Seguridad de la IA: tu ventaja competitiva en la carrera de la IA

Guía para directivos para acelerar la adopción 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 avanzar más rápido que la seguridad. Por cada iniciativa aprobada, existen innumerables casos de shadow AI entre los usuarios y los equipos. De hecho, el 85 % de los responsables de TI declaran que los empleados adoptan herramientas de IA incluso antes de que sus equipos de informática puedan evaluarlas.[1] Asimismo, el 93 % de los empleados admiten incluir información en las herramientas de IA sin aprobación.[1]

Mientras tanto, los vectores de ataque nativos de la IA están aumentando (un 47 % con respecto al año pasado),[2] ya que las herramientas de seguridad tradicionales tienen dificultades para seguir el ritmo de estas amenazas. Las nuevas formas de exposición de datos y las deficiencias de cumplimiento normativo ponen en jaque las prácticas de gobernanza actuales. Los responsables ejecutivos afrontan una cuestión fundamental: ¿Cómo gestionar los riesgos que plantea la IA sin bloquear la innovación que ofrece?

Los requisitos están 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 los abusos.

Cloudflare AI Security Suite ofrece a las organizaciones la confianza necesaria para acelerar la adopción de la IA eliminando la incertidumbre del riesgo. Nuestra plataforma unificada protege todo tu ciclo de vida de la IA mediante la detección del uso de la IA, la aplicación del acceso Zero Trust y la protección de los puntos finales de API y web con una protección proactiva contra las amenazas. Con la gobernanza de datos integrada, los equipos pueden innovar sin restricciones y sin poner en riesgo la seguridad.

85 %
de los responsables de TI declaran que los empleados adoptan herramientas de IA incluso antes de que sus equipos de TI puedan evaluarlas.[1]

El desafío de la seguridad de la IA

Los riesgos asociados a la adopción de la IA están impulsando un mercado en rápido crecimiento para las soluciones de seguridad de la IA. Por ejemplo, las plataformas de protección de aplicaciones nativas de nube (CNAPP) se han centrado en los flujos de trabajo de desarrollo de la IA. Las primeras soluciones de este tipo conforman una línea defensiva esencial, pero no son suficientes. Las empresas deben garantizar la seguridad de todo el ciclo de vida de la IA, y esto incluye proteger los sistemas de IA una vez que se ejecutan en el entorno de producción.

El alcance total del desafío que plantea la seguridad de la IA es cada vez más evidente. Hoy en día, los desarrolladores crean funciones de IA, los equipos utilizan herramientas de IA externas y los clientes interactúan con aplicaciones basadas en IA. Proteger manualmente todos estos entornos distintos es complejo, y hacerlo de forma sistemática es aún más difícil.

Evitar el uso indebido de la IA por parte de tu equipo

La adopción por parte de los usuarios de las organizaciones de las herramientas de IA (tanto autorizadas como no autorizadas) pone en riesgo los datos y las operaciones confidenciales. Los equipos de seguridad tienen que lidiar con las problemáticas siguientes:

- **Falta de visibilidad del uso de la IA:** los responsables de la seguridad no suelen tener una visión completa de qué herramientas de IA utilizan sus usuarios, dónde se procesan los datos confidenciales y cómo estos sistemas de IA se conectan a las aplicaciones empresariales. Este punto ciego genera una grave exposición al riesgo.
- **Riesgos relacionados con la seguridad y el cumplimiento normativo de los datos:** la IA cambia la forma en que los datos fluyen en tu organización. La información personal, los datos privados y los registros de clientes pueden terminar en sistemas de IA y causar así incumplimientos normativos o exponer información confidencial sobre la competencia.
- **Controles de acceso para los flujos de trabajo de la IA agéntica:** no solo se debe gestionar el acceso humano a las herramientas de IA. También es necesario gestionar el acceso de la IA agéntica a los servidores MCP y a otros sistemas críticos. Esto requiere nuevos enfoques de la gestión de identidades y los controles de acceso.

