netskope

IA segura: 5 conversaciones cruciales para los CISO

Contenido

Introducción: Un doble mandato para la IA

La gran importancia que tiene la tecnología en las organizaciones actuales ha situado al departamento de TI en el centro de atención como nunca antes. Como consecuencia, los responsables de TI se ven inmersos en una serie de conversaciones cruciales con los CEO, el consejo de administración y otros altos directivos, en un intento por determinar cómo la tecnología puede contribuir mejor al éxito de la empresa; y ningún tema ocupa un lugar más destacado en esos debates que la IA.

La IA plantea un dilema especial a los CIO, los CISO y sus equipos. En el informe «Crucial Conversations» (Conversaciones cruciales) de Netskope [1], revelamos que los CEO están encomendando a sus responsables de TI un doble mandato: integrar la IA para fomentar la experimentación y generar un valor empresarial cuantificable, pero también reducir costes, actuar como guardianes contra el gasto excesivo, evitar las exageraciones y proteger contra posibles fugas de datos o brechas de seguridad.

En resumen, los responsables de TI deben usar la IA para facilitar las innovaciones disruptivas pero, al mismo tiempo, deben defender a la empresa de los riesgos que eso conlleva. Es una dualidad que ejerce mucha presión sobre ellos.

Cada organización se encuentra en una fase diferente de madurez en cuanto a la IA. Algunas siguen identificando casos de uso en los que la IA puede tener un impacto, mientras que otras avanzan a toda máquina, desarrollando sus propias aplicaciones de IA y animando al personal a adoptar de forma generalizada estas herramientas. Todo el mundo se apresura a hacer uso de la IA para acelerar su crecimiento, pero lo hacen a ritmos distintos y partiendo de puntos de salida diferentes.

[1] https://www.netskope.com/crucial-conversations

Cinco pasos para la adopción de la IA

Independientemente del punto en el que se encuentre su empresa en la curva de madurez de la IA, hay aspectos de seguridad fundamentales que deben tenerse en cuenta. La clave está en planificar la estrategia de seguridad sabiendo plenamente cuáles son los riesgos de cada etapa.

1. **¿Podemos experimentar con las herramientas de IA** mientras gestionamos los riesgos de la IA en la sombra?
2. **¿Podemos hacer uso de la IA integrada en las plataformas SaaS** sin permitir un intercambio de datos no autorizado?
3. **¿Cómo gestionamos las aplicaciones de IA** independientes y evitamos la fuga de datos?
4. ¿Cómo podemos evitar respuestas de modelo perjudiciales o sesgadas y las vulnerabilidades habituales de las aplicaciones al **desarrollar aplicaciones de IA privadas**?
5. ¿Cómo podemos evitar conceder un acceso demasiado permisivo al **implementar agentes autónomos**?

Experimentación

IA integrada en plataformas SaaS

Aplicaciones de IA independientes autorizadas

Aplicaciones de IA privadas

Agentes autónomos

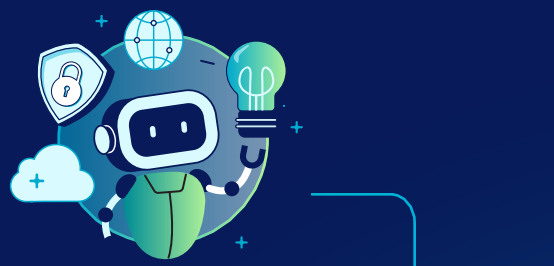
Mantener conversaciones más productivas con los altos directivos

La IA no solo es el tema principal de conversación en los círculos tecnológicos hoy en día, sino que también es un tema prioritario entre los directivos y en las salas de juntas. Según nuestra investigación con consejeros delegados[1], sabemos que se muestran entusiasmados con el potencial de la IA y desean que sus responsables de TI elaboren una estrategia para su adopción e integración cuando sea pertinente, sin dejarse llevar por las exageraciones que se observan en el sector.

El reto para los profesionales de TI, especialmente en el ámbito de la seguridad, consiste en implementar la IA de tal forma que se eviten las concesiones entre rendimiento y seguridad, y que se mantenga el cumplimiento normativo a la vez que se reducen los costes y la complejidad. Del mismo modo, a medida que los planes de implementación se van concretando, es fundamental que los beneficios y los riesgos de las empresas sigan estando a la vista.

