netskope

Cybersecurity INSIDERS

2026

Informe sobre los riesgos y la disponibilidad de la IA

Descripción general

Para la mayoría de las organizaciones, este es el año en que la IA se convierte en infraestructura. Ahora, los agentes realizan acciones de forma autónoma: modifican registros, crean cuentas y envían código mediante llamadas a las API que se completan antes de que ningún humano las revise. Esto hace que cada implementación de IA suponga un riesgo para la seguridad, independientemente de cómo lo aborden las organizaciones.

Las estructuras de seguridad que se encuentran hoy en día en la mayoría de las organizaciones se diseñaron para un mundo diferente: uno en el que los seres humanos eran los únicos agentes, los procesos eran deterministas, los datos se mantenían en formatos reconocibles y la confianza se verificaba en el navegador. Ese mundo ya no existe.

Este informe se basa en una encuesta exhaustiva realizada a 1253 profesionales de la ciberseguridad, en la que se analizan las medidas que adoptan las organizaciones para proteger la IA, teniendo en cuenta aspectos como la gobernanza, la visibilidad, la protección de los datos y el control de los agentes.

Principales hallazgos:

- **La adopción de la IA se ha adelantado a la gobernanza de seguridad**
  Las herramientas de IA se utilizan actualmente en el 73 % de las organizaciones encuestadas, pero la gobernanza que garantiza la seguridad y el cumplimiento de las políticas en tiempo real solo alcanza al 7 %. Esto crea un déficit estructural de 66 puntos, que se está ampliando a medida que la adopción de la IA se acelera y va más rápido que la implementación de los controles.

- **Los gastos aumentan, pero la confianza disminuye**
  El 90 % ha aumentado este año el presupuesto destinado a la seguridad de la IA, pero el 29 % se siente menos seguro que hace doce meses. El problema avanza más rápido que la inversión.

- **La mayor parte de la actividad de la IA pasa desapercibida para los sistemas de seguridad**
  El 94 % de los encuestados señaló que existe una falta de visibilidad en las actividades relacionadas con la IA. El 88 % no sabe distinguir entre las cuentas personales de IA y las cuentas corporativas. Solo el 6 % afirma tener una visión completa del proceso de la IA en su organización.

- **La IA está dejando obsoletas las medidas de prevención de pérdida de datos heredadas**
  Mientras que los sistemas DLP comparan patrones, la IA transforma el significado; solo el 8 % cuenta con controles que evalúan el contenido de forma semántica, independientemente de cómo se haya reescrito.

- **Los agentes actúan sin controles**
  Los agentes de IA tienen acceso de escritura a herramientas de colaboración (53 %), al correo electrónico (40 %), a repositorios de código (25 %) y a proveedores de identidad (8 %). El 91 % de las organizaciones solo es consciente de lo que ha hecho un agente después de que se haya llevado a cabo la acción.

- **Gran parte de la seguridad de la IA se basa en la confianza**
  El 31 % se basa en políticas por escrito y en el cumplimiento de los empleados para garantizar su aplicación principal. Un 11 % no dispone de nada. Solo el 23 % afirma que aplica medidas de seguridad de IA en tiempo real, en el momento de la acción.

Los riesgos relacionados con la IA están pasando de los errores humanos a la autonomía de las máquinas, pero los controles siguen centrados en hacer frente al primer caso. La encuesta señala cuatro prioridades arquitectónicas: visibilidad continua de toda la actividad de la IA, incluido el tráfico de agentes y M2M; aplicación de políticas en tiempo real sin generar fricciones ni latencia; controles de datos con reconocimiento semántico que evalúen el significado en lugar de los patrones; y la ampliación del modelo de confianza cero a identidades no humanas (NHI). En los capítulos siguientes, se analiza la madurez de la mayoría de las organizaciones y cómo acabar con los problemas en la ejecución.

