Año 2024, Vol. 21, Nº 2, pp. 58 - 71.
DOI: https://doi.org/10.35628/resup.v16i2.149

Inteligencia artificial, desinformación y defensa nacional

Artificial intelligence, disinformation and national defence

Recibido: 28 de agosto del 2024 | Aceptado: 04 de noviembre del 2024

Freddy Linares Torres
https://orcid.org/0009-0009-1268-8109

Licenciado en Administración. Magister en Ingeniería de Sistemas. Egresado de la LXXI Maestría en Desarrollo y Defensa Nacional-CAEN. Ha participado del Curso de Dirección Estratégica para la Defensa y Administración de Crisis (CEDEYAC)-Marina de Guerra del Perú, del Programa de Gestión Estratégica del Poder Aeroespacial y Ciberespacio (PROGEPAC)-Fuerza Aérea del Perú.

Email: freddy.linares@neuromatics.la

Resumen: Exacerbado por el rápido avance de la inteligencia artificial (IA), el sector defensa enfrenta nuevos desafíos en el ámbito del ciberespacio. La IA generativa, capaz de crear material falso pero altamente convincente, representa una evolución en el panorama de las amenazas cibernéticas, al representar la fácil elaboración de contenidos multimedia que pueden tener múltiples efectos (Kertysova, 2018). Desde la influencia en la opinión pública hasta la influencia en procesos electorales, estas herramientas pueden ser utilizadas para provocar desestabilizar gobiernos, sembrar discordia y erosionar la confianza en las instituciones democráticas. A medida que los actores estatales y no estatales aprovechan estas tecnologías, la línea entre la producción de información veraz y la desinformación se vuelve cada vez más difusa. Este fenómeno no solo complica los esfuerzos de ciberdefensa, sino que también plantea graves riesgos para la integridad social en una era donde las sociedades transicionan sus actividades de espacios físicos a espacios digitales (Aïmeur et al., 2023). La capacidad de generar noticias falsas, videos deepfake y otros contenidos engañosos a gran escala, plantean un reto significativo a la seguridad nacional y a la actividad del sector público para alcanzar los intereses nacionales (Sługocki & Sowa, 2021), requiriendo estrategias avanzadas de detección y mitigación. Ante el aumento de estos riesgos relacionados a la difusión de cierto tipo de contenidos en medios

digitales, distintos países optan por buscar establecer regulaciones sobre su generación, identificación y distribución. Este artículo explora las profundas implicaciones de la desinformación en internet y se centrará en analizar los avances en la IA generativa para la producción de contenido falso, su potencial como amenaza para las acciones de defensa y seguridad nacional, y el papel de las políticas públicas y la cooperación internacional en la creación de un marco robusto que proteja la soberanía y estabilidad de las naciones en el ciberespacio.

Palabras clave: Inteligencia artificial generativa, manipulación, deepfake, desinformación, ciberdefensa

Abstract: *Exacerbated by the rapid advancement of artificial intelligence (AI), the defense sector faces new challenges in the cyberspace domain. Generative AI, capable of creating fake but highly convincing material, represents an evolution in the cyber threat landscape, representing the easy development of multimedia content that can have multiple effects (Kertysova, 2018). From influencing public opinion to influencing electoral processes, these tools can be used to destabilize governments, sow discord, and erode trust in democratic institutions. As state and non-state actors take advantage of these technologies, the line between the production of truthful information and disinformation becomes increasingly blurred. This phenomenon complicates cyber defense efforts and poses severe risks to social integrity in an era where societies transition their activities from physical to digital spaces (Aïmeur et al., 2023). The ability to generate fake news, deepfake videos, and other deceptive content on a large scale poses a significant challenge to national security and public sector activity to achieve national interests (Sługocki & Sowa, 2021), requiring advanced detection and mitigation strategies. Due to the increase in the number of risks related to the dissemination of certain types of content in digital media, different countries seek to establish regulations on their generation, identification, and distribution. This article explores the profound implications of disinformation on the Internet and focuses on analyzing advances in generative AI for the production of false content, its potential as a threat to national defense and security actions, and the role of public policies and international cooperation in the creation of a robust framework to protect the sovereignty and stability of nations in cyberspace.*

Keywords: *Generative artificial intelligence, manipulation, deepfake, disinformation, cyberdefense*

