Seguridad de la IA: tu ventaja competitiva en el desarrollo de la IA

Una guía que ayuda a los ejecutivos a acelerar la implementación y gestionar el riesgo de la IA

Introducción

A medida que las organizaciones se apresuran a aprovechar el poder transformador de la IA, la adopción suele superar a la seguridad. Por cada iniciativa aprobada, existen incontables casos de de uso de la Shadow AI por parte de los individuos y de los equipos. De hecho, el 85 % de los responsables de la toma de decisiones del sector de informática afirman que los empleados están adoptando herramientas de IA incluso antes de que sus equipos informáticos puedan evaluarlas.[1] Y el 93 % de los empleados admite haber introducido información en herramientas de IA sin aprobación previa.[1]

Mientras tanto, los vectores de ataque nativos de la IA están aumentando un 47 % con respecto al año pasado,[2] a medida que las herramientas de seguridad tradicionales luchan por mantenerse actualizadas. Las nuevas formas de exposición de datos y las deficiencias en el cumplimiento normativo desafían las prácticas de gobernanza existentes. Los líderes ejecutivos se enfrentan a una pregunta crítica: ¿Cómo gestionar los riesgos introducidos por la IA sin bloquear la innovación que esta permite?

Los requisitos son claros. Las organizaciones deben crear un entorno que permita:

- Identificar todo el uso de la IA.
- Proteger todas las comunicaciones de la IA.
- Aplicar todas las políticas de la IA.
- Proteger todos los modelos de IA contra la vulneración.

Cloudflare AI Security Suite permite que las organizaciones aceleren la implementación de la IA al eliminar la incertidumbre en torno al riesgo. Nuestra plataforma unificada protege todo tu ciclo de vida de la IA mediante la detección del uso de la IA, la aplicación de acceso Zero Trust y la protección de los puntos finales web y API con una defensa proactiva contra las amenazas. Con la gobernanza de datos integrada, los equipos pueden innovar con libertad sin comprometer la seguridad.

El 85 %
de los responsables de las decisiones de informática afirman que los empleados están adoptando herramientas de IA incluso antes de que sus equipos de informática puedan evaluarlas.[1]

El desafío de la seguridad de la IA

Los riesgos de la adopción de la IA están impulsando un mercado de rápido crecimiento para las soluciones de seguridad de la IA. Por ejemplo, las plataformas de protección de aplicaciones nativas de la nube (CNAPP) se han centrado en los flujos de trabajo de desarrollo de la IA. Las primeras propuestas como estas son una línea de defensa esencial, pero no son suficientes. Las empresas deben proteger todo el ciclo de vida de la IA, lo que incluye la protección de los sistemas de IA una vez que se están ejecutando.

El alcance total del desafío que plantea la seguridad de la IA se está volviendo cada vez más claro. Hoy en día, los desarrolladores están creando funciones de IA, los empleados están utilizando herramientas de IA externas y los clientes interactúan con aplicaciones impulsadas por la IA. Proteger todos estos diferentes entornos de forma manual es complejo, y hacerlo de manera consistente es aún más difícil.

Prevenir el uso indebido de la IA por parte del personal de trabajo

A medida que el personal de la organización adopta herramientas de IA autorizadas y no autorizadas, los datos y las operaciones confidenciales están en riesgo. Los equipos de seguridad tienen que gestionar:

- **La falta de visibilidad del uso de la IA:** los líderes no suelen tener una visión completa de qué herramientas de IA están utilizando sus empleados, dónde se procesan los datos confidenciales y cómo estos sistemas de IA se conectan a las aplicaciones empresariales. Este punto ciego genera una exposición significativa al riesgo.
- **Los riesgos de seguridad de datos y de cumplimiento normativo:** la IA cambia la forma en que los datos fluyen por tu organización. La información personal, los datos de propiedad exclusiva y los registros de clientes pueden terminar en sistemas de IA y provocar infracciones de cumplimiento normativo o exponer información confidencial.
- **Los controles de acceso para flujos de trabajo de la IA generativa:** no solo se debe gestionar el acceso humano a las herramientas de IA. También se debe gestionar el acceso de la IA agéntica a los servidores MCP y a otros sistemas críticos. Esto requiere nuevos enfoques para la gestión de la identidad y los controles de acceso.

