IA Generativa y Ciberseguridad

Gobernanza, riesgos emergentes y oportunidades para un futuro digital responsable

DOI: 10.29236/sistemas.n177a11

Resumen

La Inteligencia Artificial Generativa (GenAI) se ha consolidado como el catalizador más influyente de la transformación digital contemporánea. Su capacidad para producir contenido original, automatizar análisis complejos y expandir las capacidades cognitivas humanas está redefiniendo procesos empresariales, educativos y gubernamentales. Sin embargo, esta tecnología también reconfigura el panorama de amenazas: informes recientes, como el Microsoft Digital Defense Report 2025 (Microsoft, 2025), el IBM Threat Intelligence Index 2025 (IBM, 2025) y el ENISA Threat Landscape 2025 (ENISA, 2025), evidencian un repunte en ataques hiperpersonalizados, deepfakes operativos y variantes de malware asistidas por IA. Paradójicamente, la GenAI también es la clave para fortalecer los mecanismos de defensa, pues acelera la detección de anomalías y optimiza la respuesta ante incidentes. Este artículo analiza esa dualidad bajo la óptica del NIST Cybersecurity Framework 2.0 (NIST, 2024), con énfasis en la función Govern como base de la gobernanza algorítmica. Asimismo, se examina el uso ético de la tecnología en la academia y la empresa, junto con su impacto en la brecha digital global. Finalmente, se proponen líneas de investigación críticas para un entorno donde la velocidad del cambio exige decisiones informadas.

Palabras clave

IA Generativa, Ciberseguridad, Gobernanza, Responsabilidad, Resiliencia.

Introducción

La Inteligencia Artificial Generativa ha dejado de ser una novedad para convertirse en el núcleo del ecosistema digital actual. Los modelos fundacionales, capaces de generar texto, código, audio y video, han permeado todos los sectores y han alterado no solo la operación diaria, sino la esencia misma de los modelos de negocio. Accenture (Accenture, 2024) describe este fenómeno como la "reinvención del núcleo digital", un escenario donde la IA evoluciona de ser una herramienta de soporte a un amplificador sustancial de las capacidades humanas.

No obstante, este avance ocurre en un terreno hostil donde los riesgos mutan a la misma velocidad. Los actores maliciosos instrumentalizan la GenAI para automatizar la ingeniería social, generar contenidos sintéticos indistinguibles de la realidad y escalar sus operaciones con una eficiencia sin precedentes. Investigaciones recientes confirman una aceleración de ataques que explotan sesgos cognitivos humanos e introducen desinformación a escala industrial (Microsoft, 2025).

Simultáneamente, las organizaciones despliegan GenAI para enriquecer el análisis de seguridad, contextualizar amenazas y reducir los tiempos de respuesta. Este equilibrio dinámico entre riesgo y oportunidad demanda nuevas estrategias de gobernanza, ética y formación. El presente artículo aborda esta complejidad desde una perspectiva integral que entrelaza tecnología, educación y equidad digital.

Oportunidades y riesgos: La doble cara de la GenAI

La incorporación de la IA Generativa marca un punto de inflexión en la ciberseguridad al modificar drásticamente los tiempos, la escala y la sofisticación tanto del ataque como de la defensa.

La perspectiva ofensiva

La GenAI ha reducido significativamente las barreras de entrada para campañas avanzadas. Los adversarios ahora automatizan tareas críticas de la cadena de ataque:

- **Ingeniería social a escala:** Generación de *phishing* hiper-personalizado que imita la sintáxis y el tono de directivos o entidades de confianza.
- **Engaño sintético:** Uso de *deepfakes* de audio y video para manipular decisiones corporativas, eludir verificaciones biométricas o erosionar la reputación institucional (IBM, 2025).
- **Desarrollo de amenazas:** Aunque los modelos comerciales poseen salvaguardas, la IA ayuda indirectamente a explorar vulne-

rabilidades y crear variantes de malware polimórfico, lo que acelera el desarrollo de Tácticas, Técnicas y Procedimientos (TTPs).

