2026

Centro de Respuestas a Incidentes Informáticos
CSIRT Académico UNAD

Inteligencia Artificial, Gobernanza y Ciberseguridad Institucional

RETOS ESTRATÉGICOS PARA UN CSIRT ACADÉMICO

VIEM
Vicerrectoría de Innovación y Emprendimiento

ECBTI
Escuela de Ciencias Básicas, Tecnología e Ingeniería

Semillero de Investigación
Ceros y Unos

E-boletín Informativo CSIRT Académico UNAD

Medio de divulgación: Correo Electrónico, Sitio Web

Incluye: Noticias, alertas o informes relacionados con la disciplina de la ciberseguridad

Número treinta y nueve [39]
Mayo de 2026

Universidad Nacional Abierta y a Distancia (UNAD)

Universidad Nacional Abierta y a Distancia
Calle 14 sur No. 14-23 | Bogotá D.C
Correo electrónico:
csirt@unad.edu.co
Página web: https://csirt.unad.edu.co

Estado legal:
Periodicidad: Mensual
ISSN: 2806-0164

Edición electrónica, financiada por la Universidad Nacional Abierta y a Distancia (UNAD)

Vicerrectoría de Innovación y Emprendimiento (VIEM)
Ing. Andrés Ernesto Salinas
Vicerrector

Escuela de Ciencias Básicas Tecnología e Ingeniería (ECBTI)
Ing. Claudio Camilo González Clavijo
Decano

Maestría en Ciberseguridad (ECBTI)
Ing. Sonia Ximena Moreno Molano
Líder de Programa

Especialización en Seguridad Informática
Ing. Cesar Antonio Villamizar
Lider de Programa

Semillero de Investigación Ceros y Unos, adscrito al Grupo de Byte InDesign

Centro de Desarrollo Tecnológico
CSIRT Académico UNAD
Ing. Luis Fernando Zambrano Hernández
Líder CSIRT Académico UNAD

Responsable de la Edición
Ing. Luis Fernando Zambrano Hernandez

Revisó
Adm. Libardo Cárdenas Corral
Analista CSIRT Académico UNAD

Contenido

Introducción

La inteligencia artificial (IA) se ha consolidado como uno de los principales motores de transformación tecnológica, económica e institucional del siglo XXI. Su incorporación en procesos académicos, empresariales, gubernamentales y operativos ha acelerado la automatización, el análisis de información y la toma de decisiones basada en datos, modificando profundamente la relación entre tecnología, gobernanza y seguridad digital. El (Stanford University, 2026) advierte que el crecimiento de las capacidades de IA ocurre a una velocidad superior a la capacidad institucional para gobernarlas, supervisarlas y protegerlas adecuadamente.

En este escenario, la IA presenta un comportamiento dual dentro de la ciberseguridad: mientras fortalece capacidades de detección, automatización y análisis predictivo, también amplifica amenazas relacionadas con phishing automatizado, generación de malware, deepfakes y manipulación de información. Esto obliga a las organizaciones a replantear sus modelos tradicionales de gestión del riesgo tecnológico y resiliencia institucional.

Bajo este contexto, la gobernanza de inteligencia artificial emerge como un componente estratégico para garantizar transparencia, trazabilidad, protección de datos y control sobre tecnologías cada vez más autónomas y complejas. Asimismo, los equipos CSIRT adquieren un papel fundamental en la identificación, prevención y respuesta frente a riesgos emergentes asociados a IA.

El presente documento desarrolla un análisis académico sobre la relación entre inteligencia artificial, gobernanza y ciberseguridad institucional, abordando los principales riesgos tecnológicos derivados de la IA, el rol estratégico de los CSIRT y la necesidad de fortalecer capacidades institucionales orientadas a resiliencia digital y gestión integral del riesgo tecnológico.