1. INTRODUCCIÓN

El ciberespacio es un dominio clave en defensa, tal como sucede con los dominios terrestres, marítimos, aéreos, espaciales (Espona, 2022) y cognitivos. No obstante, su importancia relativa ha crecido por la interconexión global, la transversalidad y la dependencia de las instituciones en las tecnologías de la información, factores que lo han convertido en un escenario crítico donde diversos actores pueden operar con facilidad y muchas veces en anonimato. Es por ello por lo que un desequilibrio operacional en las dinámicas y capacidades del Estado en el ciberespacio frente a agentes externos puede tener múltiples efectos para la defensa. Internet es actualmente el medio más poderoso para la transmisión de diversos tipos de contenido, destacando la información en sí misma que se comunica en distintos formatos (posts, videos, fotos, mensajes de texto, entre otros). En ese sentido, la misma ventaja que dan las herramientas para la generación y transmisión de contenidos, por otro lado presentan un riesgo para la información veraz, por lo que actualmente el discernimiento o el fast-check de lo que circula es crucial en el contexto de seguridad y estabilidad de una nación. Si bien es importante fortalecer la capacidad de defensa ante ciberataques también lo son las capacidades para enfrentar las operaciones de desinformación soportadas en IA generativa.

La IA generativa es una potente herramienta para crear contenido multimedia de alta calidad y casi indistinguible de los producidos por humanos. Actualmente existen distintas herramientas digitales basadas en IA que han demostrado su capacidad para crear artículos de noticias, imágenes artísticas, así como fotos y videos montados (conocidos como *deepfakes*) con un grado de realismo sin precedentes. La sofisticación y accesibilidad de estas herramientas han hecho que la producción de contenido sea sumamente más fácil y económica, ampliando significativamente el potencial de su uso para fines como la desinformación[1] . La llegada de la IA generativa ha amplificado esta amenaza de manera exponencial, debido a que otorga la capacidad de generar contenido digital engañoso a gran escala y con alta precisión a costos muy bajos, favoreciendo la conformación de un entorno digital donde la veracidad de la información sea constantemente cuestionada. Considerando lo popular y dinámicos que son las interacciones en los espacios virtuales, esta proliferación de desinformación tiene el peligroso

[1] Definida como la difusión intencionada de información falsa o engañosa para influir en la opinión de otros (National Museum of American Diplomacy, 2023).

potencial de comprometer las comunicaciones digitales, manipular la opinión pública y erosionar la confianza en las instituciones a favor de intereses de terceros. Esta facilidad con la que se puede producir y distribuir contenido falso plantea un desafío significativo para los esfuerzos de defensa y la protección de la integridad social.

El objetivo de este artículo es explorar las profundas implicaciones de la desinformación facilitada por la IA generativa en el contexto de defensa. Se analizarán los avances recientes en tecnologías de IA generativa y su capacidad para producir contenido falso de manera convincente. Además, se evaluarán los riesgos asociados con la desinformación en el ciberespacio y cómo estas amenazas afectan la estabilidad política y social. De esta forma, a través de una investigación exploratoria y revisión bibliográfica, se pretende proporcionar una visión integral de cómo la IA generativa está transformando el panorama de las amenazas cibernéticas y qué aspectos deben considerarse para proteger a las naciones de los riesgos asociados con la desinformación.

2. ANÁLISIS

2.1 Rol de la ciberdefensa en la sociedad digital

En los últimos años, la sociedad ha experimentado un proceso de digitalización sin precedentes. El internet y las tecnologías de la información, que estuvieron disponibles al público general desde los 90s, han revolucionado la forma en que las personas interactúan, trabajan y se informan. A inicio del siglo XXI los espacios digitales (foros, páginas de chats, redes sociales, entre otros) actuaban como espacios sociales alternativos donde solo una parte de la población participaba, pero en la actualidad se han consolidado y han desplazado los espacios físicos de tal forma, que la participación en el mundo digital es una dimensión relevante de la participación de un individuo en la sociedad actual (Laskar, 2023). Esta transición de la participación de las personas hacia nuevos espacios, implicó un cambio en la percepción sobre distintas dinámicas sociales al realizarse bajo condiciones diferentes a los espacios físicos tradicionales. Entre estas diferencias están el anonimato, el acceso a información, la personalización de las experiencias, la distancia ya no es una barrera, entre otros. En sus primeros años del internet, algunas interpretaciones sobre la evolución de los espacios e instituciones destacaban que el poder se concentraría de forma significativa en quienes posean la red y los nuevos espacios digitales donde frecuentan los ciudadanos (Castell, 1996); en cambio, otros argumentaban que internet tendría una significativa capacidad para empoderar al usuario e impulsar la búsqueda

de objetivos personales (Wellman et al., 2003). En la actualidad, internet es una inmensa red compleja que articula distintos espacios digitales y donde el poder no se concentra ni totalmente en los usuarios ni en aquellos que "controlan la red", sino que este balance difiere en distintos tipos de espacios, existiendo de esta forma tanto espacios digitales que son regulados con rigurosidad como otros espacios digitales que ofrecen una amplia libertad a los usuarios.