La perspectiva defensiva

En contraparte, la defensa experimenta mejoras sustanciales en eficiencia analítica:

- **SOC Aumentados:** Los Centros de Operaciones de Seguridad procesan volúmenes masivos de datos para detectar patrones sutiles que escaparían al análisis humano tradicional.
- **Respuesta y recuperación:** La GenAI facilita la priorización de alertas, sugiere acciones de contención y sistematiza la documentación posterior al incidente, liberando al talento humano para tareas de mayor valor estratégico (Accenture, 2025).
- Esta naturaleza dual subraya la urgencia de fortalecer los controles y da paso al análisis de los marcos de referencia que deben guiar esta adopción.

Gobernanza algorítmica y el NIST CSF 2.0

El Cybersecurity Framework 2.0 del *National Institute of Standards and Technology* (NIST, 2024) posiciona a la función *Govern* (Gobernar) en el centro de su modelo y enfatiza que la ciberseguridad es una responsabilidad estratégica de liderazgo. En la era de la GenAI, esta función se vuelve vital para orquestar una arquitectura de decisiones que gestione el "riesgo algorítmico".

Bajo la función *Govern*, las organizaciones deben:

1. Instituir políticas de IA Responsable que definan el uso aceptable y ético.
2. Delimitar roles claros a lo largo del ciclo de vida de los modelos.
3. Evaluar vectores de ataque específicos, como los documentados en el marco ATLAS de MITRE (MITRE, 2023), que incluyen envenenamiento de datos, extracción de modelos e inyección de prompts.
4. Asegurar la supervisión humana (*Human-in-the-loop*) en decisiones críticas.
5. Asegurar la trazabilidad y auditoría de los resultados automatizados.

La GenAI potencia transversalmente las demás funciones del marco: mejora la clasificación de activos en *Identify*, endurece configuraciones en *Protect*, afina la sensibilidad en *Detect*, y agiliza la comunicación en *Respond* y *Recover*. Sin embargo, sin el eje rector de *Govern*, estas mejoras carecen de sostenibilidad y seguridad.

Ética y responsabilidad: Academia y Empresa

La IA Generativa ofrece un potencial inmenso, pero introduce desafíos éticos y cognitivos que no pueden ignorarse.

En el **ámbito académico**, la GenAI puede democratizar el acceso a tutorías personalizadas y explicar conceptos complejos. Sin embargo, organismos como la UNESCO (UNESCO, 2023) y la OCDE (OECD, 2023) advierten sobre el riesgo de atrofia en el pensamiento crítico y la dificultad para verificar la autoría. Un uso responsable exige rediseñar las evaluaciones para privilegiar el razonamiento humano, la creatividad y la argumentación, habilidades que la IA aún no puede replicar con autenticidad.

En el **entorno empresarial**, la integración de GenAI impulsa la reinvención operativa. No obstante, esto requiere una gestión rigurosa para evitar sesgos algorítmicos, proteger la propiedad intelectual y evitar la fuga de datos en modelos públicos. La transparencia y la interpretabilidad (*Explainable AI*) son requisitos no negociables para mantener la confianza de clientes y empleados.

GenAI y la brecha digital global

La adopción de GenAI no ocurre en el vacío, sino en un contexto de desigualdad estructural. El Global Cybersecurity Outlook 2025 (WEF, 2025) destaca una bifurcación clara:

- **Países Desarrollados:** Cuentan con ecosistemas de innovación robustos, talento especializado y marcos regulatorios maduros, condiciones que les permiten liderar iniciativas de "IA Soberana".
- **Países Emergentes:** Enfrentan brechas de conectividad y una escasez crítica de talento en ciberseguridad, documentada por ISC2 (ISC2, 2025). La dependencia de proveedores tecnológicos externos aumenta su vulnerabilidad y limita su autonomía digital.

Pese a esto, la GenAI podría ser una herramienta para cerrar brechas si se utiliza estratégicamente en salud, agricultura y educación. El éxito dependerá de políticas públicas inteligentes, cooperación internacional y una inversión sostenida en infraestructura digital propia.

Conclusiones

La IA Generativa no es una tecnología más; se trata de un cambio de paradigma que redefine la seguridad, la educación y la gobernanza. Su impacto final dependerá de nuestra capacidad para adoptarla bajo principios de responsabilidad y estrategia. La función Govern del NIST (NIST, 2024) emerge como el pilar fundamental para gestionar esta transición.