Inteligencia Artificial, Gobernanza y Ciberseguridad Institucional:

Retos Estratégicos para un CSIRT Académico

Autores

Luis Fernando Zambrano Hernandez
Docente Investigador
Líder CSIRT Académico UNAD
Universidad Nacional Abierta y a Distancia
ORCID: 0000-0002-4690-3526

Hernando José Peña Hidalgo
Docente Investigador
CSIRT Académico UNAD
Universidad Nacional Abierta y a Distancia
ORCID: 0000-0002-3477-2645

Sonia Ximena Moreno Molano
Líder Maestría en Ciberseguridad
Universidad Nacional Abierta y a Distancia
ORCID: 0000-0003-0392-1983

Néstor Raúl Cárdenas Corral
Analista CSIRT Académico UNAD
Universidad Nacional Abierta y a Distancia
ORCID: orcid.org/0000-0003-3691-0148

Inteligencia artificial y transformación institucional

La inteligencia artificial (IA) se ha consolidado como una de las tecnologías con mayor capacidad de transformación sobre los entornos sociales, económicos, gubernamentales y educativos. Durante los últimos años, especialmente a partir del auge de los modelos generativos y de los sistemas basados en aprendizaje profundo, la IA dejó de ser un componente experimental para convertirse en un elemento estructural de la transformación digital institucional. El crecimiento acelerado de herramientas capaces de automatizar tareas cognitivas, generar contenido, analizar grandes volúmenes de información y apoyar procesos de toma de decisiones ha modificado de manera significativa la relación entre tecnología, gobernanza y seguridad digital.

El AI Index Report 2026 elaborado por el Stanford Human-Centered Artificial Intelligence (Stanford University, 2026) señala que la expansión de la IA está ocurriendo a un ritmo superior al desarrollo de mecanismos institucionales de regulación, evaluación y control. Este fenómeno ha generado preocupaciones relacionadas con la seguridad, la privacidad, la transparencia algorítmica y la capacidad de las organizaciones para gestionar adecuadamente los riesgos tecnológicos asociados a estas tecnologías.

En el ámbito institucional, la IA se está incorporando de forma progresiva en procesos administrativos, financieros, académicos, industriales y gubernamentales. Las organizaciones utilizan modelos de IA para automatizar servicios, mejorar la eficiencia operativa, fortalecer capacidades analíticas y optimizar procesos de

atención y monitoreo. Sin embargo, esta adopción también incrementa la superficie de exposición digital y genera nuevos escenarios de riesgo asociados a dependencia tecnológica, manipulación de información, vulnerabilidades algorítmicas y ataques potenciados mediante inteligencia artificial.

Estudios recientes han evidenciado que la IA posee un comportamiento dual dentro de los ecosistemas digitales: mientras fortalece capacidades defensivas y de análisis, también amplifica el alcance y sofisticación de amenazas cibernéticas. En este sentido, (Jada & Mayayise, 2024) identifican que la IA impacta la ciberseguridad organizacional en todo su ciclo de vida, mejorando procesos de automatización, inteligencia de amenazas y resiliencia digital, aunque simultáneamente introduce riesgos asociados a ataques adversariales y problemas derivados de la calidad de los datos utilizados por los modelos.

La transformación institucional impulsada por IA también ha modificado los modelos tradicionales de gobernanza tecnológica. Las organizaciones ya no solamente deben administrar infraestructura y servicios digitales, sino también modelos algorítmicos capaces de tomar decisiones automatizadas, procesar datos sensibles y generar contenido con alto nivel de realismo. Bajo este escenario, la gobernanza de IA emerge como un componente estratégico para garantizar principios de transparencia, trazabilidad, responsabilidad y seguridad.

(De Almeida & Dos Santos Júnior, 2025) sostienen que las organizaciones públicas que implementan inteligencia artificial requieren modelos multilayer de gobernanza que integren estrategias institucionales, políticas de IA, gestión de riesgos, protección de datos y procesos de formación para tomadores de decisiones y desarrolladores. Los autores identifican que la ausencia de estructuras formales de gobernanza limita significativamente la capacidad institucional para implementar IA de manera ética y segura

Así mismo, la relación entre IA y ciberseguridad ha adquirido relevancia geopolítica y estratégica. El crecimiento de tecnologías automatizadas aplicadas a defensa, vigilancia, análisis masivo de información y operaciones cibernéticas ha llevado a diversos países a considerar la IA como un componente central de seguridad nacional. (Khan et al., 2024) señalan que la IA se ha convertido en un nuevo escenario de confrontación digital, donde las capacidades ofensivas y defensivas evolucionan de forma simultánea, impulsadas por tensiones geopolíticas y competencias tecnológicas internacionales.

