiones donde la digitalización aún está en desarrollo, es donde la IA puede jugar un papel más preponderante ya que podría emplearse en la construcción de los primeros sistemas de seguridad robustos y así evitar las vulnerabilidades críticas que atenten contra la gestión pública (Cuba, 2021).

En contraste con la seguridad digital tradicional, la inteligencia artificial posee la capacidad de aprender y adaptarse continuamente a nuevas amenazas, lo que representa una oportunidad para su aplicación en los desafíos de seguridad actuales.

Aplicaciones de la IA en seguridad

Considerando la creciente digitalización de la sociedad y la integración de la IA en diversas tareas, no es sorprendente los distintos campos de aplicación que puede tomar para la seguridad. A continuación detallamos algunos de estos.

Automatización de procesos

Uno de las principales ventajas de la IA en la seguridad es que puede mejorar la capacidad de detectar, prevenir y responder a potenciales amenazas de manera más eficiente y efectiva que las técnicas tradicionales, debido a su capacidad para analizar grandes volúmenes de datos en tiempo real que le permiten detectar rápidamente patrones anómalos y posibles amenazas para alertar a los equipos de seguridad antes de que un ataque pueda causar daños significativos (Chio y Freeman, 2018).

En ese sentido, es clara la utilidad de la IA para opmizar la gestión de las tareas rutinarias de seguridad al automatizar procesos como el análisis de malware, la gestión de parches o la monitorización de la red, permitiendo con ello que los expertos humanos se centren en retos más complejos y a la vez reduciendo la probabilidad de errores humanos (Palo Alto, s. f.) Por ejemplo, la IA puede actualizar el software de toda una organización al instante, una tarea que a los humanos les llevaría mucho más tiempo. Esta eficiencia refuerza las ciberdefensas de una organización y mejora la productividad operativa, convirtiendo a la IA en un aliado clave en la lucha continua contra las ciberamenazas. Un caso de aplicación sobre la automatización de la seguridad es el grupo de compañías energéticas y petroquímicas Shell (Gerges, 2021), el cual utiliza IA, dispositivos industriales de Internet de las cosas (IoT) y equipos robóticos para monitorear sus operaciones en tiempo real con el fin de detectar fugas de gas y otros incidentes de seguridad antes de que se agraven.

Otro ejemplo de cómo se está utilizando la IA para automatizar acciones de seguridad lo encontramos en las fiscalizaciones del Instituto Nacional de Defensa de la Competencia y de la Protección de la Propiedad Intelectual (Indecopi, 2024) en Perú. En este sentido, Indecopi ha anunciado la implementación de herramientas de inteligencia artificial y análisis de datos para asegurarse de que se cumpla la prohibición de las llamadas no consentidas o el spam telefónico con fines publicitarios. De esa forma, los procesos de fiscalización se volverán más eficientes y se mejorará la identificación de infracciones.

No obstante, a pesar de las ventajas mencionadas, la aplicación de la IA en la automatización de procesos también tiene limitaciones. Una de ellas es la dependencia de grandes cantidades de datos de alta calidad para tener mayor precisión, lo cual de por sí es un reto aún mayor para el sector seguridad ya que en muchas situaciones estos conjuntos de datos etiquetados no se encuentran disponibles (Brundage et al., 2020). Por otro lado, la vulnerabilidad a ciberataques que pueden engañar al modelo basado en IA para dañar su efectividad también es otra limitante a superar (Piergallini, 2020). Un caso

Protección de datos

Este concepto se refiere a proteger el átomo de la información digital transmitida en correos electrónicos, contraseñas o documentos. En los últimos años, este concepto ha cobrado relevancia considerando la creciente oleada de ciberataques a fin de hacerse con el activo valioso que representan los datos (Vargas, 2023). En ese sentido, la IA a través de encriptación avanzada y algoritmos de aprendizaje automático pueden identificar patrones de comportamiento inusuales para prevenir estos ataques. Una de las tecnologías que puede ser impulsada y que demuestra un gran potencial en esta manera de aplicación es la biometría; gracias a los sensores especializados que permiten una identificación precisa (Linares-Torres et al., 2024).

