netskope

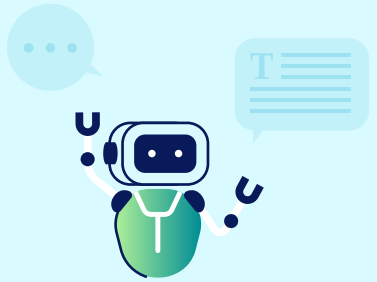
El

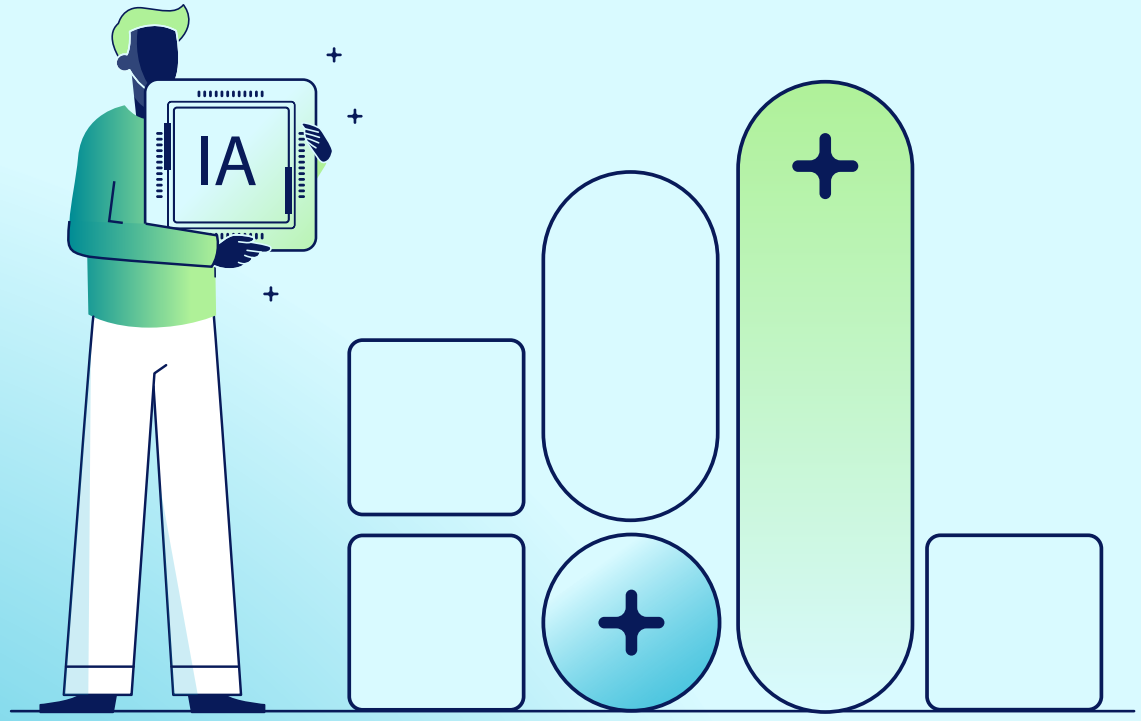
Manual de seguridad de la IA

Una guía práctica para asegurar la IA de extremo a extremo en cualquier lugar

El

Manual de seguridad de la IA





Índice

Introducción	3
Retos de seguridad en la IA	4
Pilares de la seguridad de la IA	5
Navegar por la seguridad de la IA	6
El futuro de la seguridad de la IA	12
Conclusión	13
Acerca de Netskope	14

Introducción

Las tecnologías de inteligencia artificial (IA) se han consolidado rápidamente como herramientas útiles e importantes para muchas organizaciones. Con la aparición constante de nuevas capacidades y casos de uso, la IA es ahora un elemento primordial de las pilas tecnológicas de casi todas las empresas.

El rápido ascenso de la IA ha estado marcado también por unos altos niveles de inversión. Según los analistas de IDC, se calcula que el mercado de gasto informático mundial en IA aumentará hasta casi 750 000 millones de dólares para el año 2028, con un gasto especifico en IA generativa ligeramente por encima de los 300 000 millones de dólares.[1]

Se calcula que el mercado de gasto informático mundial en IA aumentará hasta casi 750 000 millones de dólares para el año 2028, con un gasto específico en IA generativa ligeramente por encima de los 300 000 millones de dólares.