En este contexto, las amenazas potenciadas mediante IA han incrementado su complejidad. Actualmente, herramientas generativas pueden ser utilizadas para desarrollar campañas avanzadas de phishing, automatizar procesos de ingeniería social, generar código malicioso o producir contenido falso

altamente creíble mediante técnicas de deepfake y manipulación multimedia. Estos escenarios generan desafíos importantes para las capacidades tradicionales de detección, análisis y respuesta ante incidentes de seguridad.

La expansión institucional de la IA también ha generado preocupaciones relacionadas con privacidad y protección de datos. Los modelos generativos requieren grandes volúmenes de información para entrenamiento y operación, lo que incrementa los riesgos de exposición de información sensible, fugas de datos y uso indebido de información personal. (Ye et al., 2024)advierte que los actuales modelos regulatorios y mecanismos de protección de datos resultan insuficientes frente a las dinámicas de procesamiento utilizadas por sistemas de IA generativa, especialmente en escenarios donde existe reutilización masiva de datos y limitada trazabilidad sobre el origen de la información utilizada para entrenamiento.

Desde el punto de vista organizacional, este panorama obliga a replantear los modelos institucionales de gestión del riesgo tecnológico. Las instituciones educativas, entidades gubernamentales y organizaciones privadas enfrentan actualmente el desafío de adoptar tecnologías de IA sin comprometer principios fundamentales de seguridad, privacidad, continuidad operativa y gobernanza digital. En consecuencia, la IA ya no puede ser analizada únicamente como una herramienta de innovación tecnológica, sino como un componente estratégico con implicaciones directas sobre la estabilidad institucional y la ciberseguridad.

Asimismo, la creciente integración entre IA y procesos industriales ha fortalecido la necesidad de desarrollar modelos sostenibles de ciberseguridad. (Lezzi et al., 2025) destacan que la incorporación de IA en escenarios de evaluación de riesgo y protección industrial permite mejorar capacidades predictivas, clasificación de amenazas y detección de patrones anómalos, aunque también exige mecanismos robustos de control y supervisión para evitar vulnerabilidades derivadas de automatización excesiva o decisiones algorítmicas opacas.

En este escenario, las instituciones requieren fortalecer capacidades interdisciplinarias que integren gobernanza, gestión del riesgo tecnológico, protección de datos y ciberseguridad institucional. La velocidad con la que evoluciona la IA demanda estructuras organizacionales capaces de responder de manera dinámica a riesgos emergentes, manteniendo equilibrio entre innovación tecnológica y protección institucional.

Finalmente, el panorama actual evidencia que la inteligencia artificial no representa únicamente un avance tecnológico, sino una transformación estructural sobre la manera en que las organizaciones administran información, gestionan riesgos y protegen sus activos digitales. La convergencia entre IA, gobernanza y

ciberseguridad configura uno de los principales retos estratégicos contemporáneos para instituciones públicas, privadas y académicas, especialmente para equipos CSIRT encargados de fortalecer capacidades preventivas, analíticas y de respuesta frente a amenazas cada vez más automatizadas y complejas.

ILUSTRACIÓN 1.

LA DOBLE CARA DE LA INTELIGENCIA ARTIFICIAL: OPORTUNIDADES QUE FORTALECEN, RIESGOS QUE AMENAZAN.

Zambrano Hernández, L. F., Peña Hidalgo, H. J., Moreno Molano, S. X., & Cárdenas Corral, N. R. (2026). La doble cara de la inteligencia artificial: oportunidades que fortalecen, riesgos que amenazan [Imagen generada con inteligencia artificial].

Riesgos tecnológicos y amenazas de ciberseguridad asociadas a la inteligencia artificial

La rápida adopción de tecnologías basadas en inteligencia artificial ha generado importantes beneficios para las organizaciones en términos de automatización, análisis de información y fortalecimiento de capacidades operativas. Sin embargo, este crecimiento también ha incrementado significativamente los riesgos tecnológicos y las amenazas de ciberseguridad asociadas a entornos digitales cada vez más automatizados y dependientes de modelos algorítmicos avanzados.