Por otro lado, la versatilidad y practicidad de los recursos digitales disponibles en internet, como las herramientas de ofimática, la difusión de contenido multimedia, el comercio electrónico o los servicios de videollamadas, facilitaron que estos elementos se posicionaran progresivamente como nuevos estándares para actividades de distintos sectores que se podían realizar en internet (educación, entretenimiento, trabajos, etc.), volviéndolos progresivamente más digitales (Linares et al., 2023). Esta digitalización multisectorial, presente en el sector privado y público, ha impulsado la economía global, mejorando la eficiencia en múltiples áreas y procesos que han logrado integrar óptimamente estos avances tecnológicos en los últimos años. La competitividad actual entre grandes empresas se basa en una gran medida en la integración de las nuevas innovaciones para optimizar sus procesos, como la IA generativa, que se estima agregará a la economía global miles de millones de dólares en valor por mejoras en la productividad (McKinsey & Company, 2023). El desarrollo del perfil digital en la población juega un rol crítico para optimizar el uso de estas nuevas tecnologías en una sociedad, es por ello por lo que se debe profundizar en instrumentos como el Índice de Actividad Digital (InAD Perú) (Neurometrics Behavioral Lab et al., 2023), el cual explora que tan desarrollado se encuentra la dimensión digital entre la población peruana. Sin embargo, la digitalización de la sociedad también ha introducido nuevas vulnerabilidades. La dependencia de activos críticos en sistemas digitales ha creado un entorno donde las ciberamenazas pueden tener consecuencias devastadoras para una sociedad cada vez más digital. Es en este contexto que el concepto de la ciberdefensa toma relevancia.

Según el *National Institute of Standards and Technology* (NIST), el ciberespacio representa la red interdependiente de infraestructuras de tecnologías de la información conformada por internet, redes de telecomunicaciones, computadoras, sistemas de información, sistemas aislados, redes y procesadores, entre otros (NIST, s.f.). Así, la ciberdefensa se refiere a la capacidad militar que permite actuar frente a amenazas o ataques realizados en y mediante el ciberespacio, cuando estos afecten la seguridad nacional (El Peruano, 2019). La ciberdefensa incluye medidas preventivas, de detección y de respuesta a incidentes

para proteger la integridad y disponibilidad de los servicios digitales. Respecto al sector público, esta área abarca tanto la protección de infraestructuras críticas como la defensa de los sistemas militares y gubernamentales, asegurando la continuidad de los servicios públicos digitales y de las operaciones del sector privado. Esta protección es de especial relevancia, al considerar que los gobiernos están ejecutando proyectos de transformación digital que buscan expandir la integración de los recursos digitales en sus procesos y servicios, de tal forma que se transicione de una entidad física a una digital similar a una plataforma de servicios digitales disponibles para ciudadanos así como agentes privados (Linares-Torres & Contreras, 2023). Pero además se debe reconocer los atributos del ciberespacio para canalizar las amenazas tradicionales contra la seguridad nacional que forman también parte de las prioridades de la ciberdefensa, la cual representa el último bastión en el ciberespacio (Castillo, 2024). Entre las principales amenazas que atentan a la seguridad nacional se encuentran los conflictos sociales, el crimen organizado, el tráfico ilícito de drogas y el terrorismo, y todas, en mayor o menor grado, han hecho una transición a los dominios del ciberespacio donde tienen el margen para utilizar las nuevas capacidades digitales para organizarse e impulsar sus actividades (Vega, 2024). Por lo tanto, el ciberespacio no solo representa un campo de batalla para nuevas amenazas, sino lo es también para otras ya conocidas que instrumentalizan nuevos recursos disponibles y sus características a su favor. Por ejemplo, en 2007 Estonia sufrió lo que se consideró como el primer atentado digital contra un Estado: una serie de ataques cibernéticos que paralizaron bancos, medios de comunicación y agencias gubernamentales, que mostraron la necesidad de desarrollar una ciberdefensa robusta (Ottis, 2008). Este episodio fue un punto de inflexión que impulsó planes de ciberdefensa a nivel internacional, aunque principalmente entre aquellos países de primer mundo que tenían un fuerte apoyo en la tecnología. Hoy en día la integración de la ciberdefensa en las estrategias de defensa nacional, permite a los Estados protegerse contra actores malintencionados que buscan explotar las vulnerabilidades del ciberespacio, así como estar a la vanguardia ante posibles amenazas emergentes en los espacios digitales.