Con este libro electrónico, esperamos contribuir a la consecución de estos objetivos. ***IA segura: 5 conversaciones cruciales para los CISO*** tiene como objetivo ayudar a los equipos de seguridad a avanzar con confianza en su proceso de adopción de la IA, manteniendo conversaciones más productivas con sus compañeros sobre los retos y las oportunidades que plantea esta tecnología. El objetivo final es ayudar a las organizaciones a integrar principios preparados para la IA en su estrategia de seguridad, y principios de seguridad en su estrategia de IA, convirtiendo así la seguridad en un motor de crecimiento, en lugar de un obstáculo.

[1] https://www.netskope.com/crucial-conversations



Innovar con la IA y protegerse contra los riesgos: tres principios que hay que priorizar.

Independientemente de en qué punto del proceso de adopción de la IA se encuentre, hay tres principios que se deben tener en cuenta para aprovechar de forma segura el potencial de esta tecnología.

1. **Visibilidad.** Los equipos de seguridad necesitan tener una visibilidad total de su entorno de IA para saber qué herramientas de IA se están utilizando y cómo.

2. **Protección.** Los profesionales deben aplicar medidas de protección contextual para que una seguridad dinámica y adaptable proteja la empresa sin frenar la innovación.

3. **Preparación.** Al analizar de forma proactiva sus datos y aplicaciones, los profesionales de la seguridad pueden alcanzar un nivel de preparación para la IA que deje a su organización lista para el éxito.

Paso 1: Experimentación

¿Podemos experimentar con las herramientas de IA mientras gestionamos los riesgos de la IA en la sombra?

Cuando se lanzó ChatGPT en noviembre de 2022, causó sensación en todo el mundo y tomó por sorpresa a las empresas. Casi de inmediato, los empleados empezaron a utilizar versiones personales de los *chatbots* con IA para agilizar o resolver sus tareas laborales. Hoy en día, la IA en la sombra continúa siendo un problema para muchas organizaciones: Según las investigaciones de Netskope Threat Labs[1], un sorprendente 72 % de los usuarios empresariales seguirá utilizando cuentas personales para acceder a ChatGPT, Google Gemini y otras aplicaciones populares de IAGen en el trabajo en 2025.

Y este problema es cada vez más complicado. Casi todas las aplicaciones SaaS consolidadas incluyen funcionalidades de IA integradas; los modelos de IA se comunican ahora entre sí directamente; se pueden crear agentes mediante lenguaje natural, por lo que ya no son exclusivos de quienes tienen conocimientos técnicos; y todas estas instancias de IA interactúan con más datos y aplicaciones de los que un ser humano podría hacerlo jamás. Como consecuencia, la IA en la sombra se está expandiendo a un ritmo sin precedentes.

El 72 % de los usuarios de empresa sigue utilizando cuentas personales para acceder a aplicaciones de IAGen en el trabajo.

Netskope, informe Generative AI Cloud and Threat Report 2025

[1] https://www.netskope.com/resources/reports-guides/cloud-and-threat-report-generative-ai-2025

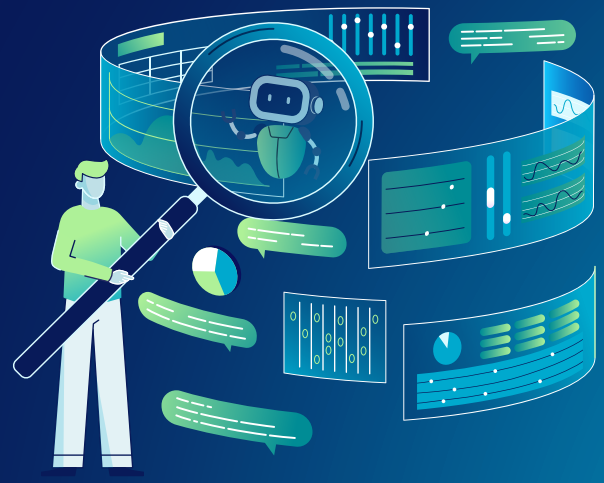
Los equipos de seguridad necesitan urgentemente disponer de una visión amplia y detallada de su entorno de IA. Deben asegurarse de tener una visión global de todas las herramientas de IA que se utilizan en su organización, incluidas las aplicaciones no gestionadas y las instancias personales. También deben profundizar y comprender qué hacen los usuarios y los agentes en el marco de esas interacciones.