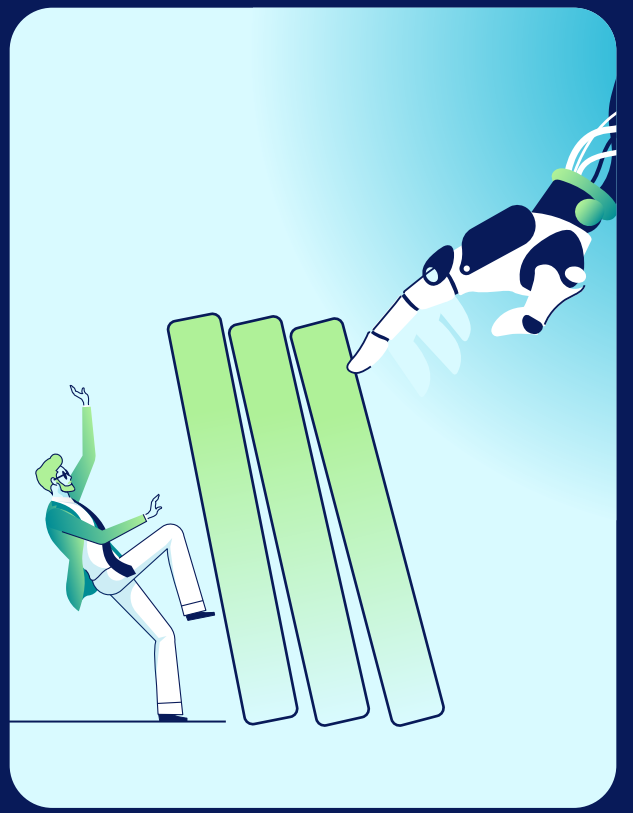
Para los profesionales de la seguridad, los posibles riesgos de las aplicaciones de IA en su entorno son obvios y cada vez mayores. En el nivel más básico de adopción, los datos se comparten con aplicaciones de terceros en la nube.

Desde el punto de vista de la seguridad, esto plantea dudas sobre qué datos introducen los empleados en estos sistemas y qué controles se han establecido para gestionarlos. El avance de los protocolos estándar destinados a facilitar aún más el intercambio de datos con las aplicaciones de IA, como el protocolo de contexto de modelos (MCP), sistematiza estos riesgos.[2]

Es probable que los retos de seguridad se intensifiquen aún más según vaya evolucionando la tecnología empresarial de la IA. Por ejemplo, los sistemas agenciales de IA pueden funcionar de forma autónoma para alcanzar objetivos concretos o ejecutar tareas definidas sin necesidad de intervención humana constante. Los analistas del sector de Gartner prevén que, para 2028, un 25 % de todas las infracciones empresariales estarán relacionadas con el mal uso de agentes de IA.[3]

Teniendo en cuenta la rapidez con la que evolucionan los riesgos a los que se enfrentan los profesionales de la seguridad, no es ninguna sorpresa que estén buscando ayuda para moverse por este nuevo panorama. En este libro electrónico describimos las principales preocupaciones de seguridad de las organizaciones actualmente y las soluciones que Netskope puede proporcionar como ayuda.

Gartner prevé que, para 2028, un 25 % de todas las infracciones empresariales estarán relacionadas con el mal uso de agentes de IA.



[1] IDC Market Forecast, Worldwide Artificial Intelligence IT Spending Forecast, 2024–2028, Rick Villars et al., octubre de 2024, Doc #US52635424.

[2] Informe Cloud and Threat Report de 2025 de Netskope https://www.netskope.com/netskope-threat-labs/cloud-threat-report/cloud-and-threat-report-2025

[3] Principales predicciones de Gartner para 2025.

Retos de seguridad en la IA

Los tres principales problemas a los que se enfrentan actualmente los equipos de seguridad

1 Aumento de la superficie de riesgo

A medida que evoluciona el uso de la IA, de herramientas de IA generativa especializadas (como ChatGPT) a capacidades de IA integradas en todas las aplicaciones de empresa y aplicaciones de IA de creación privada, continúa aumentando la superficie de ataque. Cada etapa introduce nuevos riesgos:

- Las herramientas públicas de la IAGen entrañan riesgos de divulgación inadvertida de datos confidenciales.
- Las funciones de IA integradas en las aplicaciones SaaS existentes pueden dar lugar a fugas o manipulaciones de datos.
- Los LLM de alojamiento privado y las aplicaciones de IA personalizadas introducen nuevos vectores, como controles de acceso mal configurado o vulnerabilidades en las canalizaciones de datos.
- Las conexiones entre las aplicaciones de IA y las fuentes de datos a través de nuevos protocolos, como MCP, amplían la superficie de riesgo de una posible filtración de los datos.

2 Exposición y filtración de datos confidenciales

El riesgo más inminente en la adopción de la IA es la pérdida de datos, tanto accidental como maliciosamente:

- La exposición involuntaria se produce cuando los empleados introducen datos confidenciales (p. ej., información personalmente identificable, secretos comerciales, datos regulados) en modelos públicos sin ser conscientes de las consecuencias.
- El personal interno malintencionado o los atacantes externos pueden explotar las herramientas de IA para filtrar datos o explotar maliciosamente los canales de salida del modelo.
- También existe un riesgo durante el entrenamiento: El uso de datos inadecuadamente seleccionados durante el entrenamiento del modelo puede dar lugar a modelos que filtran información confidencial.