2.2 La información en la era digital

Internet ha difundido masivamente el acceso a la información junto con la capacidad de generarla y compartirla. Así, la información es compartida en distintos tipos de formato multimedia (textos, imágenes, audio o videos). Esta transmisión de contenido es la piedra angular de la era digital actual. La evolución

en el volumen de los datos que se generan, procesan y consumen refleja cómo la actividad de los usuarios en internet se ha vuelto cada vez más audiovisual e interactiva. El volumen de datos creados, capturados, copiados y consumidos a nivel global en internet para 2025 se estima que superará la cantidad de 180 zettabytes (equivalente a mil millones de terabytes), una cantidad más de diez veces mayor que en 2015, donde este valor fue 15.5 zettabytes (Taylor, 2023). No obstante, un aspecto negativo a considerar de esta democratización digital es la propagación de información poco confiable o falsa. Por un lado, el anonimato y los recursos disponibles permiten la creación y publicación por parte de los usuarios, de distintos tipos de contenidos en línea sin mayores costos o compromisos. Por otro lado, la alta exposición que puede generarse en internet alrededor de cierto tipo de información o contenido sucede a tal velocidad que le es muy complicado a los agentes y usuarios reaccionar a tiempo. Esto convierte a internet en una corriente frenética de contenido digital de difícil regulación, donde algunos elementos pueden ser ignorados y perderse, mientras otros consiguen un alcance masivo de forma directa (compartir un video) o indirecta (comentar o generar contenido derivado). De esta forma, la velocidad y el alcance de la difusión de información en internet pueden amplificar tanto verdades como falsedades, comprometiendo la percepción y las decisiones del ciudadano.

La alta competitividad en internet por la atención de los usuarios, la aparición de figuras mediáticas independientes (reporteros, influencers, entre otros) y la influencia de los intereses comerciales e ideológicos, ha llevado a que la difusión de información falsa sea uno de los principales problemas de internet (Orús, 2024), especialmente cuando se refiere a temas de moda. Este contenido desinformador puede adoptar varias formas, incluyendo imágenes alteradas, mensajes de texto y noticias falsas ("*fake news*"), sean en formato de videos, imágenes o textos. Si bien algunas noticias falsas pueden responder a fines como la sátira o la parodia, este contenido es especialmente peligroso cuando son noticias fabricadas para la manipulación, publicidad o propaganda (Tandoc et al., 2018). Por ejemplo, durante las elecciones presidenciales de 2016 en Estados Unidos, la difusión de noticias falsas a través de las redes sociales se consideró un factor influyente en la polarización del electorado, quienes eran más propensos a creer aquellas noticias que favorecían a sus candidatos preferidos (Allcott & Gentzkow, 2017). La pandemia de la COVID-19 es otro buen ejemplo de un escenario crítico reciente, donde la información demostró ser una herramienta clave para que los países pudieran mitigar y enfrentar las consecuencias de la enfermedad, pero también donde los ciudadanos estuvieron expuestos a un gran nivel de desinformación en

los canales digitales (Sługocki, W. Ł., & Sowa, 2022). Según un estudio de 2022, la temática sobre la que la población llegó a ver más noticias falsas fue la COVID-19, destacando Norteamérica y América Latina donde la tasa de penetración de este contenido fue de cerca de 55% (Orús, 2022).

Según el Reglamento de la Ley de Ciberdefensa en Perú, un acto hostil en el ciberespacio es "toda acción en y mediante el ciberespacio que atenta contra la seguridad nacional, soberanía, los activos críticos, los intereses nacionales; y con frecuencia son no cinéticos, dificultando la determinación y atribución" (El Peruano, 2024, p3). Desde una perspectiva micro las consecuencias de la desinformación pueden ser cometer un error o tomar una mala decisión; sin embargo, desde un nivel macro la desinformación en línea puede representar una acción hostil legítima en el ciberespacio pues, pese a que no busca dañar activos críticos directamente, buscan alcanzar distintos objetivos específicos entre la sociedad, que pueden llegar a desestabilizar la gobernabilidad de las instituciones desde diferentes puntos y así ir en contra de los intereses nacionales. Cuando las personas no pueden distinguir entre información verdadera y falsa, se socava la confianza en las instituciones y en los medios de comunicación tradicionales, lo que puede llevar progresivamente a la polarización y a la radicalización, debilitando el tejido social. Regresando al caso de la pandemia, la desinformación online relacionada a la COVID-19 mediante noticias falsas y teorías conspirativas, mermaba los esfuerzos de las autoridades de salud para concientizar a las personas sobre medidas adecuadas de prevención de contagio y generó resistencia a las medidas de salud pública (Pennycook et al., 2020). Este fenómeno demuestra que la desinformación puede influir en el comportamiento de los ciudadanos y afectar a sectores específicos, como la salud, mientras compromete las capacidades operativas y esfuerzos del sector público. Si bien no todos estos actos de desinformación estuvieron articulados como un ataque organizado al sector público, no se deben subestimar las intenciones de terceros ni el potencial de sus actos para alimentar una guerra de información en el corto o largo plazo.