Un caso de aplicación de soluciones de IA para la protección de datos es el de United Family Healthcare (UFH), una red internacional de hospitales y clínicas con sede en Pekín, China, que consideraba que los ordenadores personales, las plataformas sociales y los teléfonos inteligentes representaban posibles riesgos para su infraestructura de seguridad, al igual que el uso compartido de contraseñas e información entre sus empleados. Ante ello, la UFH optó por elegir la solución de seguridad de IBM Security QRadar SIEM, la cual permitió que los miembros del equipo de la UFH con una formación de seguridad formal limitada puedan ver las amenazas priorizadas y participar en investigaciones relacionadas. Un principal aspecto que se debe destacar es que con la integración de esta solución basada en IA en sus sistemas descubrieron nombres de usuarios y contraseñas no conformes entre los empleados (IBM, 2023). De igual forma, la gestión unificada de los datos permitió a la UFH atender con más soltura las exigencias regulatorias del entorno, al añadir utilidades de creación automática de reportes que simplifican la preparación de revisiones, ya sean internas o externas, cuando sea necesario.

Al igual que con la automatización, la aplicación de la IA en la protección de datos también tiene algunos obstáculos. Ejemplo de ello fue lo sucedido con la base de datos Biostar 2 en el 2019, cuando ésta expuso más de un millón de huellas dactilares de los usuarios de diversas instituciones públicas y privadas por sistemas de seguridad deficientes (Taylor, 2019). Asimismo, los casos de suplantación de identidad donde se emplean máscaras 3D para confundir a los algoritmos de visión artificial son muestras de las falencias actuales que tienen los sistemas de reconocimiento facial (Amada et al., 2021). De igual forma, otro caso fue lo sucedido en el Perú con la filtración de datos de más de 25 millones de peruanos que estaban alojados en los servidores del RENIEC y Ministerio del Interior debido a un acceso indebido y que terminaron en un foro de comercialización ilegal de información digital (Silva, 2025)

Ciberseguridad

Se define como el resguardo de los activos digitales frente a los ataques digitales que pueden aprovechar vulnerabilidades existentes en el sistema (Chio y Freeman, 2018). Si consideramos los ciberataques cada vez más sofisticados, las tecnologías tradicionales de ciberseguridad son a menudo insuficientes, basta con que los ciberdelincuentes encuentren una vulnerabilidad explotable para que puedan generar daños. En ese sentido, el aprendizaje automático permite analizar patrones de tráfico y detectar anomalías para predecir ataques basándose en datos históricos, y así ajustar las defensas ante un atentado. De esta manera, las soluciones de IA permiten una respuesta más rápida a incidentes, reduciendo el tiempo que los atacantes tienen para causar daño y facilitando el trabajo de los equipos de seguridad.

Un ejemplo de este tipo de aplicación es la IA Cylance de BlackBerry, un *software* que a través de aprendizaje automático previene ataques de *malware* siendo hasta un 95% más ligero que otras soluciones como las firmas, lo que le permite ser implementado sin añadir sobrecarga a los sistemas. Por dichas ventajas, esta herramienta de IA ya ha sido implementada por diversas organizaciones empresariales y entidades gubernamentales a nivel mundial quienes respaldan su eficacia para predecir ataques con precisión (Giamatteo, 2023). Otra solución de ciberseguridad basada en IA es Microsoft 365 Defender, que en 2021 bloqueó más de 9,6 mil millones de amenazas de *malware*, 35,7 mil millones de correos electrónicos maliciosos y de *phishing* y 25,6 mil millones de intentos de secuestrar cuentas de clientes dirigidas a dispositivos de empresas y de consumidores (Microsoft, 2022).