3 Gobernanza responsable de la IA

Con el crecimiento de los sistemas de IA, surgen importantes cuestiones sobre ética y cumplimiento que se entrecruzan con la seguridad:

- Los modelos de IA pueden involuntariamente codificar y propagar sesgos, causando escrutinio normativo y daños reputacionales.
- La manipulación indebida de los datos de empleados y clientes utilizados en los flujos de trabajo de IA puede infringir el RGPD, la HIPAA u otras leyes sobre privacidad de datos.
- El despliegue autónomo de la IA en lugar de la toma de decisiones por seres humanos, especialmente en áreas de alto riesgo (p. ej., contratación, seguridad, finanzas) plantea dilemas éticos y brechas de responsabilidad.

Pilares de la seguridad de la IA

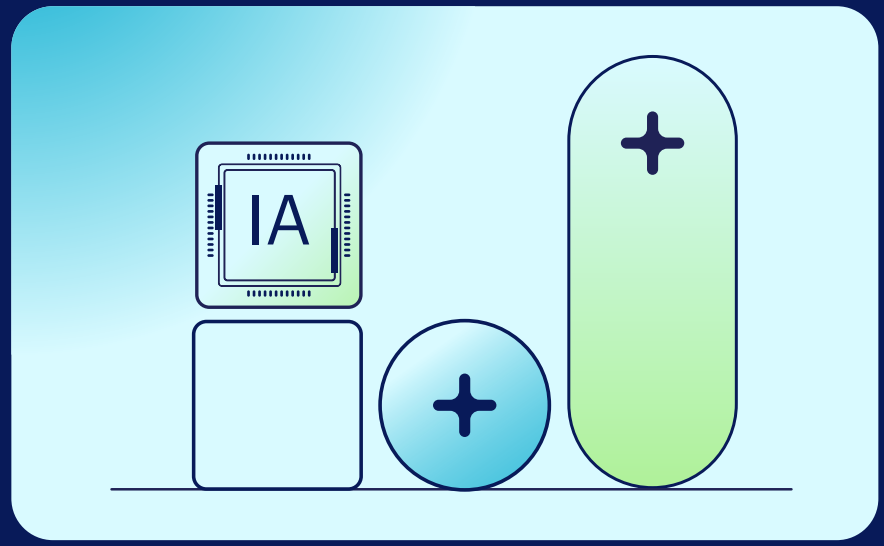
La exigencia de la confianza cero (*zero trust*)

La seguridad de la IA depende de un enfoque de confianza cero, algo muy similar a la seguridad de SaaS, pero con problemas únicos derivados del modo en el que los modelos de IA procesan los datos entrantes y generan los salientes.

Tanto la seguridad de la IA como la del SaaS exigen unos estrictos controles de acceso, una supervisión continua y una sólida protección de los datos para mitigar los riesgos. Sin embargo, mientras que la seguridad de SaaS se centra principalmente en salvaguardar las aplicaciones y las interacciones con el usuario, la seguridad de la IA también debe responder por la integridad de los datos del entrenamiento, el acceso al modelo y la posible manipulación adversa. Esto hace que el cumplimiento de unas políticas de seguridad que tengan en cuenta el contexto y la detección de amenazas en tiempo real sean requisitos esenciales en la prevención de la fuga de datos, el acceso no autorizado y la explotación del modelo de IA.

Un sólido marco de confianza cero para la seguridad de la IA garantiza que se verifiquen todas las solicitudes, se supervisen todos los flujos de datos y se garantice el acceso según unas evaluaciones dinámicas del riesgo en lugar de permisos estáticos. Este enfoque exige una visibilidad pormenorizada del movimiento de los datos y unos controles de seguridad adaptables que se ajusten según el contexto en tiempo real.

Al disponer de unos principios de confianza cero, las empresas pueden adoptar y ampliar las tecnologías basadas en IA con tranquilidad y sin comprometer la seguridad o el cumplimiento.



Consejo profesional

La seguridad de la IA depende de un enfoque de confianza cero, algo muy similar a la seguridad de SaaS, pero con problemas únicos derivados del modo en el que los modelos de IA procesan los datos entrantes y generan los salientes.

Navegar por la seguridad de la IA

Seis principales desafíos y soluciones

Desafío 1: Falta de visibilidad

Con la integración de las herramientas de IA en los flujos de trabajo diario, las organizaciones se enfrentan a un reto fundamental de seguridad: no pueden asegurar lo que no pueden ver.

Los empleados tienen acceso a aplicaciones tanto autorizadas como sin autorizar con credenciales corporativas y personales, lo que difumina la frontera entre usos aprobados y no aprobados. Esta expansión descontrolada aumenta el riesgo de fuga de datos, pérdida de PI e incumplimiento normativo, especialmente cuando se introduce información confidencial en servicios de IA sin gestionar o en la sombra.