2.3 Desarrollo de la IA generativa

Siendo IA aquellas tecnologías que simulan las capacidades de una persona para aprender y realizar una tarea, la IA generativa son aquellas tecnologías que permiten la creación de contenido multimedia (imágenes, audios, videos, texto) que simula a una obra elaborada por una persona real (McKinsey & Company, 2023). Estas IAs generativas han avanzado sostenidamente en los últimos

años, transformando la forma en que se crea contenido digital al ser una opción accesible y que puede ofrecer resultados de alta calidad. Esta tecnología utiliza modelos de aprendizaje profundo (*deep learning*) para procesar gran cantidad de contenido casi indistinguible de los creados por humanos. Uno de los avances más significativos ha sido el desarrollo de modelos de lenguaje natural como GPT-3 (Generative Pre-trained Transformer 3) por OpenAI, que puede generar texto coherente y contextualmente relevante a partir de breves indicaciones textuales, lo que lo hace útil para aplicaciones como traducción, respuestas de preguntas o asistencia en tareas puntuales (Brown et al., 2020). Esta tecnología fue la base para ChatGPT, una de las IAs generativas de texto más populares y que actualmente sigue mejorando sus capacidades con su nuevo modelo GPT-4o que permite además interactuar por voz y que la IA responda simulando una conversación real (OpenAI, 2024).

Los servicios digitales basados en IA generativa ofrecen actualmente a los usuarios crear de una forma sencilla y personalizada contenido digital de su preferencia. Una aplicación popular es la generación de imágenes con detalle similar a la de diseñadores o editores profesionales. Por ejemplo, DALL-E desarrollada por OpenAI que genera imágenes de alta calidad a partir de descripciones textuales (prompt), y por otro lado, StyleGAN de NVIDIA que se especializa en la creación de imágenes realistas de personas, paisajes y objetos que no existen en realidad (Gonzalo et al., 2024). Es este último uso de generación realista el que despierta preocupaciones, dado que si bien tiene usos beneficiosos como la creación de avatares virtuales, edición de fotos o apoyar a generar contenido visual para películas o videojuegos, podría ser usado de forma indebida para crear falsificaciones digitales convincentes como los denominados *deepfakes*, que son videos, imágenes o audios manipulados con IA para crear representaciones falsas pero convincentes de eventos o declaraciones (Chesney & Citron, 2019).

2.4 La amenaza de las falsificaciones inteligentes

Un caso notorio de *deepfake* fue el video manipulado del ex presidente de EE. UU., Barack Obama, creado en 2018 por el investigador de IA, Jordan Peele, para demostrar los peligros de esta tecnología capaz de engañar al público debido a su alto nivel de realismo. En el video, Obama parecía dar un discurso, pero en realidad sus palabras y movimientos fueron completamente generados mediante IA (Chesney & Citron, 2019). Tal como advirtió Peele, en los años recientes los deepfakes visuales comenzaron a ser utilizados para crear confusión en la población e incluso desacreditar a figuras públicas, como fue el video manipulado de la Presidenta de la Cámara de Representantes de EE.UU., Nancy Pelosi, que circuló en 2019, donde

hicieron parecer que estaba ebria o drogada al dar un discurso (The Guardian, 2019). Ese mismo año, surgió también un video *deepfake* donde se utilizó esta tecnología para manipular el rostro de Mark Zuckerber y que concuerde con un audio grabado, para hacer parecer que daba un discurso sobre cómo un hombre podría poseer datos robados de millones de personas (CNN, 2019).