La gobernanza de la IA va muy por detrás de su adopción

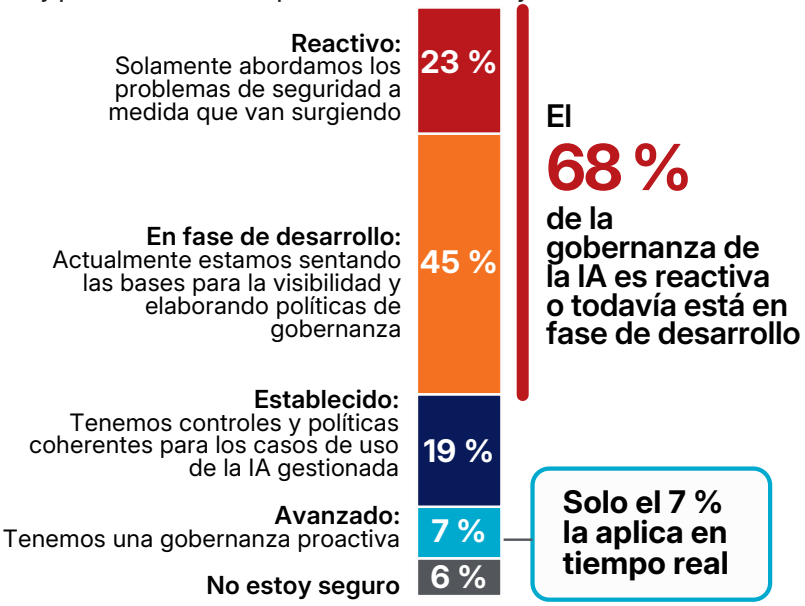
Hace doce meses, para la mayoría de las organizaciones, la gobernanza de la IA era una prioridad futura: algo que había que formalizar después de que se estabilizara su adopción. Esa adopción no esperó. Los copilotos, las herramientas de autocompletado de código y los generadores de contenido llegaron al entorno de producción en todos los departamentos y, para cuando el equipo de seguridad tuvo listo su marco de trabajo, la infraestructura de IA ya estaba en funcionamiento.

Hoy en día, el 68 % de las organizaciones describe su gobernanza de la IA como reactiva o todavía en fase de desarrollo. Solo el 7 % ha alcanzado un nivel avanzado de madurez en la aplicación de políticas en tiempo real. La diferencia de 66 puntos entre el 73 % que implementa herramientas de IA y el 7 % que las gestiona en tiempo real es un desajuste estructural: las organizaciones están avanzando a ritmo de producción sobre unos cimientos de seguridad y cumplimiento normativo que apenas existen. Y las consecuencias ya empiezan a hacerse notar: el 39 % de las organizaciones ya ha estado a punto de sufrir un incidente relacionado con la IA que ha supuesto una exposición de datos involuntaria. De estas, el 17 % no cambió nada después de que esto ocurriera.

Aunque el hecho de que el 68 % de las organizaciones cuente con un sistema de gobernanza reactivo o que aún se encuentre en fase de desarrollo no parezca alarmante, en realidad oculta una tendencia problemática en muchas organizaciones. Más de un tercio de los encuestados señalan una adopción fragmentada de la IA en la que diversos equipos implementan herramientas de forma independiente, sin un marco, normas o políticas de seguridad comunes. Una división gestiona agentes autónomos siguiendo unas directrices informales, mientras que otra ni siquiera ha documentado qué herramientas de IA utilizan los empleados. Las conversaciones sobre la gobernanza no solo van con retraso; en muchas organizaciones ni siquiera han comenzado. El 48 % prevé que los fallos de gobernanza, en concreto, la IA en la sombra y un acceso excesivamente permisivo, serán los responsables de la próxima gran infracción de seguridad relacionada con la IA. Los profesionales más cercanos al problema ya saben que uno de los mayores riesgos procede de las herramientas que sus propios empleados adoptaron el trimestre pasado sin decírselo a nadie. Esa falta de gobernanza se extiende a todas las capas de seguridad posteriores.