Estos modelos de aprendizaje y predicción operan mediante la adquisición de flujos de datos de tráfico de red, registros de actividades y diferentes fuentes de información de inteligencia de amenazas (Kwon et al., 2019). Así, utilizan algoritmos de aprendizaje automático supervisado para identificar cualquier correlación maliciosa a través de los patrones de flujo de datos que estiman con el aprendizaje histórico, mientras que emplean el aprendizaje automático no supervisado para identificar cualquier anomalía sin el requisito de un aprendizaje previo para etiquetarse, lo que permite detectar ataques (Chandola et al., 2019). Ahora bien, estas aplicaciones de IA también tienen riesgos, uno de ellos son los ataques adversariales que engañan a los sistemas de IA, obligándolos a realizar predicciones o decisiones incorrectas o imprevistas (Papernot et al., 2016).

Con la creciente ola de ciberataques cada vez más sofisticados y basados en IA a redes gubernamentales, el problema deja de ser solo privado y se convierte también en una preocupación para la defensa nacional (Persi Paoli y Afina, 2025). Como contramedida, ya varios gobiernos han empezado a integrar IA en sus acciones, que sumada a la experiencia y capacidad humana, se espera puedan permitir una reacción más eficaz frente a estas amenazas. Iniciativas lideradas por entidades como la Agencia de Seguridad Nacional de Estados Unidos y la Agencia de Ciberseguridad de la Unión Europea también han impulsado el uso de la inteligencia artificial para identificar patrones de ataque y prevenir incidentes antes de que se materialicen (European Union Agency for Cybersecurity, 2023; Departamento de Seguridad Nacional de los Estados Unidos, 2024). Todo esto refleja cómo la protección del ciberespacio y la defensa nacional están cada vez más interconectadas, y subraya la necesidad de contar con estrategias integrales que combinen tecnología avanzada con marcos regulatorios sólidos.

Defensa nacional

Esta se refiere a la protección de una nación ante amenazas externas e internas, valiéndose para ello de diversas estrategias y tecnologías llevadas a cabo por las fuerzas armadas. En ese sentido, la IA puede potenciar las capacidades ofensivas a través del análisis de inteligencia o el despliegue de sistemas autónomos como drones (Gray y Ertan, 2021). Asimismo, también puede fortalecer las capacidades defensivas al optimizar los procesos de vigilancia, la recopilación de inteligencia y la toma de decisiones ya que el analizar grandes cantidades de datos de diversas fuentes le permite generar información inmediata, lo cual es

muy relevante considerando que el rápido procesamiento y el aprendizaje de nuevos datos es clave en contextos de cambio constante.

En Japón, se ha invertido más de 240 millones de dólares en un programa público-privado para desarrollar herramientas de IA destinadas a contrarrestar ciberataques y mejorar la ciberdefensa nacional, integrando tecnologías de empresas como Mitsubishi y NEC. Este esfuerzo incluye la creación de sistemas avanzados para recopilar información sobre tácticas y procedimientos de ataques cibernéticos, reforzando la seguridad de infraestructuras críticas y militares. Además, Japón participa en proyectos de aeronaves no tripuladas y enjambres de drones asociados a cazas de próxima generación, en colaboración con Estados Unidos y Reino Unido, buscando compensar la escasez de personal militar y aumentar la capacidad operativa mediante sistemas autónomos y colaborativos (Gómez de Ágreda, 2024). La flexibilidad de estas soluciones de IA posibilita su aplicación en entrenamientos y simulaciones militares, permitiendo que los equipos se familiaricen con su funcionamiento, optimicen la toma de decisiones con mayor precisión y evalúen sus capacidades para mejorar su desempeño a través de la retroalimentación. Así, Corea del Sur ha desarrollado una "Inteligencia Artificial Generativa de Defensa" (GeDAI), que será puesta a prueba en ejercicios conjuntos con Estados Unidos, como el ejercicio anual "Freedom Shield". Esta IA generativa se integra en sistemas de comando y control para mejorar la planificación y ejecución de misiones, evaluando su utilidad en escenarios realistas. El Ministerio de Defensa surcoreano destaca que esta tecnología representa un avance pionero para modernizar sus fuerzas armadas y aumentar la efectividad en operaciones conjuntas con aliados (IPDForum, 2024)