Tratar con el contenido falso o desinformador es un reto actual de grandes redes sociales (como Facebook, Instagram o X), no obstante identificar los *deepfakes* requiere técnicas más avanzadas que analicen estos contenidos de cada vez mayor calidad, en busca de señales de falsificación (Aïmeur et al., 2023). O, en todo caso, realizar una búsqueda para identificar el material original y comprobar que hubo una edición, como la entrevista original de Nancy Pelosi sin velocidad reducida (The Guardian, 2019). Sin embargo, esta tecnología ya es capaz de crear representaciones que no se basan directamente en un elemento multimedia base para crear la falsificación. Un ejemplo son las imágenes de Donald Trump siendo arrestado que circularon en 2023 (Devlin & Cheetham, 2023) las cuales que no se basan en ninguna situación previa, sino que fueron totalmente diseñadas digitalmente.

La desinformación en internet tiene distintos efectos entre la población, como la generación de confusión, manipulación de opiniones, cambio de posturas sobre ciertos temas, entre otros. Los principales efectos del uso de la IA para la desinformación en internet, son el aumento en el volumen de su creación debido a su accesibilidad y el aumento en el nivel de credibilidad que puede alcanzar este contenido debido a la mejora de sus capacidades (Aïmeur et al., 2023). Por ejemplo, será significativamente más fácil crear una publicación falsa que desacredite a algún personaje público usando un bot que redacte una historia lo suficientemente creíble e incluyendo una imagen deepfake relacionada como supuesta prueba. Y si bien, parte la desinformación en línea que usa IA busca principalmente generar morbo y visitas con contenido aparentemente atractivo, pero poco articulado entre sí, el uso de estos recursos puede ser empleado también de forma estratégica por adversarios a los intereses nacionales, para influir en la percepción de la población y agravar amenazas en el ciberespacio, como el terrorismo o el crimen organizado (Castillo, 2023).

El primer uso de *deepfake* en un contexto de guerra fue en 2022, con la difusión en redes sociales de un video *deepfake* de baja calidad (*cheapfake*), que mostraba al presidente de Ucrania, Volodymyr Zelensky, solicitando la rendición de las fuerzas armadas de su país ante Rusia (The Telegraph, 2022). En ese caso, la confusión inicial entre la población y autoridades fue superada luego de dar las aclaraciones correspondientes por medios de comunicación oficiales, pero es un ejemplo del alcance e impacto que pueden tener estos contenidos multimedia para comprometer

el accionar de los agentes públicos durante escenarios críticos. Estas herramientas permiten preparar y ejecutar distintas acciones de desinformación en internet para alcanzar algún objetivo en el corto (crear confusión entre agentes objetivos) o largo plazo (fomentar ideas sesgadas, rumores o manipular la opinión de ciertos objetivos sobre un tema sensible), lo que compromete significativamente la credibilidad de las comunicaciones en línea en una época donde los medios digitales son una de las principales fuentes de información (García-Ull & Quirós-Fons, 2022).

Estas amenazas son globales y dada la complejidad inherente de internet, no existe una única solución para enfrentar la desinformación en línea y sus efectos para la seguridad (Kertysova, 2018), por lo que se requieren de esfuerzos conjuntos

TABLA 1
Malos usos de la IA y los deepfakes

Uso de IA/Deepfakes	Consecuencias
Creación de noticias falsas en época electoral	Sembrar desconfianza y manipular la opinión pública para influir en las elecciones favoreciendo a un candidato en específico.
Generación de videos o imágenes *deepfake* de figuras públicas	Desacreditar o sesgar la opinión sobre figuras públicas para sembrar desconfianza y erosionar la credibilidad.
Falsificación de declaraciones oficiales	Confundir a la población, crear caos en los medios y desacreditar las políticas públicas en distintos sectores.
Alteración de evidencia en juicios	Influenciar decisiones judiciales, obstruir la justicia y desacreditar a las autoridades judiciales.
Campañas de desinformación en redes	Polarizar a la sociedad, radicalizar opiniones y fomentar divisiones sociales a partir de un tema controversial (por ejemplo, migración, enfermedades, o contaminación).
Suplantación de identidades	Realizar fraudes, encubrir a los actores individuales de un ataque, inculpar o suplantar a autoridades, obstruir la justicia.
Difusión de teorías conspirativas	Deslegitimar instituciones, aumentar la desconfianza en el gobierno, generar paranoia y descreditar las políticas públicas.
Manipulación de imágenes en medios	Cambiar la percepción de eventos, difundir propaganda, alterar la narrativa pública.
Desinformación sobre emergencias	Obstaculizar esfuerzos de socorro, sembrar pánico durante desastres naturales o crisis.
Creación de contenido pornográfico falso	Extorsionar a individuos, dañar reputaciones de figuras públicas, comprometer la salud mental de los objetivos.