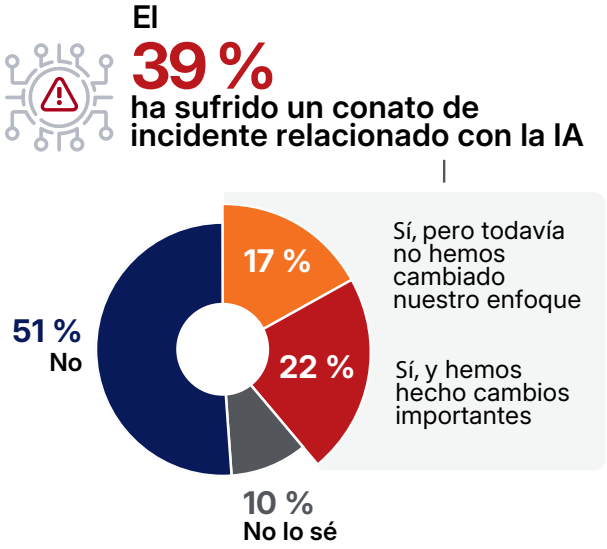
La gobernanza en tiempo real es poco habitual

¿Cuál de las siguientes afirmaciones describe mejor el nivel de madurez de su organización en lo que respecta a la gobernanza y protección de las implementaciones de IA y los datos?



Los conatos de incidentes ya son reales

¿Ha vivido alguna vez un «conato de incidente» relacionado con la IA, como una exposición o filtración involuntaria de datos confidenciales que le haya hecho replantearse su estrategia de seguridad?



La solución estructural más sencilla es identificar los tres casos de uso de la IA que presentan un mayor riesgo en el entorno, integrar políticas aplicables a esos tres casos en los controles técnicos y asignar un responsable a cada uno de ellos.

Más presupuesto, menos confianza

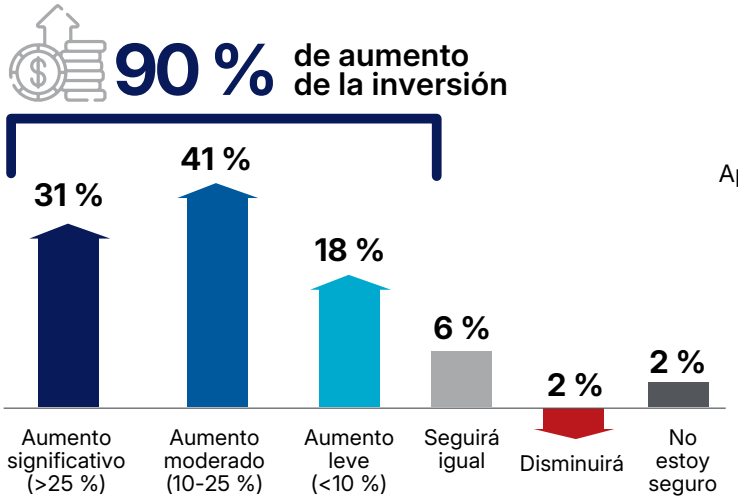
Resulta paradójico que la brecha en la gobernanza de la IA persista a pesar de que las organizaciones invierten ahora más que nunca en seguridad. El 90 % ha aumentado este año su gasto en seguridad en relación con la IA y casi una tercera parte ha incrementado sus presupuestos en más de un 25 %, aún así, el 29 % afirma sentirse menos seguro que hace doce meses. La inversión está aumentando, pero la confianza, no.

Los participantes en el estudio explican por qué: El 34 % considera que el mayor obstáculo es la presión empresarial para adoptar la IA a un ritmo más rápido del que puede llevar la seguridad. Las carencias en materia de competencias ocuparon el segundo lugar, con un 25 %, y las herramientas obsoletas que no pueden interpretar las amenazas específicas de la IA ocuparon un tercer lugar, con un 21 %. Los problemas de presupuesto quedaron en cuarto lugar, con un 14 %. Aunque muchos ya disponen del presupuesto necesario, la arquitectura que este financia sigue reflejando un modelo de amenazas anterior a la IA.

Las herramientas de seguridad existentes se diseñaron para formatos de archivo conocidos, flujos de datos predecibles e interacciones a ritmo humano. Aumentar el presupuesto destinado a eso solo sirve para comprar más cosas que ya son ineficaces frente a los riesgos relacionados con la IA. Los datos sobre la confianza lo confirman: El 51 % considera que sus controles técnicos son deficientes, el 50 % considera que la visibilidad es deficiente y el 61 % considera que la gobernanza de NHI es deficiente.