Por otro lado, la IA también tiene un potencial considerable para impulsar la mejora o desarrollo de distintas tecnologías estratégicas usadas para las operaciones de defensa, sobre todo en relación con sus capacidades de reconocimiento, permitiendo a las organizaciones que velan por la seguridad y defensa nacional el analizar data a una velocidad y precisión mayor a la posible ofrecida por una revisión tradicional (Geist y Lohn, 2018). Por ejemplo, la visión por computadora permite detectar y seguir en tiempo real movimientos que podrían indicar comportamientos sospechosos.

La tabla 2 resume las aplicaciones revisadas de la IA en distintos campos para la seguridad.

Tabla 2. Cuadro resumen de aplicaciones de IA en seguridad y defensa.

Área de Aplicación	Formas de integración de la IA
Automatización de procesos	-Monitoreo automático de operaciones industriales -Apoyo en la identificación de patrones sospechosos -Análisis de datos para fiscalizaciones
Protección de datos	-Detección de patrones anómalos en los accesos -Encriptación y resguardo de información -Prevención de filtraciones de información
Ciberseguridad	-Detección temprana de ciberataques -Mitigación de riesgos

Defensa Nacional	- Desarrollo de armamento inteligente - Desarrollo de equipo sofisticado (drones y vehículos autónomos) - Monitoreo avanzado del entorno

Implicaciones del uso de la IA en la seguridad

Las diversas aplicaciones de la inteligencia artificial en seguridad y defensa son muestra de la relevancia y utilidad que viene cobrando en los últimos años; sin embargo, su uso no solo tiene desafíos operativos sino también éticos. A continuación, se profundizará en dos de los principales dilemas: los desafíos en su integración y las consecuencias no deseadas de su uso extendido.

Desafíos en la integración de la IA

La integración de la IA en la seguridad y la defensa requiere un cambio interno de las organizaciones que implica no sólo la adopción de nuevas tecnologías, sino también medidas importantes en la infraestructura, la cultura organizacional y las habilidades de las personas a cargo (Vargas, 2023). De esa manera, la adopción de la IA exige una gran inversión en infraestructura tecnológica, incluyendo *hardware* avanzado y plataformas de *software* capaces de manejar grandes volúmenes de datos. Además, las instituciones deben desarrollar nuevas competencias para operar estas tecnologías, lo que incluye la capacitación de personal en el uso de herramientas basadas en IA y el desarrollo de habilidades en análisis de datos y gestión de riesgos. La falta de talento especializado en el manejo de IA es un desafío que muchas organizaciones enfrentan, lo que limita la efectividad y ralentiza su integración.

Asimismo, también existe la necesidad de diseñar nuevas políticas y procedimientos que aseguren la transparencia, la responsabilidad y la equidad en el uso de la IA. De esta manera, su implementación en la seguridad y la defensa debe tomar en cuenta factores como la privacidad de los datos, la protección contra el uso indebido de información y la garantía de que los sistemas operen de manera justa y no sesgada.

Si bien en el sector privado la competencia entre empresas tecnológicas puede marcar los estándares y llevar a que se impulse este proceso de cambio organizacional; en el sector público este proceso de transformación digital recae enteramente en el aparato estatal (Linares y Contreras, 2023). Esto evidencia la importancia de que los agentes públicos cuenten con las capacidades adecuadas para el manejo de este tipo de tecnologías ya que cumplen el rol de tomadores de decisiones (Vargas, 2023).