Proteger los modelos y aplicaciones de IA públicos

Los sistemas de IA, tanto si se desarrollan internamente como si se consumen a través de terceros, se están convirtiendo en un pilar esencial de la experiencia de los clientes y de los usuarios, y deben protegerse como tales. La lista OWASP Top 10 para las aplicaciones LLM destaca las amenazas a las que las herramientas tradicionales no pueden responder:

- **Ataques de consumo ilimitado:** al igual que los ataques DoS tradicionales, el objetivo de los ataques de consumo ilimitado es sobrecargar los LLM con solicitudes que consumen muchos recursos. Los modelos de precios de pago por uso en la nube pueden disparar las repercusiones económicas de este tipo de ataques, y los usuarios legítimos pueden sufrir una disminución de la calidad del servicio.
- **Envenenamiento de modelos:** los atacantes implantan puertas traseras, sesgos o vulnerabilidades en los LLM inyectando datos alterados en los conjuntos de datos o repositorios públicos que los desarrolladores utilizan para entrenar los modelos. Un modelo envenenado de esta manera se comporta con normalidad hasta que ciertos desencadenantes activan comportamientos dañinos.
- **Ataques de inyección de instrucciones:** la inyección de instrucciones, una táctica habitual para la exfiltración de datos a través de interfaces de IA, manipula los datos especificados en los LLM mediante la inserción de instrucciones maliciosas dentro de las instrucciones de los usuarios o el contenido externo, y esto provoca que los modelos ignoren las instrucciones originales y ejecuten los comandos del atacante.
- **Jailbreak:** las instrucciones diseñadas cuidadosamente (como los escenarios de juego de roles, los comandos de invalidación de instrucciones y las estrategias en varias fases) pueden eludir las protecciones de seguridad de los LLM para generar contenido prohibido o extraer información confidencial.

Para acelerar la innovación de la IA sin poner en riesgo a la organización, recomendamos a los responsables de TI adoptar un enfoque de la seguridad de la IA que abarque toda la gama de requisitos:

- Proteger cómo tu equipo usa la IA generativa
- Proteger las aplicaciones y las cargas de trabajo de IA
- Desarrollar aplicaciones basadas en IA diseñadas para ser seguras

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

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

- **Seguridad e integridad de los datos de entrenamiento:** los datos de entrenamiento no seguros pueden añadir sesgos, puertas traseras o vulnerabilidades que persistan en los modelos de los entornos de producción. Las organizaciones deben proteger el acceso a los conjuntos de datos de entrenamiento, evitar modificaciones no autorizadas y garantizar la procedencia de los datos durante todo el ciclo de vida del modelo.
- **Gestión de credenciales y secretos:** los flujos de trabajo de desarrollo de la IA requieren el acceso a distintos sistemas, como el almacenamiento en la nube para conjuntos de datos, los clústeres de proceso para el entrenamiento, los registros de modelos y los servicios de terceros. Los secretos o las claves API insuficientemente protegidos pueden exponer credenciales confidenciales y permitir así el acceso no autorizado a los modelos propios, los datos de entrenamiento o los sistemas de producción.
- **Controles de acceso al entorno de desarrollo:** los ingenieros de IA suelen necesitar privilegios elevados para experimentar con los modelos y acceder a los datos confidenciales. Sin los controles de acceso adecuados, esto puede resultar en amenazas internas, una exposición accidental de datos o una extracción no autorizada de modelos.

Unificar la seguridad con Cloudflare

Para adoptar con éxito la IA es necesario fomentar la productividad, no restringirla. Cuando los equipos tienen la certeza de que se han implementado las medidas de seguridad adecuadas, pueden utilizar la IA con confianza, innovan más rápido y emprenden proyectos más ambiciosos.

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

- Las aplicaciones externas y públicas basadas en IA
- Las cargas de trabajo y los sistemas de IA internos y privados

Cloudflare AI Security Suite se centra en todo el ciclo de vida de la IA, y aborda todas las necesidades de seguridad (desde la detección y la gestión de riesgos hasta la protección de datos), garantizando el acceso seguro de los usuarios y protegiendo las aplicaciones y los flujos de trabajo de desarrollo basados en IA. Para proporcionar una capa fundamental de seguridad del entorno de producción en tiempo real, la red global de Cloudflare se sitúa en línea para inspeccionar y filtrar todas las interacciones de la IA, protegiendo los datos de todos los usuarios y aplicaciones.

En lugar detectar los problemas cuando ya se han producido, Cloudflare los previene y bloquea las amenazas antes de que lleguen a los modelos de IA. Los equipos de seguridad se benefician de la visibilidad que necesitan para anticiparse a las nuevas amenazas.

Funciones principales de Cloudflare

Cloudflare AI Security Suite ofrece una supervisión integral, una protección en tiempo real y una gestión proactiva de los riesgos en una única ubicación para ofrecerte un enfoque integral de la seguridad de la IA.

Detección y visibilidad integrales de la IA

Una seguridad de la IA eficaz depende de tener un inventario completo y en tiempo real tanto de los recursos de IA y como de su uso (autorizado y no autorizado). Cloudflare AI Security Suite se basa en la supervisión continua y en la detección automatizada para identificar todos los modelos de IA, los asistentes, los agentes y las 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 evitar los ataques gracias a la detección y la mitigación de las vulnerabilidades, los errores de configuración y las rutas de ataque específicos de la IA, incluidas las que figuran en la lista OWASP Top 10 para los LLM. La puntuación de confianza de las aplicaciones te ayuda a priorizar la corrección para que los equipos aborden primero los riesgos más graves.