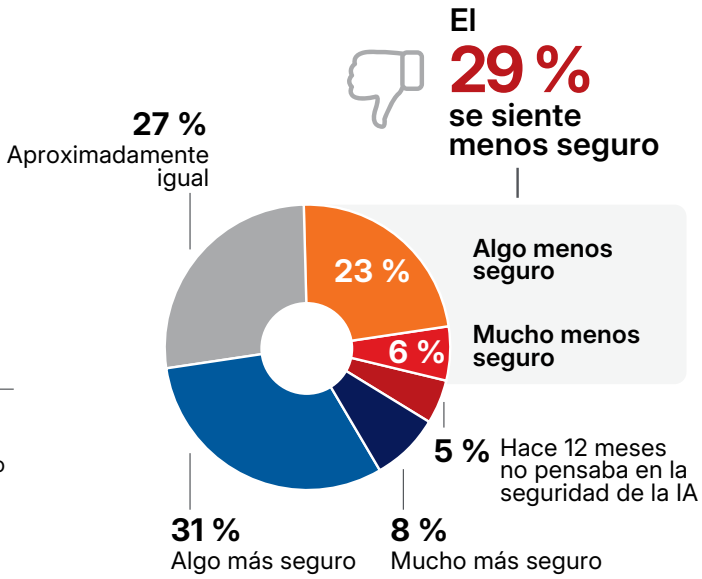
Los gastos aumentan con rapidez

¿Cómo cambiará la inversión de su organización en controles de seguridad específicos para la IA durante los próximos 12 meses?



La confianza decae

En comparación con hace 12 meses, ¿de qué forma ha cambiado la confianza que tiene en la capacidad de su organización para defender los sistemas de IA frente a los ataques y la fuga de datos?



Para reasignar el presupuesto, lo primero que hay que hacer es comparar el gasto actual en seguridad de la IA con esos tres índices de vulnerabilidad (visibilidad, controles técnicos y gobernanza de NHI) y concentrar la inversión en aquellos aspectos en los que la brecha entre el gasto y la capacidad sea mayor.

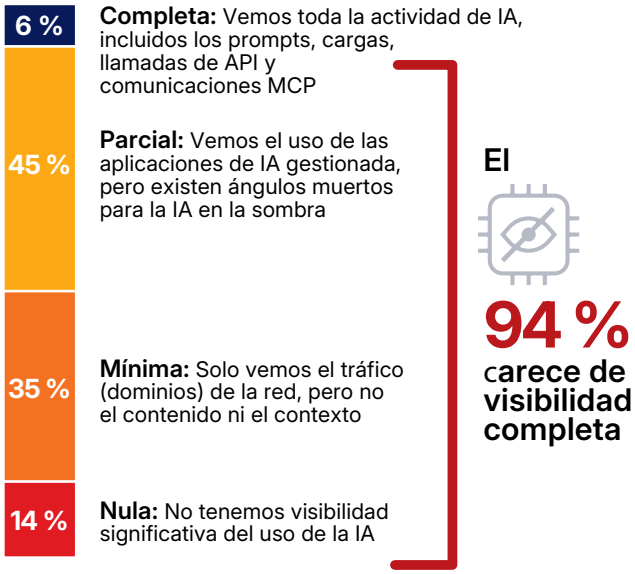
La mayor parte de la actividad de la IA pasa desapercibida para los sistemas de seguridad

Lo que no se ve no puede protegerse. Solo el 6 % de las organizaciones afirma tener una visibilidad completa del uso de la IA en todo su entorno. El 45 % tiene una visibilidad parcial limitada de las aplicaciones gestionadas, sin poder ver nada más allá de las herramientas autorizadas. El 35 % solo ve patrones de tráfico a nivel de red, lo suficiente para saber que algo está pasando, pero no de qué se trata. El 14 % restante no tiene ninguna visibilidad. El 94 % toma decisiones sobre la seguridad de la IA sin disponer de toda la información y, para la mayoría, esta es la forma habitual de operar.

Incluso cuando se tienen métodos de detección, sigue siendo difícil distinguir lo que realmente importa. El 88 % no es capaz de distinguir con certeza entre cuentas personales de IA e instancias corporativas en la misma plataforma, lo que constituye el principal ángulo muerto técnico según la encuesta. Cuando un equipo de seguridad no puede determinar si un empleado está utilizando un entorno de IA autorizado o una cuenta personal sin gobernanza de datos, las políticas de DLP, los controles de acceso y los registros de auditoría pierden toda su fiabilidad. La IA en la sombra añade otra capa más a estos ángulos muertos: el 31 % recurre a un análisis posterior de los registros para identificar herramientas de IA no autorizadas, el 21 % no es capaz de detectar en absoluto la IA en la sombra y solo el 27 % utiliza CASB o SWG para la detección en tiempo real.