Efectos no deseados del uso extendido de la IA

La implementación de la IA en sistemas de seguridad y defensa plantea también ciertos dilemas éticos como por ejemplo el sesgo algorítmico. Este dilema señala que los sistemas de IA pueden reproducir prejuicios raciales y de género debido a los datos de entrenamiento como sucede con ciertos sistemas de reconocimiento facial que presentan mayores tasas de error al identificar a personas de piel oscura y mujeres, lo que puede conducir a discriminación en operaciones de seguridad (Raji y Buolamwini, 2019). Otro caso, en el contexto de la vigilancia y la seguridad pública, son los sistemas de reconocimiento facial que muestran tasas de error

más altas para personas de ciertas etnias, lo que puede llevar a prácticas discriminatorias si se confía ciegamente en la tecnología sin reconocer este error (Linares et al., 2024). Para disminuir la probabilidad de este sesgo, se debe garantizar la diversidad en los equipos de desarrollo y utilizar datos representativos que reflejen la pluralidad de la sociedad (European Union Agency for Fundamental Rights, 2020).

Otro efecto por mencionar es la privacidad y el potencial de la IA para facilitar la vigilancia masiva ya que, sin regulaciones adecuadas, existe el riesgo de que estas tecnologías se utilicen para violar la privacidad individual. Por ejemplo, una normativa española sobre IA creó restricciones para el uso del reconocimiento facial, limitándola solo a casos específicos (Pascual, 2025).

Finalmente, un tercer efecto que se debe resaltar es la cada vez mayor dependencia de la IA a pesar de que también es susceptible a errores. Esta dependencia excesiva puede reducir la intervención humana en tareas rutinarias; pero se debe considerar que ello también afectaría la experiencia práctica del personal; quienes podrían ser incapaces de actuar por su cuenta en situaciones críticas.

Conclusiones

La IA ha cobrado mayor protagonismo en todo tipo de actividades y sectores, no siendo la seguridad y defensa una excepción como se pudo describir en las aplicaciones mencionadas previamente. Las aplicaciones inmediatas como la automatización de procesos y el procesamiento de grandes cantidades de datos son una muestra clara del potencial de la IA y su adopción para operaciones o sistemas de monitoreo.

A su vez, la protección de datos es otro aspecto importante para la seguridad y la defensa. Por ello, casos como el United Family Healthcare reflejan la creciente importancia para las empresas, y seguramente los gobiernos, sobre la necesidad de prevenir filtraciones de información y proteger los activos digitales valiéndose de herramientas basadas en IA. Asimismo, la IA también es útil para la ciberseguridad al ofrecer capacidades para predecir ataques a la infraestructura crítica como lo ejemplifican los casos de Blackberry y Palo Alto en instituciones públicas y privadas.

Ahora bien, a pesar de las múltiples ventajas, se debe mencionar también los desafíos de la integración de la IA. En primer lugar, desde un punto de vista operativo, se debe reconocer la dificultad de su implementación en países en vías de desarrollo debido a la necesidad de infraestructura y formación del personal a cargo. En segundo lugar, desde un punto de vista metodológico, se deben también considerar los efectos como la sobredependencia en las automatizaciones y la existencia de sesgos en los algoritmos en la toma de decisiones.

En conclusión, la inteligencia artificial tiene el potencial para fortalecer las prácticas de seguridad y defensa en todos sus niveles; no obstante está lejos de ser una solución infalible sin la adecuada calibración. Una integración eficaz sólo podrá ser tal si sopesa los beneficios y los potenciales riesgos tanto operativos como éticos.

Referencias

Amada, T., Liew, S. P., Kakizaki, K., y Araki, T. (2021). *Universal adversarial spoofing attacks against face recognition*. [Ataques de suplantación adversarial universal contra sistemas de reconocimiento facial]. En IEEE International Joint Conference on Biometrics (IJCB) (pp. 1-7). IEEE. https://doi.org/10.1109/IJCB52358.2021.9484380

Brundage, M., Avin, S., Wang, J., Belfield, H., Krueger, G., Hadfield, G., ... y Anderljung, M. (2020). *Toward trustworthy AI development: mechanisms for supporting verifiable claims*. [Hacia un desarrollo fiable de la IA: mecanismos para respaldar afirmaciones verificables]. https://arxiv.org/pdf/2004.07213

Brynjolfsson, E., y McAfee, A. (2014). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. [La segunda era de las máquinas: trabajo, progreso y prosperidad en una época de tecnologías brillantes]. W. W. Norton & Company.