Uno de los principales desafíos actuales radica en el carácter dual de la IA dentro de la ciberseguridad. Mientras las organizaciones utilizan sistemas inteligentes para fortalecer mecanismos de defensa, detección de anomalías y análisis predictivo, actores maliciosos también aprovechan estas tecnologías para desarrollar ataques más sofisticados, automatizados y difíciles de detectar. (Nobles, 2024) advierte que la "weaponización" de la IA se ha convertido en una preocupación creciente debido a la capacidad de los ciberdelincuentes para explotar vulnerabilidades en modelos de aprendizaje automático y utilizar herramientas generativas para potenciar operaciones ofensivas.

En este contexto, las amenazas impulsadas mediante IA han evolucionado rápidamente. Actualmente, herramientas generativas permiten automatizar campañas de phishing altamente personalizadas, producir deepfakes convincentes y generar código malicioso con niveles de complejidad superiores a los observados en ataques tradicionales. Estas capacidades reducen las barreras técnicas para la ejecución de ciberataques y amplifican el impacto potencial sobre instituciones públicas, privadas y académicas.

Adicionalmente, la integración masiva de modelos generativos ha introducido nuevas superficies de ataque asociadas directamente a la arquitectura y funcionamiento de los sistemas de IA. Entre los principales riesgos emergentes se encuentran los ataques de prompt injection, el model poisoning, la manipulación adversarial y la extracción indebida de información sensible desde modelos entrenados con grandes volúmenes de datos. (Xu et al., 2024) señalan que los grandes modelos de lenguaje (LLM) presentan importantes desafíos relacionados con privacidad, interpretabilidad, calidad de datos y seguridad operacional, especialmente en escenarios donde los modelos son utilizados para automatizar procesos críticos de ciberseguridad.

ILUSTRACIÓN 2.

ECOSISTEMA DE AMENAZAS IMPULSADAS POR IA: NUEVOS RIESGOS QUE DESAFÍAN LA SEGURIDAD INSTITUCIONAL.

Nota. Imagen generada con apoyo de inteligencia artificial

[1] Prompt Injection: Manipulación de entradas en modelos de IA para alterar su comportamiento, evadir restricciones o extraer información sensible.

[2] Deepfakes: Generación de contenido audiovisual falso altamente

realista para suplantar identidades, desinformar o manipular decisiones.

[3] Model Poisoning: Envenenamiento de datos de entrenamiento para degradar el rendimiento del modelo o inducir comportamientos maliciosos.

[4] Data Leakage: Exposición o filtración accidental de datos sensibles a través de modelos, respuestas o almacenamiento inseguro.

[5] Malware IA: Uso de inteligencia artificial para generar código malicioso más evasivo, adaptable y capaz de automatizar ataques.

[6] APIs Vulnerables: Interfaces mal protegidas que permiten accesos no autorizados, abuso de funcionalidades o extracción masiva de datos.

Otro aspecto crítico corresponde al impacto que la IA tiene sobre el desarrollo de software y la seguridad de aplicaciones. La generación automática de código mediante modelos de IA ha acelerado procesos de programación y automatización, aunque simultáneamente ha incrementado riesgos relacionados con vulnerabilidades, errores de configuración y reutilización insegura de fragmentos de código. (Negri-Ribalta et al., 2024) identifican que el código generado mediante IA puede introducir debilidades de seguridad difíciles de detectar, afectando la integridad del ciclo de desarrollo seguro de software y aumentando la exposición institucional frente a ataques cibernéticos.

La expansión de IA en entornos institucionales también ha fortalecido riesgos relacionados con privacidad y protección de datos. Los sistemas generativos requieren grandes volúmenes de información para entrenamiento y funcionamiento, lo que incrementa la probabilidad de exposición de datos sensibles, fugas de información y uso indebido de contenido protegido. Esta situación resulta especialmente relevante para instituciones educativas y organizaciones públicas que administran datos personales, financieros y académicos de gran sensibilidad.

De igual manera, las organizaciones enfrentan riesgos derivados de dependencia tecnológica y pérdida de control sobre infraestructuras digitales críticas. Muchas soluciones basadas en IA operan mediante servicios cloud o plataformas propietarias administradas por terceros, lo que genera desafíos relacionados con soberanía digital, continuidad operativa y trazabilidad sobre el procesamiento de información institucional. Este escenario incrementa la complejidad de la gestión del riesgo tecnológico y obliga a replantear modelos tradicionales de seguridad institucional.