En su mayor parte, la visibilidad es incompleta

¿Qué grado de visibilidad tiene su equipo de seguridad sobre el uso de la IA?



La identidad de instancias no está clara

¿Pueden sus herramientas de seguridad distinguir entre las instancias personales y corporativas de las aplicaciones de IA (p. ej., ChatGPT personal frente a ChatGPT empresarial)?

La visibilidad es el requisito estructural del que dependen todos los demás controles. Tanto la DLP como las políticas de acceso y la aplicación de un uso aceptable parten de la base de que la organización pueda ver la actividad que pretende gobernar. Las interacciones más difíciles de controlar son también las que más rápido crecen: las integraciones de API, las conexiones de agentes basadas en MCP y la comunicación M2M están a la cabeza en cuanto a dificultad, mientras que el ***chat*** directo entre el usuario y la IA ocupa el último lugar, con un 6 %.

> Para subsanar esa falta de visibilidad, es necesario ampliar el control de actividad de esos canales, comenzando por establecer una distinción a nivel de cuenta entre las cuentas de IA personales y las corporativas, ya que de ello depende todo lo que vendrá después.

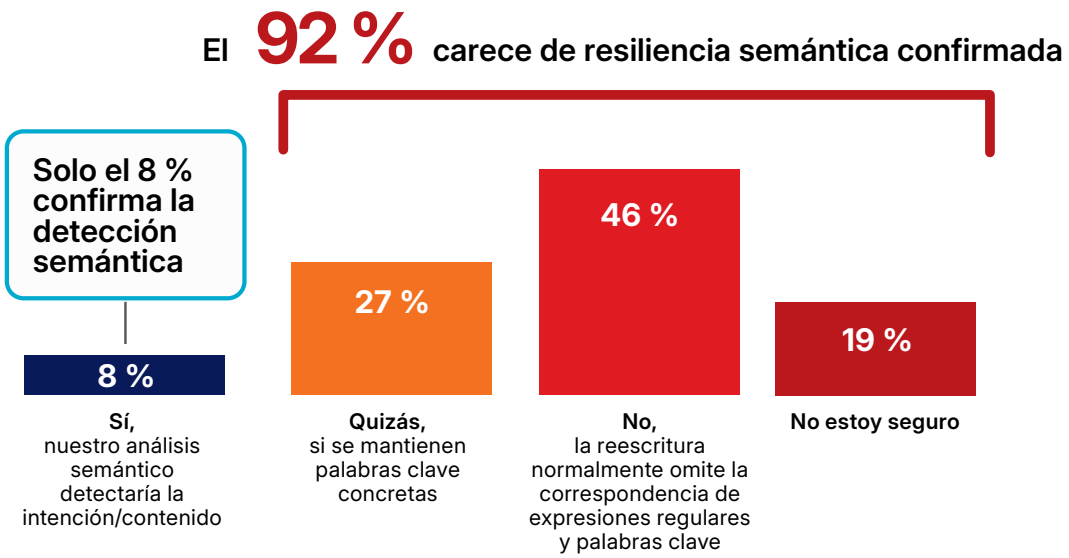
La IA deja indefensa a la DLP tradicional

Incluso cuando las organizaciones pueden ver la actividad de la IA, la herramienta principal encargada de capturar datos en movimiento se diseñó para un tipo de movimiento básicamente distinto. La DLP se creó para encontrar patrones concretos: formatos de tarjetas de crédito, secuencias en los números de la seguridad social y coincidencias de expresiones regulares con contenido confidencial conocido. Aunque la DLP puede bloquear la subida o el copiado-pegado de datos confidenciales en los ***prompts*** al detectar estos patrones, si la IA consigue acceder a los datos, reformulará el contenido confidencial, conservando su significado, y eliminará su huella digital original.