BBC. (2017, 15 de mayo). *Ciberataque masivo: ¿quiénes fueron los países e instituciones más afectados por el virus WannaCry?* https://www.bbc.com/mundo/noticias-39929920

Chandola, V., Banerjee, A., y Kumar, V. (2010). Anomaly detection for discrete sequences: A survey. [Detección de anomalías en secuencias discretas: un estudio] *IEEE transactions on knowledge and data engineering*, *24*(5), 823-839. https://doi.org/10.1109/TKDE.2010.235

Chio, C., y Freeman, D. (2018). *Machine Learning and Security: Protecting Systems with Data and Algorithms*. [Aprendizaje automático y seguridad: protegiendo sistemas con datos y algoritmos]. O'Reilly Media.

Chui, M., Roberts, R., Yee, L., Hazan, E., Singla, A., Smaje, K., Sukharevsky, A. y Zemmel, R. (2023). The economic potential of generative AI: The next productivity frontier. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier

Cuba, J. C. (2021). Inteligencia artificial en la Seguridad Nacional: Límites legales. En Vera (Ed.), *Ambiente Estratégico 2021: Seguridad, Desarrollo y Defensa Nacional* (pp 25-36). Centro de Altos Estudios Nacionales – Escuela de Posgrado y Centro de Estudios Estratégicos del Ejército del Perú. https://ceeep.mil.pe/2021/12/17/ambiente-estrategico-2021-seguridad-desarrollo-y-defensa-nacional/

Departamento de Seguridad Nacional de los Estados Unidos. (2024). *Leveraging AI to enhance the nation's cybersecurity*. [Aprovechando la IA para mejorar la ciberseguridad del país]. https://www.dhs.gov/science-and-technology/news/2024/10/17/feature-article-leveraging-ai-enhance-nations-cybersecurity

Doshi-Velez, F., y Kim, B. (2017). *Towards a rigorous science of interpretable machine learning*. [Hacia una ciencia rigurosa del aprendizaje automático interpretable]. https://arxiv.org/abs/1702.08608

El Mundo (2015). *Un ciberataque compromete datos de millones de empleados del Gobierno de EEUU*. https://www.elmundo.es/internacional/2015/06/05/5570e5cce2704e635b8b4592.html

European Union Agency for Cybersecurity. (2023). *Threat Landscape Report: AI in Cybersecurity and Emerging Risks. [Informe sobre el panorama de las amenazas: La IA en la ciberseguridad y los riesgos emergentes]*. https://www.enisa.europa.eu/topics/cyber-threats/threat-landscape

European Union Agency for Fundamental Rights. (2020). *Getting the future right – Artificial intelligence and fundamental rights*. [Acertar con el futuro - Inteligencia artificial y derechos fundamentales]. https://fra.europa.eu/en/publication/2020/artificial-intelligence-and-fundamental-rights

Geist, E., y Lohn, A. J. (2018). *How Might Artificial Intelligence Affect the Risk of Nuclear War?* [¿Cómo puede afectar la inteligencia artificial al riesgo de guerra nuclear?]. RAND Corporation. https://doi.org/10.7249/PE296

Gerges, A. (2021). *Extending the boundaries of safety*. [Ampliando los límites de la seguridad]. Shell. https://www.shell.com/what-we-do/digitalisation/digitalisation-in-action/pushing-the-boundaries-of-safety

Goodfellow, I., Bengio, Y., y Courville, A. (2016). *Deep Learning*. [Aprendizaje profundo]. MIT Press. https://www.deeplearningbook.org/

Giamatteo, J. (2023). *Endpoint Security Evolution: Protection and the Rise of Prevention*. [Evolución de la seguridad de los endpoints: La protección y el auge de la prevención]. https://blogs.blackberry.com/en/2023/04/endpoint-security-evolution-protection-and-prevention