Proteger los modelos y las aplicaciones de IA de acceso público

Ya sea que se desarrollen internamente o se consuman a través de terceros, los sistemas de IA se están convirtiendo en un pilar esencial de la experiencia del cliente y del usuario, y deben protegerse como tales. La lista OWASP Top 10 for LLM Applications destaca las amenazas que las herramientas tradicionales no pueden abordar:

- **Ataques de consumo ilimitado:** similar a los ataques DoS tradicionales, los ataques de consumo ilimitado buscan sobrecargar los LLM con solicitudes que consumen muchos recursos. Los modelos de precios de pago por uso en la nube pueden disparar el impacto financiero de este tipo de ataques, mientras que los usuarios legítimos se enfrentan a una disminución en la calidad del servicio.
- **Envenenamiento de modelos:** los atacantes implantan puertas traseras, sesgos o vulnerabilidades en los LLM al inyectar datos corruptos en conjuntos de datos públicos o repositorios utilizados por los desarrolladores para el entrenamiento de modelos. Un modelo envenenado de esta manera se comporta con normalidad hasta que ciertos desencadenantes específicos activan los comportamientos dañinos.
- **Ataques de inyección de prompts:** una táctica popular para la exfiltración de datos a través de las interfaces de la IA. La inyección de prompts manipula las entradas de los LLM al insertar instrucciones maliciosas dentro de los prompts del usuario o contenido externo, lo que provoca que los modelos ignoren las instrucciones originales y ejecuten los comandos del atacante.
- **Jailbreaking:** los prompts cuidadosamente diseñados, como los escenarios de juego de roles, los comandos de anulación de instrucciones y las estrategias de varias fases, pueden evadir las medidas de seguridad de los LLM para generar contenido prohibido o extraer información confidencial.

Para permitir una rápida innovación en la IA sin poner en riesgo a la organización, recomendamos que los líderes adopten un enfoque de seguridad de la IA que abarque todo el conjunto de requisitos:

- Protección del uso de la IA generativa por parte de los usuarios
- Protección de las aplicaciones y cargas de trabajo con tecnología de IA
- Desarrollo de aplicaciones con IA seguras según el principio de seguridad por diseño

Proteger los flujos de trabajo de desarrollo y entrenamiento de la IA

Los proyectos de IA implican conjuntos de datos masivos, recursos costosos y experimentación iterativa, todo esto crea nuevas superficies de ataque y vulnerabilidades. Los equipos de seguridad deben abordar los siguientes desafíos:

- **La seguridad e integridad de los datos de entrenamiento:** los datos de entrenamiento comprometidos pueden introducir sesgos, puertas traseras o vulnerabilidades que persisten en los modelos de producción. Las organizaciones deben proteger el acceso a los conjuntos de datos de entrenamiento, evitar las modificaciones no autorizadas y garantizar la fuente de los datos durante todo el ciclo de vida del modelo.
- **La gestión de credenciales y de información confidencial:** los flujos de trabajo de desarrollo de la IA requieren acceso a sistemas que incluyen almacenamiento en la nube para grupos de datos, clústeres de procesamiento para entrenamiento, registros de modelos y servicios de terceros. La información confidencial o las claves API protegidas de forma inadecuada pueden exponer credenciales confidenciales, lo que permite el acceso no autorizado a modelos propietarios, datos de entrenamiento o sistemas de producción.
- **Los controles de acceso al entorno de desarrollo:** los ingenieros de IA suelen necesitar privilegios elevados para experimentar con modelos y acceder a datos confidenciales. Sin controles de acceso adecuados, esto puede provocar amenazas internas, exposición accidental de datos o la extracción no autorizada de modelos.

Consolidación de la seguridad con Cloudflare

La adopción exitosa de la IA se centra en facilitar la productividad, no en restringirla. Cuando los equipos utilizan la IA con confianza y con las protecciones adecuadas, innovan con más velocidad y emprenden proyectos más ambiciosos.