La distinción está en la arquitectura. La DLP opera en la capa sintáctica, haciendo coincidir secuencias de caracteres con reglas predefinidas. La IA opera en la capa semántica, transformando el contenido mientras conserva la intención. Una sencilla prueba de transformación lo deja claro: si un empleado utiliza la descripción de un proyecto secreto y pide a la IA «hazme un resumen de esto en un correo electrónico profesional», la IA podrá sustituir «Proyecto X» por «nuestra iniciativa estratégica futura». Para un filtro de expresiones regulares, el término «iniciativa estratégica» parece totalmente inofensivo, aunque el valor semántico (el secreto) sigue siendo el mismo. Nos encontramos con el mismo tipo de problemas cuando traducimos asuntos confidenciales del inglés a otro idioma y viceversa. La IA genera una versión de los datos en los que cada palabra clave original se ha sustituido, pero el riesgo subyacente permanece, no ha cambiado. La DLP tradicional tampoco sirve en materia de inferencia. Podrá analizar dos listas distintas, una de nombres y otra de enfermedades, y considerarlas inofensivas por separado; pero una IA puede reformular el documento para vincular explícitamente ambas listas, lo que daría lugar a una infracción de la HIPAA que un filtro de DLP basado en patrones no podría detectar. El 46 % afirmó que sus controles no detectarían este tipo de incumplimientos de las políticas, ya que la reescritura suele omitir la coincidencia de expresiones regulares y palabras clave. Otro 27 % dijo que la detección solo funciona si determinadas palabras clave sobreviven a la transformación (un mecanismo de control que solo es eficaz si coopera el adversario). Si a esto añadimos el 19 % que no está seguro de su protección, el 92 % de las organizaciones carece de un sistema de DLP cuya eficacia haya sido confirmada tras la reformulación del contenido por parte de la IA.

La reescritura anula los controles de patrones

La prueba de transformación: Si un empleado le pide a la IA: «reescribe este documento como una entrada de blog genérica», ¿son capaces sus controles de seguridad de detectar los datos confidenciales en la respuesta?



La DLP sigue detectando infracciones basadas en patrones, pero el contenido transformado por la IA pasa desapercibido. El problema se ha trasladado a una capa para cuya inspección la DLP nunca fue diseñada. La forma de medir la exposición es concreta: realice la prueba de transformación anterior con su propio sistema, elija un documento clasificado, pida a una herramienta de IA que lo vuelva a formular y compruebe si sus controles detectan el resultado. Ese resultado se convierte en la base para ejecutar una inspección basada en el contenido que evalúe el significado en el momento de la transferencia.

Agentes de IA sin supervisión

Aunque la fuga de datos a través de herramientas de IA es el riesgo que más conoce la mayoría de las organizaciones, el mayor peligro es que los sistemas de IA actúan ahora por su cuenta: muchos de ellos operan en modo oculto, fuera del alcance de los sistemas de seguridad.

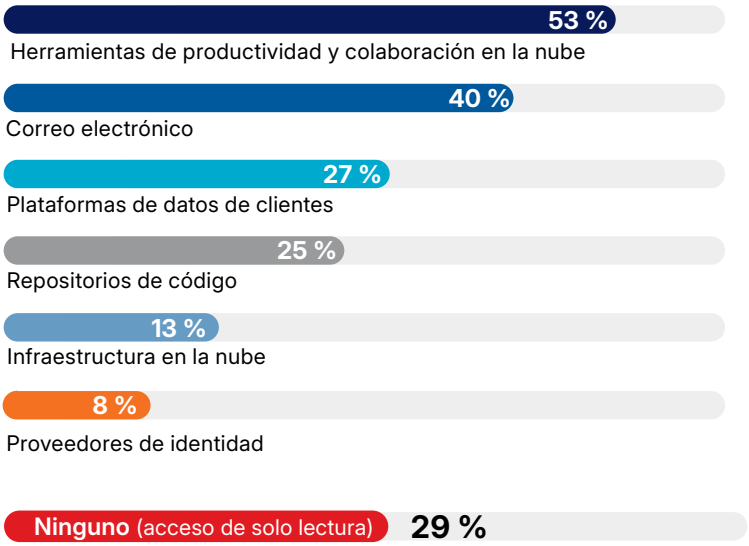
La encuesta cuantifica el alcance de esta situación. El 56 % afirma estar expuesto a los riesgos de la IA agéntica: el 24 % en entornos de producción limitada, el 9 % en entornos a gran escala que gestionan la lógica central del negocio y el 23 % a través de implementaciones en la sombra de las que el departamento de TI no tiene conocimiento. El 32 % no tiene ninguna visibilidad sobre las acciones de los agentes, mientras que el 36 % desconoce en su totalidad el tráfico de IA M2M.