Solo ese nivel de conocimientos en profundidad permitirá a los equipos de seguridad ir más allá de la confianza ciega y adquirir un control estratégico sobre las actividades de IA de su organización.

La cruda realidad es que no se puede proteger lo que no se ve.

Cómo lo hacemos

A través del agente de seguridad para el acceso a la nube (CASB) y la pasarela web segura de próxima generación (NG-SWG) de Netskope One, las empresas pueden obtener una visibilidad detallada de las actividades de la IA en su entorno. Nuestro panel de control basado en IA ofrece información detallada sobre qué usuarios acceden a qué aplicaciones, así como sobre las acciones que realizan. Además, ayuda a las empresas a realizar un análisis y una inspección en tiempo real de todas las interacciones públicas con los LLM (datos en movimiento), incluidas las interacciones entre el usuario y la aplicación.



Paso 2: IA integrada en plataformas SaaS

¿Podemos hacer uso de la IA integrada en las plataformas SaaS sin permitir un intercambio de datos no autorizado?

Los LLM y las aplicaciones de IA específicas ya no son los únicos vectores de riesgo desde el punto de vista de la IA. A medida que la tecnología evoluciona, las capacidades de la IA se están incorporando a cada vez más aplicaciones SaaS, desde plataformas de videollamadas hasta herramientas de productividad y sistemas de gestión de ventas.

Estas herramientas SaaS suelen estar profundamente integradas en las empresas modernas, que dependen de ellas para las funciones clave del día a día del negocio, lo que hace que sea casi imposible bloquearlas o eliminarlas. Además, las funciones de IA suelen impulsar la productividad de una forma que ninguna organización querría frenar.

A menudo, las nuevas funciones de la IA se incorporan sin apenas obstáculos; por ejemplo, se incluyen en una actualización general sin proporcionar apenas información sobre los términos y condiciones de uso de los datos. Una aplicación de videollamadas, por ejemplo, podría activar de forma predeterminada una función de toma de notas basada en IA, que grabaría y almacenaría información confidencial de la empresa. Esto puede pillar fácilmente desprevenidos a los equipos de seguridad.

Una aplicación SaaS ya existente podría incorporar nuevas funciones de IA, e incluso activarlas automáticamente, lo que podría pillar desprevenidos a los equipos de seguridad.

Los profesionales de la seguridad deben estar al tanto de su entorno de aplicaciones SaaS, con una visión clara de las funciones de IA, su funcionamiento y los términos y condiciones contractuales relacionados con la gobernanza de datos. Esto debe incluir la comprensión de aspectos como la forma en que cada aplicación utiliza la IA, si usa los datos de la empresa para entrenar sus modelos, si cumple con las normativas clave y si es posible desactivar sus funciones de IA.

Las organizaciones también deben plantearse seriamente la posibilidad de categorizar y clasificar los datos confidenciales para poder aplicar políticas específicas que protejan, por ejemplo, la propiedad intelectual de la empresa o los datos sujetos a regulación, a la vez que se aplican normas menos estrictas en lo que respecta a la información que no sea confidencial.

Cómo lo hacemos

El índice de confianza de la nube (CCI) de Netskope es una base de datos que contiene más de 85 000 aplicaciones SaaS, ofrece un amplio contexto sobre los riesgos y permite a los equipos de seguridad tomar decisiones fundamentadas sobre qué aplicaciones basadas en IA deben permitirse, restringirse o bloquearse.

Paso 3: Aplicaciones de IA independientes gestionadas

¿Cómo gestionamos las aplicaciones de IA independientes y evitamos la fuga de datos?