La mayoría de las organizaciones carecen de la visibilidad pormenorizada necesaria para diferenciar entre un uso arriesgado de la IA y otro legítimo. Las herramientas tradicionales se quedan cortas a la hora de identificar interacciones concretas con el modelo de IA, distinguir entre cuentas personales y de empresa, y proporcionar información en tiempo real a nivel de usuario, aplicación o actividad. Sin una visibilidad a fondo sobre cómo y dónde se utiliza la IA, los equipos de seguridad no pueden ver los posibles puntos de exposición.

Cómo resuelve esto Netskope

Con el aumento de la adopción de herramientas de IA por parte de las organizaciones, el mantenimiento de la visibilidad y el control de su uso se convierten en factores cruciales. Netskope ofrece una solución completa para hacer un seguimiento de aplicaciones de IA tanto gestionadas como sin gestionar (en la sombra), ofreciendo a los equipos de seguridad la información que necesitan para garantizar una supervisión adecuada.

Entre sus principales funciones se incluyen:

- **Conocimiento avanzado sobre las instancias:** Distinga entre instancias personales y corporativas de las aplicaciones de IA como ChatGPT, Gemini y Copilot.
- **Panel de control de IA**: Obtenga conocimientos detallados sobre las tendencias de uso de la IA, las principales aplicaciones, la frecuencia de acceso y acciones pormenorizadas de los usuarios, como inicios de sesión, publicaciones, cargas y descargas.
- **Análisis del comportamiento de usuarios y entidades (UEBA):** Detecte anomalías y comportamientos de riesgo usando aprendizaje automático para identificar amenazas, como filtración de datos, riesgos internos e infracciones de políticas.
- **Visibilidad fundacional:** Obtenga una visión global de su ecosistema de IA a través del tráfico de usuarios a aplicaciones, API y MCP. Netskope unifica la visibilidad del uso, el inventario y los flujos de datos.

Esta visibilidad global permite a los equipos de seguridad actuar con rapidez y mitigar los riesgos vinculados al uso de la IA en toda la empresa.

Desafío 2: Comprender el riesgo de las aplicaciones de IA

El panorama del riesgo evoluciona a la misma velocidad que las capacidades de la IA. Lo que antes era una sencilla aplicación de SaaS ahora puede introducir silenciosamente funciones integradas de IA, como la generación de copias, respuestas inteligentes y copilotos de IA, sin avisar de ello a los usuarios ni a los equipos de seguridad. Esta tendencia en crecimiento hace que sea cada vez más difícil comprender qué aplicaciones están usando IA, cómo la están utilizando y qué riesgos introducen en la organización.

Los equipos de seguridad necesitan poder evaluar dinámicamente el riesgo en función de cómo se integren las funciones de IA, si conservan o se entrenan con datos de la empresa y cómo se ajustan a los requisitos de cumplimiento. Sin este nivel de información, las organizaciones corren el riesgo de quedar expuestas a fugas de datos, robo de propiedad intelectual, infracciones de la normativa e incluso manipulación del modelo de IA. Puesto que la presencia de la IA en SaaS sigue aumentando, comprender el riesgo de las aplicaciones no es tan solo una buena práctica, sino que es una necesidad para cualquier organización que desee adoptar la IA de forma segura.

Cómo resuelve esto Netskope

Netskope aborda la evolución de la complejidad del riesgo de las aplicaciones de IA con su Índice de Confianza en la Nube (CCI), que ofrece información en tiempo real y continuamente actualizada sobre más de 85 000 aplicaciones en la nube y SaaS. Gracias a sus evaluaciones dinámicas del riesgo que tienen en cuenta la IA, el CCI ayuda a los equipos de seguridad a adelantarse a los posibles riesgos y garantizar el cumplimiento.

Entre sus principales funciones se incluyen:

- **Puntuación del riesgo en tiempo real, teniendo en cuenta la IA:** Identifique aplicaciones con capacidades de IA integradas y comprenda los riesgos asociados con estas funciones.
- **Información sobre cómo se gestionan los datos de empresa:** Evalúe cómo las aplicaciones gestionan los datos de empresa, incluida su conservación, el entrenamiento del modelo y la distribución a terceros.
- **Seguimiento del cumplimiento:** Adáptese a los requisitos reglamentarios, como los del RGPD, SOC 2 e ISO 27001.
- **LLM y MCP seguros:** Evalúe más de 85 000 aplicaciones SaaS, incluidas las aplicaciones de IA y las funciones de IA integradas, así como los servidores MCP públicos, identificando los atributos de riesgo, los tipos de autenticación y las versiones de protocolo.

Con el CCI, los equipos de seguridad pueden enfrentarse con confianza a las complejidades de los riesgos de las aplicaciones de IA y garantizar que su organización siga estando segura y cumpliendo la normativa.