Las organizaciones que no pueden ver a sus agentes no pueden saber que si hay alguno en la sombra. El 10 % afirma haber prohibido el uso de la IA agéntica, pero, en el conjunto de los encuestados, el 23 % admite que la utiliza en la sombra. En la práctica, las prohibiciones a veces promueven la actividad clandestina, lo que dificulta el control y, sobre todo, la contención cuando surge algún problema.

La exposición al acceso de escritura es mayor de lo que se espera la mayoría de los equipos de seguridad. El 53 % concede a las herramientas de IA acceso de escritura a las soluciones de productividad y colaboración en la nube; el 40 %, al correo electrónico; el 25 %, a los repositorios de código; y el 13 %, a la infraestructura en la nube. Luego están los aspectos que modifican la naturaleza del riesgo: el 8 % concede acceso a los proveedores de identidad. Un agente con acceso de escritura a la capa de identidad puede crear cuentas de servicio, aumentar los privilegios en los sistemas federados y otorgarse a sí mismo acceso externo mediante llamadas a la API que nunca pasan por el perímetro de la red.

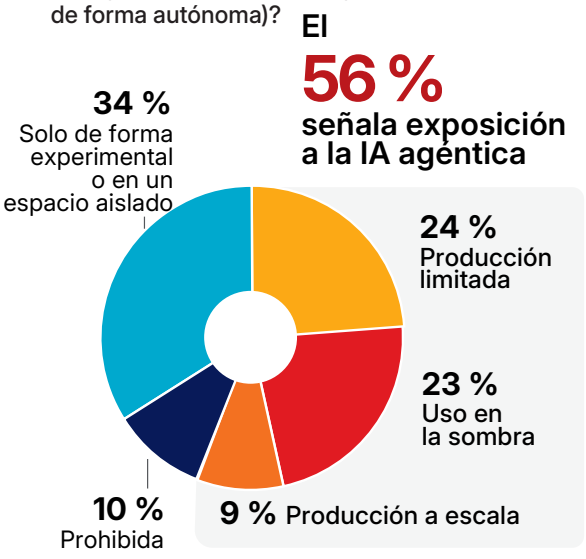
Los agentes ya tienen permiso para escribir

¿A qué sistemas internos tienen acceso de escritura sus herramientas/agentes de IA?



La implementación en la sombra es habitual

¿Cómo describiría su adopción de la «IA agéntica» (IA que persigue objetivos de forma autónoma)?



Solo el 29 % de las organizaciones restringe el acceso de las herramientas de IA a la lectura. En cuanto al 71 % restante, el camino a seguir está claro: hay que auditar qué herramientas de IA tienen actualmente acceso de escritura y establecer controles de autorización para cualquier acción que implique la creación de cuentas, la modificación de permisos o la transferencia de datos al exterior.

Cuando los agentes actúan, nadie los detiene

Los agentes pueden acceder ampliamente a los sistemas empresariales y apenas pueden ser interceptados. Una vez que un agente inicia una acción dañina, solo el 9 % de las organizaciones puede intervenir antes de que la acción finalice. El 91 % restante se desglosa en distintos grados de impotencia: el 24 % puede bloquear algunas acciones del agente, pero no todas; el 35 % solo detecta la acción en los registros una vez que esta ha finalizado, y el 32 % no tiene ningún tipo de visibilidad sobre las acciones del agente. De cada diez organizaciones que utilizan la IA agéntica, menos de una es capaz de impedir que un agente borre un repositorio, modifique un registro de cliente o amplíe privilegios antes de que se ejecute la acción.