Fuente: Elaboración propia.

y articulados entre países aliados. La tabla No.1 incluye diversas consecuencias de uso de estas tecnologías aplicadas a la desinformación. El reconocimiento de la relevancia de la amenaza de la desinformación inteligente para la seguridad en el ciberespacio, es el primer paso que deben hacer los países para preparar las medidas adecuadas. Sin embargo, ello no solo debe enfocarse en la formación de capital humano especializado, sino también involucrar a la ciudadanía, pues son el principal objetivo y medio de difusión de este contenido. Por lo que, así como es necesario desarrollar normativas que regulen el mal uso de estas tecnologías, lo es considerar las contramedidas necesarias para hacer frente a las posibles operaciones de desinformación con IA.

3. CONCLUSIONES

La desinformación basada en inteligencia artificial generativa es un reto global que afecta a la seguridad de las naciones. Estas campañas buscan manipular la opinión pública, influir en procesos electorales y desestabilizar sociedades al sembrar desconfianza en las instituciones. El impacto de estas operaciones no solo se limita a la esfera digital, sino que tiene repercusiones tangibles en la política, la economía y la gobernabilidad. Por ello, el avance de los *deepfakes* plantean nuevos desafíos para el sector interior y defensa. Se debe buscar reforzar y mantener la confianza de la población en sus instituciones, fortaleciéndolas, a través de la formación, meritocracia y eficiencia del gasto. Es crucial que los gobiernos y la sociedad en general reconozcan la gravedad de la amenaza y trabajen juntos para desarrollar estrategias efectivas.

4. RECOMENDACIONES

Para minimizar los riesgos derivados de la desinformación facilitada por la inteligencia artificial generativa, se recomiendan las siguientes acciones:

- Compartir buenas prácticas de detección y neutralización de contenido adulterado.
- Establecer marcos normativos que permitan la cooperación transfronteriza para perseguir el delito.
- Invertir en tecnologías de detección y neutralización de desinformación.
- Fomentar una cultura de verificación de la información y pensamiento crítico en los usuarios de internet, para que puedan discernir mejor entre contenido veraz y falso, así como ser conscientes sobre los riesgos de la desinformación digital y riesgos de la IA.

Referencias

Aïmeur, E., Amri, S. & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: a review. Soc. Netw. Anal. Min. 13, 30. https://doi.org/10.1007/s13278-023-01028-5

Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, 31(2), 211-236. DOI: 10.1257/jep.31.2.211

Castells, M. (1996). The rise of the network society: The information age: Economy, society, and culture (Vol. 1). Oxford: Blackwell Publishers.

Castillo, E. (2024). Impacto del Simposio Internacional "CIBERDEFENSA , CIBERSEGURIDAD, CIBERINTELIGENCIA, DOMINIO COGNITIVO, RETOS Y AMENAZAS DEL CIBERESPACIO" (SIC3RAC). VII SIC3RAC. Ministerio de Defensa del Perú.

Chesney, R., & Citron, D. (2019). Deepfakes and the New Disinformation War: The Coming Age of Post-Truth Geopolitics. Foreign Affairs. https://www.foreignaffairs.com/articles/world/2018-12-11/deepfakes-and-new-disinformation-war

CNN (2019). A deepfake video of Mark Zuckerberg presents a new challenge for Facebook. *CNN Business*. https://edition.cnn.com/2019/06/11/tech/zuckerberg-deepfake/index.html

Devlin, K., & Cheetham, J. (2023). Donald Trump: cómo detectar imágenes creadas por inteligencia artificial como las fotos falsas del arresto del expresidente. *BBC News*. https://www.bbc.com/mundo/noticias-65071726

Espona, J. (2022). Ámbito cognitivo y seguridad nacional: una perspectiva de seguridad interior. En Dirección General de la Guardia Civil, Cuadernos de la Guardia Civil: *Revista de seguridad pública*, Nº 68, 2022, págs. 31-51. https://biblioteca.guardiacivil.es/cgi-bin/koha/opac-detail.pl?biblionumber=23137

García-Ull, F., & Quirós-Fons, A. (2022). CAPÍTULO 4. INTELIGENCIA ARTIFICIAL Y POSVERDAD EN TIEMPOS DE GUERRA. En Romero-Domínguez, L.; Sánchez-Gey Valenzuela, N. (Eds). *Sociedad digital, comunicación y conocimiento: retos para la ciudadanía en un mundo global*, p 73- 90. Madrid, Dykinson, 2022. http://digital.casalini.it/9788411220828