Desafío 3: Integridad del modelo de IA

Como las organizaciones aprovechan cada vez más las herramientas de IA generativa (tanto en modelos creados a medida como en aplicaciones de empresa como Microsoft Copilot), asegurar la integridad de los datos utilizados para entrenar estos modelos se convierte en una cuestión fundamental. Estos sistemas de IA suelen entrenarse con enormes conjuntos de datos que pueden incluir documentos corporativos, correos electrónicos, presentaciones, hojas de cálculo e información comercial privada de carácter confidencial.

Si los datos privados o confidenciales se incorporan involuntariamente a los conjuntos de datos de entrenamiento, esto puede dar lugar a una exposición no solo a través de las salidas del modelo, sino también mediante consultas adversas, fuga de datos y posibles infracciones del cumplimiento. Como la adopción de la IAGen se extiende a varios departamentos, a los equipos de seguridad les resulta cada vez más difícil controlar cómo se obtienen, validan y protegen los datos de entrenamiento.

+ Microsoft Copilot puede entrenarse con el contenido del paquete Office de un usuario, lo que puede incluir desde documentos de Word a hojas de cálculo de Excel. Si los datos confidenciales o sensibles se guardan en estos lugares, y los controles de acceso no están configurados adecuadamente, entonces existe un posible riesgo de que Copilot saque a la luz estrategias comerciales confidenciales, datos financieros o información de los clientes en sus respuestas.

Cómo resuelve esto Netskope

Netskope One DSPM (Gestión de la Posición de Seguridad de Datos) capacita a las organizaciones para que puedan supervisar y proteger los datos confidenciales a través de entornos de nube y repositorios de datos. Al detectar y clasificar datos críticos, como informes financieros, información personalmente identificable y propiedad intelectual, Netskope garantiza que esta información no se utilice para entrenar modelos de IA sin la debida autorización.

Entre sus principales funciones se incluyen:

- **Supervisión continua de los entornos de nube:** Detecte y clasifique datos confidenciales en tiempo real, asegurándose de que no se produzca ningún uso no autorizado durante el entrenamiento del modelo de IA.
- **Visibilidad del acceso y la distribución de los datos:** Obtenga información en tiempo real sobre cómo se accede a los datos y cómo se comparten estos en la nube, lo que permite aplicar medidas correctivas inmediatas cuando sea necesario.
- **Cumplimiento y prevención de las fugas de datos:** Proteja los datos confidenciales para garantizar el cumplimiento, evitar fugas de datos y mantener el control sobre la propiedad intelectual.
- **Gestión sólida de la posición de seguridad:** Garantice una gestión adecuada de los datos e identifique, etiquete y clasifique los datos estructurados y no estructurados.

Con Netskope One DSPM, las organizaciones pueden proteger de forma proactiva sus datos confidenciales, garantizando que el entrenamiento de su modelo de AI se realice de forma segura, esté controlado y cumpla la normativa.

Desafío 4:
Amenazas dirigidas a los sistemas de IA

Los adversarios están haciendo evolucionar sus tácticas para explotar vulnerabilidades específicas de la IA usando inyecciones de consultas, envenenamiento de datos y entradas adversas, todo ello diseñado para distorsionar los resultados o filtrar datos confidenciales. Además, las aplicaciones de IA suelen integrarse en sistemas comerciales más amplios, convirtiéndolos en posibles puntos de entrada para movimiento lateral, escalada de privilegios o robo de datos.

Tanto si un actor de amenaza trata de manipular los resultados de un modelo de IA, como de extraer datos de entrenamiento o explotar la debilidad de los controles de acceso alrededor de las API de IA, la superficie de ataque se está ampliando rápidamente. El problema se complica por la ausencia de marcos de seguridad estándar para proteger los sistemas de IA, lo que deja a muchas organizaciones desprevenidas a la hora de defenderse frente a nuevos vectores de ataque. Con el aumento de la adopción de la IA existe una mayor necesidad de que los equipos de seguridad detecten y disminuyan de forma proactiva las amenazas dirigidas específicamente a los entornos de IA, antes de que esas amenazas pongan en peligro los datos, las operaciones o los procesos de toma de decisiones de carácter confidencial.

Cómo resuelve esto Netskope

Netskope aborda el aumento de las amenazas dirigidas a los sistemas de IA con un enfoque de seguridad de múltiples capas que integra una protección avanzada frente a amenazas, una profunda visibilidad y defensas específicas para la IA.

Entre sus principales funciones se incluyen:

- **Defensa unificada de la IA:** Netskope One AI Guardrails reduce los ataques sofisticados, incluidos los intentos de inyección de *prompts* y de fuga de datos, mediante un análisis exhaustivo y en tiempo real de todo el tráfico.
- **Protección avanzada frente a amenazas:** Utilice aprendizaje automático, espacios seguros y análisis heurísticos para detectar y bloquear amenazas tanto conocidas como de día cero, que incluyen *malware* oculto en archivos enviados a las herramientas de IA.
- **Equipo rojo y evaluaciones de vulnerabilidad:** Automatice las simulaciones de ataques para detectar vulnerabilidades y garantizar que sus modelos privados sean seguros, cumplan con la normativa y sean resistentes frente a amenazas avanzadas con el equipo rojo de Netskope One.
- **Supervisión proactiva de la actividad de IA:** Detecte las amenazas y vulnerabilidades emergentes con supervisión en tiempo real de las interacciones de IA para garantizar una exhaustiva estrategia de defensa.