Cloudflare AI Security Suite crea un entorno seguro para la innovación en la IA. Al unificar las capacidades de seguridad del perímetro de servicio de acceso seguro (SASE) y de las aplicaciones web, los directores pueden conectar y proteger dos dominios fundamentales:

- Las aplicaciones externas públicas habilitadas para la IA
- Las cargas de trabajo y sistemas de IA privados e internos

Cloudflare AI Security Suite, que se centra en el ciclo de vida completo de la IA, aborda las necesidades de seguridad desde la detección y la gestión de riesgos hasta la protección de datos, y garantiza el acceso de los usuarios y protege las aplicaciones habilitadas para la IA y los flujos de trabajo de desarrollo. Para proporcionar una capa esencial de seguridad de producción en tiempo real, la red global de Cloudflare se sitúa en línea para inspeccionar y filtrar cada interacción de la IA, para proteger los datos de todos los usuarios y las aplicaciones.

En lugar de solo detectar problemas después de que ocurren, Cloudflare evita problemas y bloquea amenazas antes de que lleguen a los modelos de IA. Los equipos de seguridad obtienen la visibilidad que necesitan para adelantarse a las amenazas emergentes.

Funciones principales de Cloudflare

Cloudflare AI Security Suite ofrece una supervisión integral, una protección en tiempo real y una gestión proactiva de riesgos en un servicio unificado para un enfoque holístico de la seguridad de la IA.

Detección y visibilidad integrales de la IA

La seguridad eficaz de la IA depende de un inventario completo y en tiempo real de los recursos y el uso de la IA, que incluye a los recursos de IA autorizados y los no autorizados. La base de Cloudflare AI Security Suite es la supervisión continua y la detección automatizada para identificar todos los modelos de IA, asistentes, agentes e implementaciones de Shadow AI en todos los entornos: público, privado o interno.

Gestión proactiva de los riesgos de la IA

Cloudflare AI Security Suite ayuda a las organizaciones a prevenir ataques mediante la detección y mitigación de vulnerabilidades, configuraciones erróneas y rutas de ataque específicas de la IA, incluidas las que figuran en la lista OWASP Top 10 for LLMs. La puntuación de confianza de las aplicaciones ayuda a priorizar la corrección para que los equipos aborden primero los riesgos más importantes.

Seguridad de las aplicaciones para las aplicaciones con IA

Para ayudar a las operaciones de seguridad (SecOps) a mantenerse al día sobre los últimos vectores de amenazas de la IA, Cloudflare AI Security Suite incluye la detección y mitigación proactivas de amenazas para vulnerabilidades específicas de la IA, errores de configuración y rutas de ataque dentro de los procesos de la IA, incluyendo la protección contra la inyección de prompts, el envenenamiento de datos y el ataque de modelos.

Los firewalls de IA especializados detectan y etiquetan los puntos finales de la API de la IA generativa o agéntica, detectan intentos de filtración de información de identificación personal y bloquean los prompts maliciosos antes de que afecten al rendimiento del modelo de IA o contaminen el modelo con contenido dañino o información errónea.

Acceso Zero Trust para los flujos de trabajo de la IA generativa o agéntica

Los principios de Zero Trust, como el de mínimo privilegio, se aplican tanto al personal de trabajo como a los agentes de IA. Cloudflare AI Security Suite puede aplicar políticas de acceso a la red Zero Trust (ZTNA) tanto para las interacciones humano-IA como para las interacciones IA-IA. El registro y los controles centralizados para los servidores MCP garantizan que la IA agéntica solo acceda a aquello para lo que ha sido autorizada, incluyendo los flujos de trabajo justo a tiempo (JIST).