En el ámbito defensivo, la IA también está transformando las capacidades de detección y respuesta frente a incidentes. (Ofusori et al., 2024) destacan que las técnicas de IA aplicadas a ciberseguridad permiten mejorar procesos de detección de anomalías, identificación de malware y

análisis predictivo de amenazas mediante modelos avanzados de aprendizaje automático. No obstante, los autores advierten que estos sistemas todavía presentan limitaciones importantes relacionadas con explicabilidad, sesgos algorítmicos y dependencia de grandes volúmenes de datos de calidad.

Asimismo, la creciente complejidad de amenazas potenciadas por IA ha comenzado a superar capacidades tradicionales de monitoreo y análisis utilizadas por muchos equipos de seguridad institucional. Los CSIRT y centros de operaciones de seguridad enfrentan actualmente incidentes más dinámicos, automatizados y adaptativos, donde la velocidad de propagación y capacidad de evasión de ataques incrementan significativamente la dificultad de contención y respuesta.

En respuesta a este panorama, diferentes investigaciones resaltan la necesidad de fortalecer enfoques de gobernanza, gestión integral del riesgo y seguridad por diseño en sistemas de inteligencia artificial. (Lezzi et al., 2025) sostienen que las organizaciones requieren integrar mecanismos de supervisión continua, evaluación de riesgos y controles de seguridad específicos para IA con el fin de garantizar resiliencia digital frente a amenazas emergentes.

En consecuencia, los riesgos tecnológicos asociados a la inteligencia artificial no pueden ser abordados únicamente desde perspectivas técnicas tradicionales. La convergencia entre IA, automatización y ciberseguridad demanda capacidades interdisciplinarias que integren gestión institucional, gobernanza digital, protección de datos y análisis avanzado de amenazas. Bajo este escenario, las instituciones deben evolucionar hacia modelos de ciberseguridad adaptativos capaces de responder a riesgos cada vez más complejos, automatizados y potenciados mediante inteligencia artificial.

Gobernanza y rol estratégico del CSIRT

Gobernanza de inteligencia artificial y ciberseguridad institucional

La acelerada incorporación de inteligencia artificial en entornos institucionales ha generado la necesidad de fortalecer mecanismos de gobernanza orientados a garantizar seguridad, transparencia y control sobre tecnologías cada vez más autónomas y complejas. La gobernanza de IA comprende el conjunto de políticas, lineamientos, estructuras organizacionales y controles destinados a regular el uso responsable de modelos inteligentes dentro de las instituciones.

Actualmente, uno de los principales desafíos consiste en que la evolución tecnológica de la IA avanza más rápido que las capacidades regulatorias y de supervisión institucional. El (Stanford University, 2026) advierte que organizaciones públicas y privadas enfrentan

crecientes dificultades para administrar riesgos asociados a privacidad, seguridad y toma automatizada de decisiones. Este panorama exige fortalecer modelos institucionales de gobierno digital y gestión integral del riesgo tecnológico.

La gobernanza de IA también implica establecer mecanismos de trazabilidad y supervisión sobre modelos capaces de procesar información sensible y automatizar procesos críticos. (De Almeida & Dos Santos Júnior, 2025) señalan que las organizaciones requieren modelos integrales de gobernanza que articulen gestión del riesgo, protección de datos, supervisión algorítmica y fortalecimiento de capacidades institucionales. Los autores identifican que la ausencia de marcos formales de gobernanza limita significativamente la capacidad organizacional para implementar IA de forma ética y segura.

Asimismo, la expansión de plataformas generativas y servicios basados en IA incrementa riesgos relacionados con dependencia tecnológica, pérdida de soberanía digital y exposición de información institucional. Muchas soluciones operan bajo infraestructuras cloud y servicios propietarios administrados por terceros, lo que incrementa la complejidad del control institucional sobre datos y procesos automatizados. Bajo este escenario, la gobernanza de IA se convierte en un componente estratégico de ciberseguridad institucional y continuidad operativa.