AI combinar estas tecnologías, Netskope ofrece una solución integrada que ayuda a las organizaciones a asegurar sus sistemas de IA frente a las sofisticadas ciberamenazas y la evolución de los vectores de ataque.

Desafío 5: Exposición de los datos

Uno de los desafíos más urgentes y peligrosos en la seguridad de la IA es el riesgo de exposición de los datos. Mientras los empleados de todos los departamentos adoptan las herramientas de IA para aumentar la productividad, puede que suban o compartan sin darse cuenta datos confidenciales como código fuente, registros de clientes, documentos financieros o PI patentada con los modelos de la IA pública. Una vez expuestos, estos datos pueden conservarse, utilizarse para entrenamiento de modelos o incluso filtrarse, según las políticas de privacidad de las aplicaciones y las prácticas de manipulación de los datos.

Al contrario de los canales tradicionales de distribución de datos, las plataformas de IA pueden actuar como cajas negras, con poca transparencia sobre cómo se guardan, se consultan o se utilizan los datos. Si se carece de protecciones, las organizaciones se enfrentan a riesgos graves que van desde infracciones reglamentarias a robo de PI, pasando por daños a la reputación y desventaja competitiva.

+ Netskope Threat Labs observó la exposición de código fuente en casi el 50 % de las infracciones de políticas relacionadas con la IA. Esto subraya la facilidad con la que los activos críticos para una empresa pueden verse comprometidos mediante unas acciones aparentemente inofensivas, como pegar un fragmento de código en un *bot* de chat de IA para depurarlo u optimizarlo.

Cómo resuelve esto Netskope

Netskope proporciona una protección exhaustiva y rica en contexto para los datos de la empresa, tanto en reposo como en tránsito. Al combinar evaluaciones del riesgo en tiempo real, controles en línea y basados en API, así como comprobaciones de la posición, las políticas de seguridad unificadas de Netskope permiten una gobernanza precisa de las interacciones, tanto de los usuarios como de los datos en toda la organización.

Entre sus principales funciones se incluyen:

- **Prevención avanzada de la pérdida de datos (DLP):** Proteja la información confidencial frente a la filtración mediante herramientas de IA, tanto si los usuarios se encuentran en la oficina, como en casa o fuera de ella.
- **Control pormenorizado:** Bloquee o limite acciones de alto riesgo, como subir código fuente o documentos confidenciales.
- **Formación del usuario en tiempo real:** Eduque a los usuarios sobre las infracciones de las políticas con indicaciones visuales, lo que ayuda a reducir la repetición de las ofensas.
- **Revisión de cada solicitud y respuesta:** Identifique y bloquee la transmisión de datos protegidos por patentes o derechos de autor en las respuestas de la IA para defenderse de forma proactiva contra los riesgos relacionados con la propiedad intelectual asociados a los resultados de los modelos generativos.
- **Seguridad del tráfico API:** Autentique y centralice la gestión del tráfico y la inspección de contenidos entre aplicaciones privadas y LLM.

Con estas capacidades, Netskope asegura una protección exhaustiva y adaptable de los datos que se amplía a toda la IA y el entorno de nube de una organización.

Desafío 6:
Gobernanza, cumplimiento y uso ético

Con la aceleración de la adopción de la IA, las organizaciones se enfrentan a una presión cada vez mayor para adaptarse a las normas emergentes de gobernanza, los requisitos reglamentarios y las expectativas éticas, especialmente en industrias fuertemente reguladas como las de finanzas, servicios de salud y gobierno. Países de todo el mundo están introduciendo rápidamente marcos y órdenes específicas para la IA, como puede verse en la Ley de IA de la UE, el Marco de Gestión de Riesgos de IA del NIST y los decretos sobre seguridad de la IA en Estados Unidos. Estos reglamentos tienen como objetivo garantizar el desarrollo y despliegue responsable de los sistemas de IA, exigiendo transparencia, privacidad de los datos, explicaciones y la ausencia de discriminación.

No obstante, satisfacer estas normas no es nada fácil. Los equipos de seguridad y cumplimiento deben entender cómo se utiliza la IA en todo su entorno, garantizar que los datos confidenciales no se conserven ni se aprenda de ellos de forma indebida, además de demostrar el cumplimiento de las cambiantes directrices legales y éticas.

Cómo resuelve esto Netskope

Netskope garantiza la gobernanza de la IA y la preparación para el cumplimiento a través de una profunda visibilidad, control de las políticas y conocimientos en tiempo real del uso de la IA en toda la empresa.