La desigualdad digital plantea el reto más grande: asegurar que la GenAI funcione como un puente hacia el desarrollo y no como un muro que aísle a las economías emergentes. Frente a un horizonte donde lo sintético amenaza la confianza, nuestra respuesta defensiva debe ser **REAL**: basada en la **R**esponsabilidad de la gobernanza,

la **E**valuación continua de riesgos, una **A**rquitectura de seguridad robusta y la **L**ucidez humana para discernir la verdad. Solo combinando estos elementos transformaremos el riesgo algorítmico en una ventaja estratégica.

Interrogantes para el próximo año

- ¿Cómo evolucionarán los ataques impulsados por agentes totalmente autónomos (Accenture, 2025)?
- ¿Qué mecanismos de supervisión humana serán indispensables para preservar la confianza institucional?
- ¿Podrán los países emergentes desarrollar estrategias que reduzcan su dependencia tecnológica?
- ¿Qué estándares globales prevalecerán para garantizar un uso transparente de la GenAI?

Líneas de investigación futura sugeridas

- Modelos de gobernanza algorítmica adaptados a países de baja madurez institucional.
- Implementación de prácticas MLSecOps para asegurar el ciclo de vida completo de la IA.
- Desarrollo de herramientas de auditoría algorítmica asistidas por la propia GenAI.
- Impacto de los agentes autónomos en infraestructuras críticas.
- Evolución del talento en ciberseguridad frente a la automatización.

Referencias

Accenture. (2024, January 16). *Securing the digital core: Elevating cybersecurity for the AI-driven enterprise*. Retrieved from Accenture: https://www.accenture.com/us-en/insights/cybersecurity/securing-digital-core

Accenture. (2025, July 10). *Empowering a secure autonomous AI future*. Retrieved from Accenture Security Blog: https://www.accenture.com/us-en/blogs/security/empowering-secure-autonomous-ai-future

Accenture. (2025, June 26). *State of Cybersecurity Resilience 2025*. Retrieved from Accenture: https://www.accenture.com/us-en/insights/security/state-cybersecurity-2025

ENISA. (2025, October 01). *ENISA Threat Landscape 2025*. Retrieved from ENISA: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2025

IBM. (2025, February 24). *X-Force Threat Intelligence Index 2025*. Retrieved from IBM: https://www.ibm.com/reports/threat-intelligence

ISC2. (2025, September 09). *2025 Cybersecurity Hiring Trends: Skills Deep Dive*. Retrieved from ISC2: https://www.isc2.org/Insights/2025/09/cybersecurity-hiring-trends-skills-deep-dive

Microsoft. (2025, October 16). *Microsoft Digital Defense Report 2025*. Retrieved from Microsoft Security Insider: https://www.microsoft.com/en-us/security/security-insider/intelligence-reports/microsoft-digital-defense-report-2025

MITRE. (2023, August 01). *ATLAS: Adversarial Threat Landscape for AI Systems*. Retrieved from MITRE ATLAS: https://atlas.mitre.org/

NIST. (2024, February 26). *The NIST Cybersecurity Framework (CSF) 2.0*. Retrieved from NIST: https://www.nist.gov/cyberframework

OECD. (2023, November 22). *OECD Digital Education Outlook 2023: Global Standards for AI in Education*. Retrieved from OECD iLibrary: https://www.oecd.org/en/publications/oecd-digital-education-outlook-2023_c74f03de-en.html

UNESCO. (2023, September 07). *Guidance for generative AI in education and research*. Retrieved from UNESCO Digital Library: https://unesdoc.unesco.org/ark:/48223/pf0000386693

WEF. (2025, January 13). *Global Cybersecurity Outlook 2025.* Retrieved from World Economic Forum: https://www.weforum.org/publications/global-cybersecurity-outlook-2025/

Juan Mario Posada *es Senior Manager en Accenture con más de 20 años de experiencia en ciberseguridad IT/OT y gestión de riesgos en Latinoamérica y EE. UU. Especialista en sectores críticos, cuenta con certificaciones como CISSP, CISM, CDPSE, CRISC y CISA. Ha asesorado empresas a nivel regional y global en la transformación estratégica de programas de ciberseguridad. Actualmente lidera la innovación y expansión de servicios de seguridad ciberfísica para Latinoamérica desde Houston.*