Rol estratégico del CSIRT frente a riesgos emergentes de IA

La evolución de amenazas potenciadas mediante inteligencia artificial ha redefinido el papel de los equipos CSIRT (Computer Security Incident Response Team) dentro de las instituciones. Tradicionalmente, los CSIRT se enfocaban en monitoreo, análisis técnico y respuesta frente a incidentes de seguridad; sin embargo, la incorporación masiva de IA ha ampliado considerablemente sus responsabilidades.

Actualmente, los CSIRT deben participar activamente en procesos de gestión del riesgo tecnológico, inteligencia de amenazas y evaluación de riesgos asociados a modelos generativos. Las amenazas impulsadas mediante IA incluyen campañas automatizadas de phishing, generación de malware, deepfakes, manipulación multimedia y ataques adversariales contra modelos inteligentes. Es así, que como se menciona de forma anterior (Nobles, 2024) identifica que la "weaponización" de la IA ha incrementado significativamente la sofisticación y capacidad de automatización de operaciones ofensivas en ciberseguridad.

En respuesta a estas amenazas, los CSIRT requieren fortalecer capacidades analíticas y predictivas basadas en automatización inteligente. (Ofusori et al., 2024) destacan que las técnicas de aprendizaje automático permiten mejorar procesos de detección de anomalías, clasificación de incidentes

y análisis de comportamiento malicioso en tiempo real. No obstante, los autores advierten que estos sistemas aún presentan limitaciones relacionadas con explicabilidad, sesgos algorítmicos y dependencia de datos confiables.

De igual manera, los CSIRT institucionales deben asumir funciones preventivas orientadas a reducir riesgos derivados del uso inseguro de plataformas generativas. Esto implica promover campañas de concientización, apoyar la construcción de políticas institucionales de uso de IA y fortalecer mecanismos de protección de datos sensibles. En entornos académicos, estas funciones adquieren especial relevancia debido al uso creciente de herramientas de IA por parte de estudiantes, docentes e investigadores.

Otro reto importante corresponde al monitoreo de riesgos asociados a grandes modelos de lenguaje (LLM). (Xu et al., 2024) señalan que estos modelos presentan desafíos permanentes relacionados con privacidad, interpretabilidad y seguridad operacional, especialmente cuando son utilizados en procesos automatizados de análisis y toma de decisiones. En consecuencia, los CSIRT deben fortalecer capacidades interdisciplinarias que integren ciberseguridad, análisis de riesgos y gobernanza digital.

Estrategias institucionales para la gestión del riesgo tecnológico asociado a IA

La gestión institucional del riesgo tecnológico asociado a inteligencia artificial requiere enfoques integrales que combinen controles técnicos, organizacionales y humanos. Las organizaciones necesitan desarrollar estrategias de ciberseguridad adaptativas capaces de responder a amenazas cada vez más automatizadas y dinámicas.

Desde el componente técnico, las instituciones deben implementar mecanismos de seguridad específicos para entornos de IA, incluyendo monitoreo continuo, segmentación de servicios críticos, protección de APIs, validación de modelos y supervisión de contenido generado automáticamente. (Lezzi et al., 2025) sostienen que la integración de IA dentro de estrategias sostenibles de ciberseguridad fortalece capacidades de resiliencia digital y protección de infraestructuras críticas, aunque exige mecanismos robustos de supervisión y evaluación continua de riesgos.

En el componente organizacional, resulta fundamental construir políticas institucionales orientadas al uso seguro y responsable de herramientas generativas. Estas políticas deben contemplar clasificación de información, gestión de terceros tecnológicos, lineamientos sobre protección de datos y mecanismos de auditoría para servicios basados en IA.

Adicionalmente, la gestión del riesgo asociado a IA requiere fortalecer capacidades humanas y cultura institucional de ciberseguridad. La alfabetización digital sobre riesgos emergentes, desinformación, deepfakes y manipulación automatizada se convierte en un

elemento esencial para reducir vulnerabilidades derivadas del factor humano.