Seguridad de las aplicaciones basadas en IA

Para ayudar a los equipos de SecOps a mantenerse al día sobre los últimos vectores de amenaza de la IA, Cloudflare AI Security Suite incluye la detección y la mitigación proactivas de las amenazas que representan las vulnerabilidades, los errores de configuración y las rutas de ataque específicos de la IA dentro de las canalizaciones de IA, como la protección contra la inyección de instrucciones, el envenenamiento de datos y el abuso de modelos.

Los firewalls de IA especializados identifican y etiquetan los puntos finales de API y de IA generativa o agéntica, detectan los intentos de exfiltración de información de identificación personal y bloquean las instrucciones maliciosas antes de que afecten al rendimiento de los modelos de IA o envenenen el modelo con contenido tóxico o información errónea.

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

Los principios Zero Trust, como el de privilegio mínimo, se aplican tanto a los usuarios humanos como a los agentes de IA. Cloudflare AI Security Suite puede aplicar las políticas de acceso a la red Zero Trust (ZTNA) tanto para las interacciones entre los usuarios humanos como entre distintos servicios de IA. El registro y los controles centralizados para los servidores MCP garantizan que la IA agéntica solo acceda a los recursos a los que está autorizada, incluidos los flujos de trabajo Just In Time (JIST).

Protección de datos con reconocimiento de IA

Una seguridad de la IA eficaz requiere una función de prevención de pérdida de datos que utilice varios modelos lingüísticos para comprender tanto el contenido de una instrucción como la intención subyacente. Cloudflare AI Security Suite incorpora funciones de prevención de pérdida de datos (DLP) en el entrenamiento, las instrucciones y las respuestas para evitar la exposición de información de identificación personal, la fuga de datos y el acceso no autorizado en los modelos y canalizaciones de IA. La seguridad en tiempo de ejecución, implementada en línea y centrada en las API, actúa como una primera capa de seguridad rápida y sencilla para complementar el enfoque de desplazamiento a la izquierda ("Shift Left") basado en una CNAPP.

Localización de datos

Los LLM y las aplicaciones de IA se rigen por las mismas normativas 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 los límites geográficos y jurisdiccionales gracias a políticas que mantienen los datos de entrenamiento y las solicitudes de inferencia dentro de las regiones aprobadas.

Seguridad por diseño para el desarrollo de la IA

Al igual que con todos los demás tipos de software, la seguridad de las aplicaciones de IA debe integrarse desde el principio del ciclo de desarrollo, no añadirse a posteriori. Cloudflare AI Security Suite proporciona a los desarrolladores las herramientas y los marcos necesarios para crear aplicaciones basadas en IA diseñadas para ser seguras.

El impacto empresarial de Cloudflare AI Security Suite

El auge de la IA no es simplemente una evolución; constituye una revolución radical equivalente a lo que significó la industrialización o la informatización. Por ello, su adopción rápida y eficaz no solo supone fomentar el crecimiento empresarial. También depende de ella la viabilidad organizativa. Actualmente, su adopción lenta o no segura representa una amenaza existencial para las organizaciones de todos los sectores e industrias.

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

- **Aceleración de la innovación de la IA:** cuando los servicios de seguridad protegen el uso de la IA, y no bloquean su adopción, los usuarios y los equipos pueden explorar nuevas aplicaciones de IA para ser más productivos en sus tareas cotidianas. Los marcos de seguridad adecuados brindan a los desarrolladores la confianza necesaria para crear ambiciosas funciones de IA sin poner en riesgo los datos o los sistemas confidenciales.
- **Reducción del riesgo relacionado con la IA:** las organizaciones pueden gestionar toda la variedad de riesgos que conlleva la adopción de la IA, incluidos los riesgos relacionados con la IA inherentes a las aplicaciones web basadas en 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 la IA. Tu equipo de SecOps puede identificar y mitigar de forma proactiva las amenazas y las vulnerabilidades específicas de la IA, minimizando la superficie de ataque y protegiendo los datos y modelos críticos.
- **Optimización de las operaciones de seguridad:** la visibilidad y el control centralizados de tu postura de seguridad de la IA simplifican la gestión y agilizan la respuesta a los incidentes. Tu equipo de SecOps puede centrarse en las iniciativas estratégicas en lugar de tener que estar constantemente solucionando los incidentes relacionados con la IA.
- **Mayor eficacia de la gobernanza y el cumplimiento normativo de los datos:** los controles de protección de datos específicos de la IA te ayudan a proteger la información confidencial y a cumplir los requisitos normativos en constante cambio durante todo el ciclo de vida de la IA.
- **Reducción del coste total de propiedad (TCO):** el uso de una plataforma consolidada que integra las inversiones en seguridad existentes resulta más económico que implementar soluciones específicas individuales para cada desafío relacionado con la seguridad de la IA.