Gonzalo J. Aniano Porcile, Jack Gindi, Shivansh Mundra, James R. Verbus, Hany Farid (2024). Finding AI-Generated Faces in the Wild. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) *Workshops*, 2024, pp. 4297-4305

Kertysova, K. (2018). Artificial Intelligence and Disinformation: How AI Changes the Way Disinformation is Produced, Disseminated, and Can Be Countered. Security and Human Rights, 29(1-4), 55-81. https://doi.org/10.1163/18750230-02901005

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. doi: 10.3389/fsoc.2023.1145221

Linares-Torres, F., & Contreras, K. (2023). Presencia del Estado y Plataforma de Servicios Digitales. *Revista De Ciencia E Investigación En Defensa – CAEN*, 4(2), 19–36. https://doi.org/10.58211/recide.v4i2.103

Linares-Torres, F., Contreras-Salazar, K., & Salazar-Curichimba (2023). Ciudadanía digital: definición y construcción de un índice nacional basado en actividades. *Revista De Ciencia E Investigación En Defensa – CAEN*, 4(3), 6–21. https://doi.org/10.58211/recide.v4i3.144

McKinsey & Company (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

Murphy G, Ching D, Twomey J, Linehan C (2023). Face/Off: Changing the face of movies with deepfakes. PLoS ONE 18(7): e0287503. https://doi.org/10.1371/*journal.pone*.0287503

National Institute of Standards and Technology (s.f.). Cyberspace. COMPUTER SECURITY RESOURCE CENTER, Glossary. https://csrc.nist.gov/glossary/term/cyberspace

National Museum of American Diplomacy (2023). What Is Disinformation? https://diplomacy.state.gov/teacher-resources/what-is-disinformation-video/

Neurometrics Behavioral Lab, Linares-Torres, F., Contreras-Salazar, K., Salazar-Curichimba, B., Contreras-Pulache, H., & Monge, M. (2023). Índice de Actividad Digital (InAD Perú). *Neurometrics*. https://doi.org/10.5281/zenodo.10208356

OpenAI (2024). Hello GPT-4o. https://openai.com/index/hello-gpt-4o/

Orús, A. (2022). Porcentaje de población que vio información falsa o engañosa sobre temas seleccionadas a nivel mundial en 2022, por región. *Statista*. https://es.statista.com/estadisticas/1346841/tasa-de-penetracion-de-las-noticias-falsas-o-enganosas-por-region-y-tematica/

Orús, A. (2024). Noticias falsas y desinformación en el mundo - Datos estadísticos. *Statista*. https://es.statista.com/temas/10126/fakes-news-y-desinformacion/#topFacts

Ottis, R. (2008). Analysis of the 2007 Cyber Attacks against Estonia from the Information Warfare Perspective. Proceedings of the 7th European Conference on Information Warfare and Security, Plymouth, 2008. Reading: Academic Publishing Limited, pp 163-168. https://ccdcoe.org/library/publications/analysis-of-the-2007-cyber-attacks-against-estonia-from-the-information-warfare-perspective/

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. *Psychological Science*, 31(7), 770-780. https://doi.org/10.1177/0956797620939054

Sługocki, W. Ł., & Sowa, B. (2021). Disinformation as a threat to national security on the example of the COVID-19 pandemic. *Security and Defence Quarterly*, 35(3), 63-74. https://doi.org/10.35467/sdq/138876

Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining "Fake News": A Typology of Scholarly Definitions. *Digital Journalism*, 6(2), 137-153. https://doi.org/10.1080/21670811.2017.1360143

Taylor P. (2023). Amount of data created, consumed, and stored 2010-2020, with forecasts to 2025. *Statista*. https://www.statista.com/statistics/871513/worldwide-data-created/

The Guardian (2019). Real v fake: debunking the 'drunk' Nancy Pelosi footage - video | Nancy Pelosi. https://www.theguardian.com/us-news/video/2019/may/24/real-v-fake-debunking-the-drunk-nancy-pelosi-footage-video

The Telegraph (2021). Deepfake video of Volodymyr Zelensky surrendering surfaces on social media. *Youtube*. https://www.youtube.com/watch?v=X17yrEV5sl4

Vega, N. (2024). Operaciones cibernéticas del ejercito en la seguridad nacional. VII SIC3RAC. Ministerio de Defensa del Perú.

Wellman, B., Quan-Haase, A., Boase, J., Chen, W., Hampton, K., Díaz, I., & Miyata, K. (2003). The social affordances of the Internet for networked individualism. *Journal of Computer-Mediated Communication*, Volume 8, Issue 3. http://dx.doi.org/10.1111/j.1083-6101.2003.tb00216.x