Entre sus principales funciones se incluyen:

- **Aplicación pormenorizada de las políticas:** Controle cómo se comparten los datos con las herramientas de IA, garantizando que ningún dato confidencial o regulado se utilice para el entrenamiento no autorizado de modelos de terceros.
- **Controles del cumplimiento en tiempo real:** Bloquee las cargas de información de salud protegida a aplicaciones que no cumplan los requisitos o interrumpa el procesamiento de datos financieros en herramientas que carezcan de las certificaciones adecuadas.
- **Respaldo al marco regulador:** Facilite el cumplimiento de marcos como la Ley de IA de la UE y el Marco de Gestión de Riesgos de IA del NIST.
- **Moderación del contenido en tiempo real:** Filtre y controle automáticamente los contenidos peligrosos o discriminatorios, incluidos los discursos de odio, los delitos, las armas y la violencia.
- **Dominio de la gobernanza de los datos de IA:** Garantice la seguridad de todo el ciclo de vida de los datos mediante la detección y clasificación automatizadas, así como el refuerzo proactivo previo a la implementación, para asegurar que su propiedad intelectual permanezca protegida y cumpla con la normativa.

Al combinar visibilidad, inteligencia sobre cumplimiento y la aplicación de políticas adaptables, Netskope permite a las organizaciones adoptar la innovación de la IA de forma responsable a la vez que se cumplen las exigencias éticas y reglamentarias, tanto actuales como futuras.

El futuro de la seguridad de la IA

Tecnologías y amenazas emergentes

Con el crecimiento de la adopción de la IA y la aplicación generalizada de los nuevos casos de uso, desde copilotos a agentes de IA creados a medida, el panorama de las amenazas evoluciona a la misma velocidad. Aunque una gran parte del enfoque de seguridad actualmente se centra en la protección de los datos y la integridad de los modelos, existen dos áreas emergentes de desarrollo tecnológico que están listas para presentar unos retos aún más grandes de cara al futuro.

En primer lugar, están aumentando los sistemas agenciales de IA, capaces de tomar decisiones y aplicar medidas con el mínimo de supervisión humana. Según Gartner, para el año 2028, al menos un 15 % de todas las decisiones de negocios diarias se tomará de forma autónoma por un sistema agencial de IA, desde prácticamente el 0 % actual.[4] Este cambio aumenta drásticamente la superficie de ataque, especialmente si se concede a estos agentes acceso a los sistemas y datos de la empresa a través de MCP o A2A (protocolo de agente a agente).

En segundo lugar, la IA física, como puede verse en vehículos y robots autónomos, está acelerándose en industrias como la logística, el transporte y la manufactura. Estos sistemas introducen riesgos de seguridad reales, ya que una IA vulnerada o que funcione mal no solo causa pérdidas de datos, sino que también podría causar daños a personas e infraestructuras.

Como las capacidades de la IA aumentan y son cada vez más avanzadas y están profundamente integradas en las operaciones comerciales diarias, los líderes de la seguridad deben establecer una gobernanza estratégica y con visión de futuro.

Aquí tiene algunas consideraciones para mantenerse a la vanguardia:

- **Visibilidad del uso de la IA:** Sepa qué equipos están creando o usando modelos de IA, tanto en la TI abierta como en la sombra. Garantice una visibilidad central y supervisión sin imponer límites a la innovación.
- **Fiabilidad de los datos:** Asegúrese de que los modelos se entrenan con conjuntos de datos seguros, de alta integridad y que cumplen las normativas. Unos datos deficientes o contaminados darán unos resultados incorrectos, sesgados o con fugas.
- **Límites de la autonomía y el riesgo:** Con el aumento de la capacidad de los sistemas agenciales de IA, se deben definir unas protecciones para la autonomía. No espere a que los agentes comiencen a tomar decisiones de gran impacto antes de aplicar la gobernanza.
- **Gestión del ciclo de vida del modelo:** Trate los modelos de IA como el código: con control de versiones, detección de vulnerabilidades, controles de acceso y registros de auditoría.
- **Preparación cultural:** La seguridad no es solo un aspecto técnico, también es conductual. Eduque a los empleados y ejecutivos sobre los riesgos de la IA, su uso seguro y el cambiante entorno reglamentario.

El futuro de la seguridad de la IA se definirá no solo por lo bien que las organizaciones se protejan ante las amenazas actuales, sino también por lo cuidadosamente que se preparen para lo que ocurrirá a continuación.

[4] Gartner 2024 https://www.gartner.com/en/newsroom/press-releases/2024-10-21-gartner-identifies-the-top-10-strategic-technology-trends-for-2025

Predicción

La compañía de análisis Gartner prevé que, para el año 2028, al menos un 15 % de todas las decisiones de negocios diarias se tomará de forma autónoma por un sistema agencial de IA, en comparación con prácticamente el 0 % en 2024.