Casos prácticos de modelos para Cloudflare AI Security Suite

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

- **Proteger cómo tus equipos utilizan las herramientas de IA:** aplica políticas Zero Trust para el acceso de los usuarios tanto a herramientas públicas de IA generativa (p. ej., ChatGPT) como a las aplicaciones internas basadas en IA.
- **Proteger tus aplicaciones públicas basadas en IA:** protege tus aplicaciones web y tus API que incluyen modelos de IA (como bots de chat y motores de recomendación) contra ataques que podrían causar la exposición de datos o el abuso del modelo.
- **Gestión del shadow IA:** detecta automáticamente las herramientas de IA no autorizadas en toda tu organización y aplica los controles adecuados para permitir la innovación continua dentro de unos límites de riesgo gestionado.
- **Prevención de pérdida de datos (DLP) basada en IA para las interacciones con la IA:** evita la exposición de tus datos confidenciales en las instrucciones o las respuestas de la IA, garantizando la protección de la información confidencial y de identificación personal.
- **Seguridad en el desarrollo de la IA:** proporciona a los equipos de ingeniería los marcos necesarios para que el desarrollo seguro sea el enfoque por defecto, y que puedan así crear funciones de IA con rapidez y sin poner en riesgo la seguridad.

Consideraciones para la implementación

La seguridad de la IA es una estrategia básica de seguridad, y como tal debe abordarse cuidadosamente.

Comienza con tu infraestructura actual	Las implementaciones de seguridad de la IA que logran los mejores resultados se estructuran sobre las herramientas existentes de seguridad de las aplicaciones y SASE, no las reemplazan. Este enfoque aprovecha las inversiones actuales al mismo tiempo que amplía la protección para responder a los riesgos específicos de la IA.
Implementa protecciones unificadas en línea	Para la seguridad de la IA, es fundamental poder bloquear las actividades maliciosas en el perímetro de la red, en tiempo real. Implementa controles en línea en tiempo real y complétalos con la supervisión basada en API.
Garantiza una cobertura integral	Tu estrategia de seguridad de la IA debe abordar toda la gama de requisitos: cómo tus equipos utilizan las herramientas de IA generativa, la protección de las aplicaciones y los flujos de trabajo basados en IA, cómo proteger los flujos de trabajo de IA agéntica y la seguridad de los flujos de trabajo de desarrollo de la IA.
Planifica a escala empresarial	Elige soluciones que puedan crecer a medida que adoptas la IA. Lo que funciona en un programa piloto también debe ser escalable a toda la organización a medida que se amplíe el uso de la IA.
Valida tus criterios de éxito	Antes de comprometerte por completo, valida la solución en escenarios reales. Elige una plataforma que permita la activación de autoservicio y gratuita de sus funciones para empresas. Esto te permitirá probar todo el conjunto de soluciones de seguridad a pequeña escala (p. ej., para un solo equipo o una única aplicación), con el fin de demostrar rápidamente su valor y asegurarte de que cumple con tus criterios de éxito.

Avanzar con Cloudflare AI Security Suite

La adopción de la IA se acelera en todo el mundo, y las organizaciones que puedan escalar la IA con total seguridad se beneficiarán de una considerable ventaja competitiva. Lo más importante es reconocer que la seguridad de la IA no consiste en impedir su uso, sino en hacerlo posible, pero de manera inteligente.

Cloudflare AI Security Suite permite a las organizaciones innovar con confianza. Nuestra solución AI Security Suite se basa en nuestras plataformas de seguridad para las aplicaciones y SASE, pero va más allá. Proporciona funciones integradas de detección de IA, controles de acceso Zero Trust, protección proactiva contra las amenazas (basada en nuestra información sobre amenazas) y una eficaz gobernanza de datos. Las organizaciones, al proteger el uso y los modelos de IA desde todas las perspectivas, pueden proporcionar a los desarrolladores las herramientas necesarias para crear más rápido, así como mejorar la productividad de los equipos. Y todo ello sin sacrificar la experiencia de los usuarios finales.

Solicita una consulta

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

+34 518 880 290

enterprise@cloudflare.com

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

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