Protección de datos con tecnología de IA

Una seguridad de IA eficaz requiere una capacidad de prevención de pérdida de datos que aproveche varios modelos lingüísticos para comprender tanto el contenido de un prompt como la intención que subyace a ella. Cloudflare AI Security Suite incorpora capacidades de DLP en el entrenamiento, los prompts y las respuestas para evitar la exposición de la información de identificación personal, la exposición de datos y el acceso no autorizado dentro de los modelos y los procesos de la IA. La seguridad en tiempo de ejecución implementada en línea y centrada en las API, actúa como una primera capa de defensa rápida y sencilla para complementar el enfoque de desplazamiento a la izquierda (“Shift Left”) habilitado por una CNAPP.

Localización de los datos

Los LLM y las aplicaciones de IA están sujetas a las mismas regulaciones que cualquier otro tipo de entorno de datos. Cloudflare AI Security Suite ayuda a las organizaciones a garantizar que las cargas de trabajo de IA respeten las fronteras geográficas y jurisdiccionales mediante políticas que mantienen los datos de entrenamiento y las solicitudes de inferencia dentro de las regiones aprobadas.

Principio de seguridad por diseño para el desarrollo de la IA

Como ocurre con todo tipo de software, la seguridad de las aplicaciones de IA debe integrarse desde el principio del ciclo de desarrollo, no añadirse luego. Cloudflare AI Security Suite ofrece a los desarrolladores herramientas y marcos para crear aplicaciones con tecnología de IA que se basan en el principio de seguridad por diseño.

El impacto comercial de Cloudflare AI Security Suite

El auge de la IA no es solo una evolución, sino una disrupción fundamental del orden de la industrialización o la informatización. Por lo tanto, la adopción rápida y eficaz se convierte no solo en una cuestión de impulsar el crecimiento, sino también de la viabilidad de la organización. La adopción lenta o insegura ahora representa una amenaza existencial para las organizaciones en todas las industrias y sectores.

Cloudflare AI Security Suite permite una transformación segura, controlada y eficiente para satisfacer las crecientes expectativas de los clientes y los requisitos del mercado en entornos altamente competitivos.

- **Aceleración de la innovación de la IA:** cuando la seguridad permite un uso seguro en lugar de bloquear la adopción, los empleados y los equipos pueden explorar nuevas aplicaciones de IA para ser más productivos en su trabajo diario. Los marcos de seguridad adecuados brindan a los desarrolladores la confianza necesaria para crear funciones ambiciosas de IA sin poner en riesgo los datos o los sistemas confidenciales.
- **Reducción del riesgo relacionado con la IA:** las organizaciones pueden gestionar todo el conjunto de riesgos que conlleva la adopción de la IA, incluidos los riesgos relacionados con la IA inherentes a las aplicaciones web habilitadas para la IA. Las aplicaciones de IA en desarrollo se pueden proteger para garantizar que los empleados no filtren datos confidenciales ni los expongan en conjuntos de datos de entrenamiento de IA. Las operaciones de seguridad (SecOps) pueden identificar y mitigar de manera proactiva las amenazas y vulnerabilidades específicas de la IA, y minimizar la superficie de ataque y proteger los datos y modelos críticos.
- **Optimización de las operaciones de seguridad:** la visibilidad y el control centralizados de su estado de seguridad de IA simplifican la gestión y agilizan la respuesta a incidentes. Tu equipo de SecOps puede centrarse en iniciativas estratégicas en lugar de tener que estar constantemente apagando incendios relacionados con la IA.
- **Gobernanza y cumplimiento normativo de datos sólidos:** los controles de protección de datos específicos de la IA ayudan a proteger la información confidencial y a cumplir con los requisitos normativos cambiantes durante todo el ciclo de vida de la IA.
- **Reducción del costo total de propiedad (TCO):** aprovechar una plataforma consolidada que integra las inversiones de seguridad existentes resulta más económico que implementar soluciones específicas separadas para cada desafío de seguridad de la IA.

Casos prácticos de modelos para Cloudflare AI Security Suite

Cloudflare AI Security Suite es ideal para abordar los requisitos esenciales en torno a la adopción de la IA.