A estas alturas, muchas organizaciones ya han elegido su herramienta de IA preferida, como ChatGPT de OpenAI, Copilot de Microsoft, Gemini de Google o Claude de Anthropic. La estandarización en torno a un único sistema en toda la empresa ofrece ventajas evidentes en cuanto a capacidades adaptadas a las necesidades empresariales, un aprendizaje reforzado y medidas de seguridad. Y si la organización bloquea además otros sistemas de IA, también reduce la superficie de ataque potencial.

Sin embargo, este enfoque no acaba de eliminar todo el riesgo. Una herramienta de IA corporativa solo puede resultar verdaderamente útil si se conecta con otros documentos y fuentes de información dentro de la empresa. Esto podría permitir que algunos usuarios extraigan datos de documentos internos a los que no deberían tener acceso; en otras palabras, se provoca una fuga de datos dentro de la organización.

Una persona que trabaje en el departamento de marketing podría preguntar a una IA empresarial con permisos excesivos sobre las próximas funciones de la hoja de ruta de productos y recibir información extraída de documentos confidenciales a los que no debería tener acceso.

Cómo lo hacemos

Netskope protege activamente contra amenazas específicas de la IA durante la ejecución. Si un usuario intenta introducir información confidencial, la solución de prevención de pérdida de datos (DLP) de Netskope One interviene de inmediato, impidiendo que los datos de identificación personal, el código fuente o los secretos comerciales lleguen al modelo de IA. Esto también puede hacer que aparezca una ventana emergente con instrucciones para el usuario.

Al mismo tiempo, Netskope One AI Guardrails ofrece moderación de los contenidos en tiempo real para cada interacción. Analiza la intención que subyace a las solicitudes y respuestas para bloquear automáticamente ataques maliciosos y sofisticados, como la inyección de *prompts* y los intentos de *jailbreak*. Además, Guardrails fomenta un uso responsable de la IA filtrando contenidos peligrosos o discriminatorios y bloqueando la distribución de materiales protegidos por derechos de autor. Al combinar estas funciones de DLP y Guardrail, las organizaciones pueden instruir de forma proactiva a los usuarios y, al mismo tiempo, proteger todo el ecosistema de la IA frente a fugas de datos y amenazas emergentes.

Paso 4: Aplicaciones de IA privadas

¿Cómo podemos evitar respuestas de modelo perjudiciales o sesgadas y las vulnerabilidades habituales de las aplicaciones al desarrollar aplicaciones de IA privadas?

Las organizaciones de sectores altamente regulados (como la sanidad, los servicios financieros y la administración pública) están a la vanguardia en el desarrollo de aplicaciones de IA de uso privado. A medida que aumenta su confianza en la IA, muchas empresas están optando por modelos gestionados localmente y entrenados con los datos propios de la organización para reducir los riesgos relacionados con la residencia de los datos, la privacidad, el cumplimiento normativo y la exposición a terceros, a la vez que mejoran la pertinencia y la fiabilidad.

Una tercera parte de las organizaciones ya utiliza los servicios de OpenAI a través de Azure, el 27 % utiliza Amazon Bedrock y el 10 % se basa en Google Vertex AI[1]. Todas estas plataformas de nivel empresarial ofrecen servicios de IA seguros y basados en la nube que proporcionan controles de privacidad más estrictos y opciones de integración más completas que sus versiones públicas.

Sin embargo, el desarrollo de una IA privada también traslada la responsabilidad sobre seguridad a la organización. Más allá de la protección en tiempo de ejecución frente a amenazas específicas de la IA y al uso indebido por parte de los empleados, existe una superficie de ataque adicional en las herramientas utilizadas para diseñar e implementar estos sistemas, que pueden carecer de medidas de seguridad integradas.

Un aspecto importante a tener en cuenta al utilizar un modelo de IA a nivel interno es si este es susceptible de sufrir vulnerabilidades. Si una organización personaliza un modelo de código abierto, por ejemplo, el equipo de seguridad debe seguir probando rigurosamente el código y asegurarse de que no se hayan introducido componentes maliciosos, como un código que pudiera capturar o transmitir *prompts* a una fuente externa.

Otra cuestión a tener en cuenta es asegurarse de que los datos de entrenamiento de la organización no contengan información no deseada. Los equipos deben comprobar si hay algún contenido sesgado, confidencial o perjudicial en estos conjuntos de datos, que suelen ser muy voluminosos.