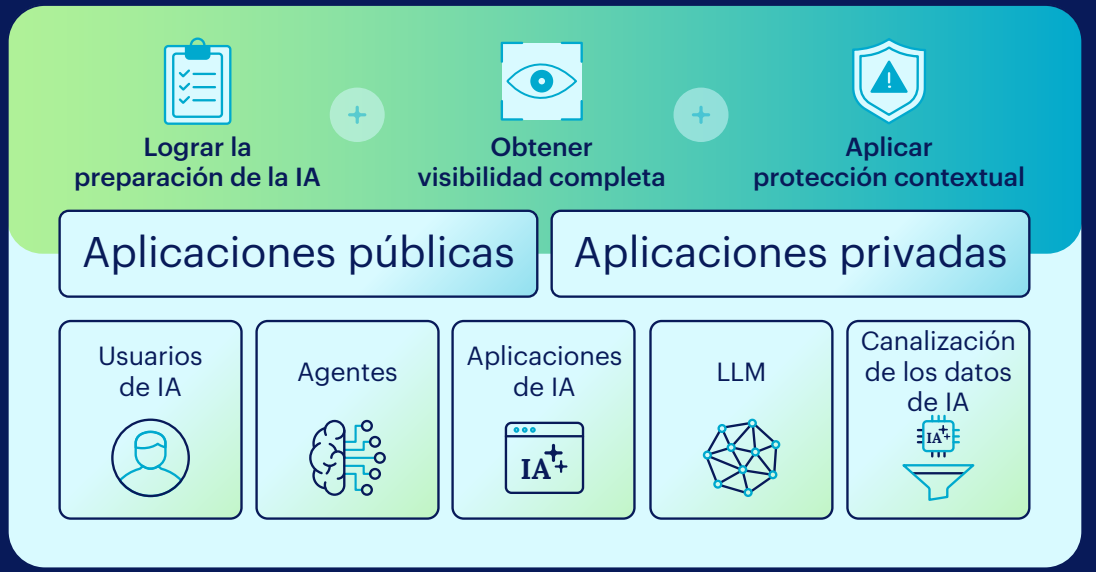
Conclusión

Asegure la IA de un extremo a otro y en cualquier lugar con Netskope One

Mientras las empresas compiten por adoptar rápidamente la IA, los líderes de la seguridad se enfrentan a una presión cada vez mayor para proteger los datos confidenciales y mantenerse por delante de los nuevos riesgos dirigidos a su ecosistema de IA. Desde la falta de visibilidad sobre el uso de la IA, pasando por la exposición de los datos y las necesidades de cumplimiento, hemos explicado seis desafíos importantes que los equipos de seguridad tendrán que superar para permitir el uso seguro de la IA en toda la empresa:

- Falta de visibilidad
- Comprender el riesgo de las aplicaciones de IA
- Integridad del modelo de IA
- Amenazas dirigidas a los sistemas de IA
- Exposición de los datos
- Gobernanza, cumplimiento y uso ético

Netskope One AI Security ofrece una solución única para gestionar su ecosistema de IA y proteger sus datos. Protege a los usuarios y a los agentes automatizados en aplicaciones SaaS públicas, herramientas de IA privadas y flujos de trabajo agénticos. Al combinar un alto rendimiento con controles de confianza cero adaptados al contexto, Netskope permite a las organizaciones aprovechar las ventajas de la IA de forma segura.



Investigación

La compañía de análisis Forrester confirmó que Netskope ofrece una reducción del 80 % en el riesgo de vulneración grave de datos por ataque externo, equivalente a un ahorro de 2 millones de dólares en costes anualizados por infracción material.[5]

[5] Forrester Report: The Total Economic Impact™ of Netskope SSE
https://www.netskope.com/resources/analyst-reports/forrester-the-total-economic-impact-of-netskope-sse

Acerca de Netskope

Netskope (NASDAQ: NTSK), líder en seguridad y redes modernas en la era de la nube y la IA, atiende las necesidades tanto de los equipos de seguridad como de redes proporcionando acceso optimizado y seguridad basada en contexto en tiempo real para el ecosistema de la IA, que incluye agentes, aplicaciones, herramientas, LLM, personas, dispositivos y datos. Miles de clientes, incluidas más de 30 empresas de Fortune 100, confían en la plataforma Netskope One, su motor Zero Trust y su potente red NewEdge para reducir riesgos y obtener una visión completa de cualquier actividad en la nube, la IA, la web y las aplicaciones privadas, ofreciendo siempre seguridad y acelerando el rendimiento sin renunciar a nada.

Recursos

Cómo asegurar la IA con Netskope One

Blog: Domine la adopción de la IA con seguridad de extremo a extremo, en cualquier lugar

Netskope Threat Labs: Informe sobre amenazas de la IA generativa en la nube

Cómo asegurar la IA generativa Para Dummies

¿Le gustaría obtener más información?

Solicite una demostración

netskope