Gómez de Ágreda, A. (2024). *IA en la defensa de la República de Corea y de Japón*. Documento de Análisis IEEE 78/2024. https://www.defensa.gob.es/documents/2073105/2278118/ia_en_la_defensa_de_la_republica_de_corea_y_de_japon_2024_dieeea78.pdf/a6305482-09cb-49e6-5c7b-af118fe7b85f?t=1732709904989

Gray, M., y Ertan, A. (2021). *Artificial Intelligence and Autonomy in the Military: An Overview of NATO Member States' Strategies and Deployment*. [Inteligencia Artificial y Autonomía en las Fuerzas Armadas: Una visión general de las estrategias y el despliegue de los Estados miembros de la OTAN]. NATO Cooperative Cyber Defence Centre of Excellence.

IBM. (2023). *Proteger los datos de los pacientes: un acto de cuidado*. https://www.ibm.com/es-es/case-studies/united-family-healthcare

Instituto Nacional de Defensa de la Competencia y de la Protección de la Propiedad Intelectual (2024). *El Indecopi inicia fiscalización por llamadas sin consentimiento o spam telefónico aplicando inteligencia artificial*. https://www.gob.pe/institucion/indecopi/noticias/951375-el-indecopi-inicia-fiscalizacion-por-llamadas-sin-consentimiento-o-spam-telefonico-aplicando-inteligencia-artificial

IPDForum. (2024). South Korea launches a defense AI center to enhance technology capabilities. Indo-Pacific Defense Forum. https://ipdefenseforum.com/2024/07/south-korea-launches-defense-ai-center-to-enhance-technology-capabilities/

Kemp S. (2024). *Digital 2024 April Global Statshot report*. [Informe Digital 2024 April Global Statshot]. Data Reportal. https://datareportal.com/reports/digital-2024-april-global-statshot

Kwon, D., Kim, H., Kim, J., Suh, S. C., Kim, I., y Kim, K. J. (2019). A survey of deep learning-based network anomaly detection. [Un estudio sobre la detección de anomalías en la red basada en el aprendizaje profundo]. *Cluster Computing*, *22*, 949-961. https://doi.org/10.1007/s10586-017-1117-8

Laskar, M. H. (2023). Examining the emergence of digital society and the digital divide in India: A comparative evaluation between urban and rural areas. Front. Sociol. 8:1145221. https://doi.org/10.3389/fsoc.2023.1145221

LeCun, Y., Bengio, Y., y Hinton, G. (2015). Deep learning. *Nature*, *521*, 436-444. https://doi.org/10.1038/nature14539

Linares-Torres, F. (2024). Inteligencia Artificial y Ciberdefensa. *Revista Seguridad y Poder Terrestre – CEEEP, 3*(4), 139–150. https://revistas.ceeep.mil.pe/index.php/seguridad-y-poder-terrestre/issue/view/11

Linares-Torres F., y Contreras-Salazar, K. (2023). Presencia del Estado y plataforma de servicios digitales. Revista de ciencia e investigación en defensa, *4*(2), 19-36. https://doi.org/10.58211/recide.v4i2.103

Linares-Torres, F., Contreras-Salazar, K., & Salazar-Curichimba, B. (2023). Ciudadanía digital: definición y construcción de un índice nacional basado en actividades. *Revista de ciencia e Investigación en defensa*, *4*(3), 6-21. https://doi.org/10.58211/recide.v4i3.144

Linares-Torres, F., Contreras-Salazar, K., y Salazar-Curichimba, B. (2024). Potencial estratégico de la biometría en la seguridad nacional. *Revista Científica Seguridad y Desarrollo*, 2(1). https://doi.org/10.58211/syd.v2i1.16

Microsoft. (2022). *Cyber Signals*. [Ciberseñales]https://news.microsoft.com/cyber-signals/

Paiva, A. (2023). *Reportan que hackers militares chinos tuvieron acceso a las redes de defensa de Japón: el Pentágono respondió*. La Tercera. https://www.latercera.com/tendencias/noticia/reportan-que-hackers-militares-chinos-tuvieron-acceso-a-las-redes-de-defensa-de-japon-el-pentagono-respondio/BCGR5J7REJEVXC7LIY2WQQNVOY/