Cómo lo hacemos

Al centralizar la autenticación, la gestión del tráfico y la inspección de contenidos entre las aplicaciones privadas y los LLM, Netskope One AI Gateway garantiza que los flujos de datos autónomos y agénticos sigan estando controlados y protegidos. Asimismo, el equipo rojo de Netskope One AI somete a pruebas de estrés de forma proactiva los modelos personalizados mediante la automatización de simulaciones adversarias dentro de los flujos de CI/CD para detectar vulnerabilidades como las inyecciones de *prompts*.

Netskope One AI Guardrails reduce los ataques sofisticados, incluidos la inyección de *prompts* y los intentos de *jailbreak*, mediante el análisis en tiempo real de todo el tráfico, a la vez que actúa como moderador de contenidos para identificar y controlar los contenidos peligrosos o discriminatorios, tanto en las interacciones humanas como en las agénticas.

Además, con Netskope One DSPM, los responsables de seguridad pueden obtener visibilidad y control de sus datos, independientemente de donde se encuentren. Esto les ayuda a detectar y clasificar información confidencial, por ejemplo, que podría utilizarse para entrenar un modelo de IA.

[1] *Netskope Threat Labs, informe Netskope Cloud and Threat Report 2026*

Paso 5: Agentes autónomos

¿Cómo podemos evitar conceder un acceso demasiado permisivo al implementar agentes autónomos?

La IA agéntica es la nueva estrella del momento en el mundo de la IA. De hecho, muchos expertos la han señalado como un elemento clave para el futuro de la tecnología empresarial, y la consultora Gartner® prevé que, para 2028, al menos el 15 % de las decisiones empresariales diarias se tomarán de forma autónoma mediante IA agéntica, frente al 0 % en 2024[1].

Aunque la implementación de esta tecnología se encuentra aún en una fase inicial, un estudio realizado por Netskope Threat Labs en agosto de 2025 reveló que ya existe un número considerable de usuarios en diversas organizaciones que están desarrollando agentes de IA o aprovechando las funciones de la IA agéntica en las soluciones SaaS.

Por ejemplo, GitHub Copilot se utiliza actualmente en el 39 % de las organizaciones, y el 5,5 % cuenta con usuarios que ejecutan agentes generados a partir de marcos de trabajo de agentes de IA populares a nivel local. Según los investigadores, el 66 % de las organizaciones cuenta con usuarios que realizan llamadas a la API de api.openai.com y el 13 % a api.anthropic.com[2].

El 39 % de las organizaciones utiliza GitHub Copilot y el 5,5 % ejecuta agentes de IA generados a partir de marcos de trabajo populares a nivel local.

Netskope, informe Cloud and Threat Report: La IA y la IA agéntica en la sombra, 2025

[1] *Comunicado de prensa de Gartner, Gartner identifica las 10 principales tendencias tecnológicas estratégicas para 2025, 21 de octubre de 2024*

[2] *https://www.netskope.com/resources/cloud-and-threat-reports/cloud-and-threat-report-shadow-ai-and-agentic-ai-2025*

En la actualidad, muchas empresas no tienen una idea clara del alcance de su infraestructura de IA agéntica. Dado que este campo evoluciona a un ritmo vertiginoso y que constantemente se añaden nuevas funcionalidades, la IA agéntica en la sombra es una faceta cada vez más importante del problema general de la IA en la sombra.

A medida que crece la adopción de agentes de IA, y se amplía el alcance de sus capacidades en toda la organización, se multiplicarán los riesgos de seguridad que estos conllevan. Será imprescindible que los equipos comprendan las medidas adoptadas por cada agente y establezcan los controles y las políticas adecuados para gestionar los permisos y las actividades que se lleven a cabo.