(Negri-Ribalta et al., 2024) advierten que los sistemas de IA aplicados al desarrollo de software también introducen riesgos relacionados con vulnerabilidades de código y debilidades de seguridad difíciles de identificar. Por esta razón, las instituciones necesitan incorporar procesos de validación, revisión segura y evaluación permanente sobre soluciones automatizadas implementadas dentro de sus ecosistemas digitales.

ILUSTRACIÓN 3.

ESTRATEGIA INSTITUCIONAL DE CIBERSEGURIDAD BASADA EN IA: MODELO POR CAPAS PARA FORTALECER LA RESILIENCIA DIGITAL.

Nota. Imagen generada con apoyo de inteligencia artificial a partir de indicaciones de los autores. La figura representa un modelo institucional de ciberseguridad basado en cinco capas estratégicas: personas, procesos, tecnología, gobernanza y monitoreo continuo, orientadas al fortalecimiento de la resiliencia digital institucional.

Resultados esperados:

- Reducción de riesgos e incidentes de ciberseguridad.
- Respuesta más rápida y efectiva ante amenazas.
- Mayor eficiencia operativa mediante automatización inteligente.
- Fortalecimiento de la cultura institucional de seguridad.
- Incremento de la confianza, cumplimiento normativo y resiliencia digital.

En consecuencia, la convergencia entre inteligencia artificial, gobernanza y ciberseguridad redefine los modelos tradicionales de protección institucional. La gestión del riesgo tecnológico asociado a IA exige capacidades interdisciplinarias, supervisión continua y estructuras organizacionales capaces de responder dinámicamente a amenazas emergentes, manteniendo equilibrio entre innovación tecnológica, seguridad digital y resiliencia institucional.

Conclusiones

La inteligencia artificial se ha consolidado como una de las principales tecnologías de transformación institucional contemporánea, impactando procesos académicos, administrativos, financieros y operativos. Su incorporación dentro de los ecosistemas digitales no solamente mejora capacidades de automatización y análisis, sino que también incrementa la complejidad de los riesgos tecnológicos y de ciberseguridad asociados a las organizaciones.

El análisis desarrollado permitió identificar que la IA posee un comportamiento dual dentro de la ciberseguridad: mientras fortalece capacidades defensivas como detección de anomalías, análisis predictivo y automatización de procesos, también amplifica amenazas relacionadas con phishing avanzado, generación automatizada de malware, deepfakes y ataques adversariales contra modelos inteligentes. Este escenario obliga a las instituciones a replantear sus modelos tradicionales de gestión del riesgo tecnológico y protección digital.

Asimismo, se evidenció que la gobernanza de inteligencia artificial representa un componente estratégico para garantizar transparencia, trazabilidad, responsabilidad y seguridad sobre el uso institucional de tecnologías emergentes. La ausencia de marcos formales de gobernanza limita la capacidad organizacional para controlar adecuadamente riesgos asociados a privacidad, protección de datos y dependencia tecnológica frente a plataformas externas.

En este contexto, los equipos CSIRT adquieren un rol fundamental dentro de las estrategias institucionales de resiliencia digital. Su función ya no se limita exclusivamente a la respuesta ante incidentes, sino que se amplía hacia capacidades preventivas, inteligencia de amenazas, monitoreo de riesgos emergentes y fortalecimiento de cultura organizacional en ciberseguridad. La evolución de amenazas potenciadas mediante IA exige equipos interdisciplinarios capaces de integrar análisis técnico, gobernanza digital y gestión del riesgo tecnológico.

Finalmente, la convergencia entre inteligencia artificial, gobernanza y ciberseguridad demuestra que la protección institucional frente a riesgos emergentes no puede depender únicamente de controles tecnológicos tradicionales. Las instituciones requieren modelos integrales y adaptativos que permitan equilibrar innovación, seguridad y continuidad operativa dentro de entornos digitales cada vez más automatizados y complejos.

Recomendaciones

Las instituciones deben establecer políticas formales de gobernanza de inteligencia artificial que definan lineamientos claros sobre uso seguro, protección de datos, trazabilidad y supervisión de herramientas basadas en IA. Estas políticas deben integrarse con los modelos institucionales de gobierno TI y gestión del riesgo tecnológico.