- **Protección del uso de herramientas de IA por tu personal de trabajo:** aplica políticas Zero Trust para el acceso del personal tanto a herramientas públicas de la IA generativa, como ChatGPT, como a aplicaciones internas con IA.
- **Protección de las aplicaciones públicas habilitadas para la IA:** protege las aplicaciones web y las API que incorporan modelos de IA, como los bots de chat y los motores de recomendación, de los ataques que podrían exponer datos confidenciales o vulnerar el modelo.
- **Gestión de la Shadow AI:** detecta de forma automática las herramientas de IA no autorizadas en toda tu organización y aplica los controles adecuados para permitir la innovación continua dentro de los límites de riesgo gestionado.
- **Prevención de pérdida de datos (DLP) con IA para interacciones con la IA:** evita la exposición de datos confidenciales en las solicitudes o respuestas de la IA, y garantiza la protección de la información de identificación personal y la información confidencial.
- **Seguridad en el desarrollo de la IA:** proporciona a los equipos de ingeniería marcos que hagan que el desarrollo seguro sea el enfoque predeterminado, que les permita crear funciones de IA rápidamente sin comprometer la seguridad.

Consideraciones para la implementación

Como estrategia de seguridad fundamental, la seguridad de la IA debe abordarse de manera reflexiva.

Iniciar con la infraestructura actual	Las implementaciones de seguridad de IA más exitosas se basan en las herramientas SASE y de seguridad de aplicaciones existentes, en lugar de reemplazarlas. Este enfoque aprovecha las inversiones actuales mientras que amplía la protección para cubrir los riesgos específicos de la IA.
Implementar protecciones unificadas en línea	La capacidad de bloquear la actividad maliciosa en el extremo de la red, en tiempo real, es crucial para la seguridad de la IA. Implementa controles en línea en tiempo real y complementalos con la supervisión basada en API.
Garantizar una cobertura completa	Tu estrategia de seguridad de la IA debe abordar el espectro completo de requisitos: el uso de herramientas de la IA generativa por parte del personal de trabajo, la protección de las aplicaciones y los flujos de trabajo con IA, la protección de flujos de trabajo de la IA agéntica y la seguridad para flujos de trabajo de desarrollo de la IA.
Planificar para una escala empresarial	Elige soluciones que puedan crecer a medida que adoptas la IA. Lo que funciona para un programa piloto debe también ser escalable en toda la organización a medida que se expande el uso de la IA.
Validar los criterios de éxito	Antes de comprometerte por completo, valida la solución en situaciones del mundo real. Elige una plataforma que permita la activación gratuita y la activación de auto-servicio de sus capacidades empresariales. Esto permite probar el paquete de seguridad completo a una escala pequeña, como para un solo equipo o aplicación, para demostrar de forma rápida su valor y garantizar que cumpla con los criterios de éxito.

Próximos pasos con Cloudflare AI Security Suite

A medida que la adopción de la IA se acelera a nivel mundial, las organizaciones que puedan escalar la IA de manera segura tendrán una ventaja competitiva significativa. La clave es reconocer que la seguridad de la IA no se trata de impedir el uso de la IA, sino de habilitarlo de forma inteligente.

Cloudflare AI Security Suite permite a las organizaciones innovar con confianza. Desarrollada sobre la base de nuestras plataformas de seguridad de aplicaciones y SASE de amplia adopción, nuestra solución AI Security Suite ofrece detección de IA integrada, controles de acceso Zero Trust, defensa proactiva contra amenazas basada en inteligencia y gobernanza de datos sólidos. Al proteger el uso y los modelos de la IA desde todos los ángulos, las organizaciones pueden empoderar a los desarrolladores para construir con más velocidad y permitir que los empleados sean más productivos, todo sin sacrificar la experiencia del usuario final.

Solicitar una consulta

Descubre cómo Cloudflare AI Security Suite puede transformar el enfoque de tu organización para asegurar la adopción de la IA.

+55 (11) 3230 4523

enterprise@cloudflare.com

https://www.cloudflare.com/es-la/

1. ManageEngine. The Shadow AI Surge in Enterprises: Insights from the US & Canada
2. Investigación de Check Point. Q1 2025 Global Cyber Attack Report from Check Point Software: An Almost 50% Surge in Cyber Threats Worldwide, with a Rise of 126% in Ransomware Attacks