Las aplicaciones basadas en IA dependen de una comunicación autenticada entre aplicaciones internas, agentes autónomos y LLM alojados de forma privada. Para ello utilizan el protocolo de contexto de modelos (MCP) y las API pero, aunque los protocolos en sí mismos son vías de comunicación seguras, estas interacciones no humanas crean un punto ciego crítico en materia de seguridad. Las API y el MCP permiten a los agentes de IA interactuar directamente con datos y herramientas confidenciales, eludiendo los sistemas de seguridad tradicionales centrados en las personas. Esta brecha conlleva el riesgo de que se produzcan interacciones autónomas sin supervisión, lo que puede dar lugar a fugas de credenciales, la introducción de herramientas maliciosas y la filtración no autorizada de datos.

Cómo lo hacemos

Netskope One AI Gateway funciona como una puerta de enlace definida por *software* para interceptar y gestionar el tráfico de API entre aplicaciones internas, agentes autónomos y LLM alojados de forma privada, garantizando que solo los agentes autenticados puedan comunicarse con los LLM al exigir un token válido generado por la puerta de enlace de IA para cada solicitud.

Netskope One Agentic Broker ofrece una visibilidad unificada y protección en tiempo real para las aplicaciones compatibles con MCP, incluidos los editores de código de IA, las interfaces de *chat* y las herramientas para desarrolladores, mediante la decodificación y la protección del tráfico MCP entre los agentes de IA y las fuentes de datos, tendiendo así un puente entre las interacciones entre humanos y LLM, y los flujos de trabajo de IA de máquina a máquina. Esto garantiza un nivel de seguridad uniforme que protege los datos corporativos confidenciales, a la vez que permite aprovechar la rapidez y la escala de la automatización agéntica.

Conclusión: Gestionar los riesgos de la IA sin tener que renunciar a nada

La IA plantea un reto y una oportunidad que marcan una nueva era para los CIO, los CISO y sus equipos. Los une más estrechamente que nunca a la estrategia empresarial y al crecimiento, lo que amplía su influencia y su impacto. Pero también plantea riesgos importantes que evolucionan con rapidez para los datos, los ingresos y la reputación de las organizaciones, lo que aumenta las consecuencias de cada filtración y ataque informático.

Para saber cómo orientarse ante este panorama, los responsables de TI necesitan un marco claro que les permita comprender las posibilidades disruptivas y los requisitos de protección de la IA, lo que les dará la confianza necesaria para hablar sobre la IA con compañeros que no tengan conocimientos técnicos, con el fin de gestionar los posibles riesgos de la IA y cumplir los rigurosos estándares de cumplimiento normativo.

Para los profesionales actuales de TI y seguridad, la clave está en garantizar la adopción de la IA de principio a fin para impulsar una innovación segura, a la vez que se mantienen unas operaciones empresariales resilientes.

Netskope One es una plataforma única y consolidada que permite gestionar los riesgos de la IA sin comprometer el rendimiento ni la experiencia del usuario, a la vez que reduce la complejidad y garantiza el cumplimiento normativo.

A medida que más empresas avanzan en su proceso de adopción de la IA, desde la experimentación inicial hasta la implementación activa, los responsables de TI pueden desempeñar un papel fundamental a la hora de promover y facilitar la innovación empresarial. Al aprovechar de forma segura las ventajas de la IA, los CIO y los CISO de hoy en día pueden generar un impacto empresarial que impulse a su organización al siguiente nivel.

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 sensibles al contexto, Netskope permite a las organizaciones pasar de la fase experimental de la IA a aprovechar todas las ventajas que esta ofrece.

Para obtener más información sobre lo que los CEO esperan de sus responsables de TI, consulte el informe «Conversaciones cruciales» de Netskope aquí.

Acerca de Netskope

Netskope es líder en seguridad, redes y análisis modernos para la era de la nube y la IA. La arquitectura única de su plataforma Netskope One ofrece seguridad en tiempo real y basada en el contexto a personas, dispositivos y datos, estén donde estén, y optimiza el rendimiento de la red, sin concesiones ni sacrificios. Miles de clientes y socios confían en la plataforma Netskope One, su motor «Zero Trust» patentado y su potente red NewEdge para reducir los riesgos, simplificar la infraestructura convergente y ofrecer una visibilidad y un control totales sobre la actividad en la nube, la IA, SaaS, la web y las aplicaciones privadas.

¿Le gustaría obtener más información?

Solicite una demostración

netskope

04/26 EB-950-2-ES