Se recomienda fortalecer las capacidades de los equipos CSIRT mediante procesos de formación especializada en amenazas asociadas a inteligencia artificial, análisis de modelos generativos y automatización defensiva. La evolución del panorama de amenazas exige capacidades técnicas y analíticas orientadas a incidentes cada vez más dinámicos y automatizados.

Igualmente, resulta necesario implementar estrategias institucionales de monitoreo continuo sobre plataformas y servicios basados en IA, incluyendo evaluación de vulnerabilidades, supervisión de APIs, protección de infraestructuras críticas y controles sobre intercambio de información sensible con herramientas externas.

Desde el componente organizacional, se recomienda promover programas permanentes de concientización sobre riesgos emergentes asociados a IA, especialmente en temas relacionados con phishing automatizado, deepfakes, manipulación de información y privacidad digital. La reducción del riesgo institucional requiere fortalecer cultura de ciberseguridad y alfabetización digital en todos los niveles organizacionales.

Finalmente, las instituciones académicas y organizaciones públicas deben impulsar espacios interdisciplinarios de investigación, simulación y análisis estratégico sobre inteligencia artificial y ciberseguridad. La velocidad de evolución tecnológica exige capacidades continuas de actualización, evaluación y adaptación frente a amenazas emergentes que impactan directamente la estabilidad y resiliencia institucional.

Referentes

De Almeida, P. G. R., & Dos Santos Júnior, C. D. (2025). Artificial intelligence governance: Understanding how public organizations implement it. *Government Information Quarterly*, *42*(1), 102003. https://doi.org/10.1016/j.giq.2024.102003

Jada, I., & Mayayise, T. O. (2024). The impact of artificial intelligence on organisational cyber security: An outcome of a systematic literature review. *Data and Information Management*, *8*(2), 100063. https://doi.org/10.1016/j.dim.2023.100063

Khan, K., Khurshid, A., & Cifuentes-Faura, J. (2024). Is artificial intelligence a new battleground for cybersecurity? *Internet of Things*, *28*, 101428. https://doi.org/10.1016/j.iot.2024.101428

Lezzi, M., Montefusco, P., Lazoi, M., & Corallo, A. (2025). AI-based cybersecurity for a sustainable digital industry: Systematic literature review and future research directions. *Journal of Industrial Information Integration*, *48*, 100980. https://doi.org/10.1016/j.jii.2025.100980

Negri-Ribalta, C., Geraud-Stewart, R., Sergeeva, A., & Lenzini, G. (2024). A systematic literature review on the impact of AI models on the security of code generation. *Frontiers in Big Data*, *7*, 1386720. https://doi.org/10.3389/fdata.2024.1386720

Nobles, C. (2024). The Weaponization of Artificial Intelligence in Cybersecurity: A Systematic Review. *Procedia Computer Science*, *239*, 547-555. https://doi.org/10.1016/j.procs.2024.06.206

Ofusori, L., Bokaba, T., & Mhlongo, S. (2024). Artificial Intelligence in Cybersecurity: A Comprehensive Review and Future Direction. *Applied Artificial Intelligence*, *38*(1), 2439609. https://doi.org/10.1080/08839514.2024.2439609

Stanford University. (2026). *Artificial Intelligence Index Report 2026*. https://hai.stanford.edu/assets/files/ai_index_report_2026.pdf

Xu, H., Wang, S., Li, N., Wang, K., Zhao, Y., Chen, K., Yu, T., Liu, Y., & Wang, H. (2024). *Large Language Models for Cyber Security: A Systematic Literature Review* (Versión 5). arXiv. https://doi.org/10.48550/ARXIV.2405.04760

Ye, X., Yan, Y., Li, J., & Jiang, B. (2024). Privacy and personal data risk governance for generative artificial intelligence: A Chinese perspective. *Telecommunications Policy*, *48*(10), 102851. https://doi.org/10.1016/j.telpol.2024.102851

Contáctenos

Correo electrónico: csirt@unad.edu.co
Página web: https://csirt.unad.edu.co

El CSIRT Académico UNAD está siempre disponible para apoyarte ante consultas o inquietudes relacionadas con la protección de la información en la universidad. No dudes en ponerte en contacto con nuestro equipo para recibir asesoría, reportar incidentes o recibir orientación en temas de seguridad digital. ¡Tu seguridad es nuestra prioridad!