Palo Alto Networks. (s. f.). What Is the Role of AI in Security Automation? https://www.paloaltonetworks.com/cyberpedia/role-of-artificial-intelligence-ai-in-security-automation

Papernot, N., McDaniel, P., Goodfellow, I., Jha, S., Celik, Z. B., & Swami, A. (2016). Practical black-box attacks against deep learning systems using adversarial examples. https://doi.org/10.48550/arXiv.1602.02697.

Pascual, M. (2025). *¿Hay que etiquetar los videos generados con IA? Preguntas y respuestas sobre la nueva normativa española*. El País. https://elpais.com/tecnologia/2025-03-20/hay-que-etiquetar-los-videos-generados-con-ia-que-pasa-si-no-lo-hago-preguntas-y-respuestas-sobre-la-nueva-normativa-espanola.html

Persi Paoli, G., y Afina, Y. (2025). *AI in the military domain: A briefing note for states*. [La IA en el ámbito militar: Nota informativa para los Estados]. Instituto de las Naciones Unidas de Investigación sobre el Desarme. https://unidir.org/publication/ai-military-domain-briefing-note-states/

Piergallini, F. (2020). *Adversarial attacks: The enemy of artificial intelligence*. [Ataques adversarios: El enemigo de la inteligencia artificial]. Telefónica Tech. https://telefonicatech.com/en/blog/adversarial-attacks-enemy-artificial-intelligence

Raji, I. D., y Buolamwini, J. (2019). *Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial ai products*. [Auditoría procesable: investigación del impacto de la publicación de resultados sesgados sobre el rendimiento de productos comerciales de inteligencia artificial]. En Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (pp. 429-435). https://doi.org/10.1145/3306618.3314244

Silva, D. S. (2025). *Reniec niega hackeo masivo y denuncia al Mininter como responsable del uso indebido de datos*. Infobae. https://www.infobae.com/peru/2025/04/04/reniec-niega-hackeo-masivo-y-denuncia-al-mininter-como-responsable-del-uso-indebido-de-datos/

Taylor, J. (2019). *Major breach found in biometrics system used by banks, UK police and defence firms*. [Importante fallo en un sistema biométrico utilizado por bancos, la policía británica y empresas de defensa]. The Guardian. https://www.theguardian.com/technology/2019/aug/14/major-breach-found-in-biometrics-system-used-by-banks-uk-police-and-defence-firms

Taylor P. (2023). *Amount of data created, consumed, and stored 2010-2020, with forecasts to 2025*. [Cantidad de datos creados, consumidos y almacenados 2010-2020, con previsiones hasta 2025]. Statista. https://www.statista.com/statistics/871513/worldwide-data-created/

Tornaghi, C. (2023). *El dramático ciberataque que puso a América Latina en alerta*. Americas Quarterly. https://americasquarterly.org/article/el-dramatico-ciberataque-que-puso-a-america-latina-en-alerta/

Vargas, F. (2023). Desafíos globales y tendencias para la transformación de las instituciones de seguridad y defensa. En J. Avalos (ed.), *Brújula hemisférica: Desarrollo de capacidades prospectivas y estudios de futuros para las decisiones en seguridad de América Latina* (pp. 153-216). Centro de Estudios Superiores Navales e Instituto de Investigaciones Estratégicas de la Armada de México.

Wünsch, S., y Machhaus, C., (2013). “Octubre Rojo” ataca a computadoras en todo el mundo. DW. https://www.dw.com/es/octubre-rojo-ataca-a-computadoras-en-todo-el-mundo/a-16525533

Zetter, K. (2014). *Countdown to Zero Day: Stuxnet and the Launch of the World's First Digital Weapon*. [Cuenta atrás para el Día Cero: Stuxnet y el lanzamiento de la primera arma digital del mundo]. Crown Publishers.