Las consecuencias ya se están dejando ver. El 37 % ha sufrido problemas operativos provocados por agentes de IA en los últimos doce meses y, de estos, un 8 % fueron lo suficientemente graves como para provocar interrupciones de servicio o daños a los datos. El 38 % señala como su principal preocupación que un agente transfiera datos de forma autónoma a una ubicación no fiable y el 24 % teme que un agente elimine configuraciones o código críticos. Estas preocupaciones se reflejan en los acontecimientos de 2025-2026 recogidos en informes independientes. A mediados de 2025, la vulnerabilidad EchoLeak (CVE-2025-32711, CVSS 9.3) permitió una inyección de ***prompts*** sin clics por parte del usuario en Microsoft 365 Copilot, lo que facilitó la filtración de datos corporativos sin la interacción del usuario. A comienzos de 2026, los investigadores dieron a conocer el ataque «Reprompt», que combinaba tres técnicas para convertir Copilot Personal en un canal de filtración de datos con un solo clic.

En algún lugar de ese grupo que carece de visibilidad sobre los agentes de IA (32 %) hay un analista de SOC que, al llegar el lunes por la mañana, rastreará un cambio anómalo en los privilegios hasta una cuenta de servicio creada por un agente 72 horas antes y descubrirá que ese agente ha estado comunicándose con los sistemas de producción durante todo el fin de semana. Los registros mostrarán cada acción. No se habrá activado ninguna alerta porque no existía ninguna regla de detección para comportamientos iniciados por el agente.

La prevención es la excepción

¿Puede su organización evitar una acción arriesgada impulsada por la IA (p. ej., que un agente elimine un repositorio) antes de que esta ocurra?

Las consecuencias operativas comienzan a verse

Durante el último año, ¿ha causado alguna herramienta de IA algún problema operativo?

Hay una solución para ese problema: definir qué se considera una anomalía en las acciones de los agentes dentro del entorno; crear reglas de detección para esos patrones y exigir la aprobación de un operador humano para las acciones de alto riesgo de los agentes, como la creación de cuentas, los cambios de permisos y las transferencias externas de datos. La interceptación automatizada en la capa de solicitud es el objetivo que se pretende alcanzar a medida que las herramientas van madurando.

La confianza cero se interrumpe en la máquina

El 91 % de las organizaciones no puede detener a un agente antes de que este actúe. Esto se debe a la arquitectura: la confianza cero se creó en torno a un usuario con un dispositivo, a un lugar, a un patrón de comportamiento y a una puntuación de riesgo. Un agente de IA tiene una credencial, un alcance y una tarea. El 62 % aplica de alguna manera los principios de confianza cero a la seguridad de la IA, por eso es el enfoque más extendido. Sin embargo, el 65 % afirma que sus actuales controles de confianza cero no pueden proteger las identidades no humanas (NHI).

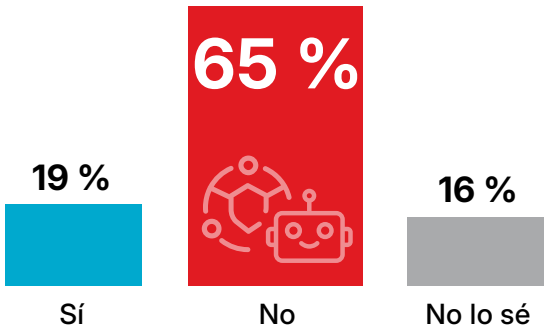
La gobernanza de NHI obtiene la puntuación más baja en todas las dimensiones evaluadas, ya que el 61 % la califica de deficiente; sin embargo, el 78 % espera que el crecimiento de las NHI supere al de las identidades humanas durante el próximo año. Cada nuevo agente, microservicio y flujo de trabajo de automatización crea cuentas de servicio y claves de API para cuya gestión no fue diseñada la gobernanza tradicional de identidades. Estos marcos se han diseñado para entidades que existen más allá de los trimestres fiscales. La identidad de un agente puede durar unos minutos.

Los protocolos que usan los agentes para comunicarse crean una segunda brecha. El MCP ha surgido como un conector común entre los agentes de IA y las herramientas empresariales. En muchas implementaciones actuales, la interoperabilidad ha tenido prioridad sobre la verificación de la identidad integrada, la aplicación del principio del mínimo privilegio o la visibilidad de auditorías independientes. Según la encuesta, solo el 8 % de las organizaciones cuenta con políticas para gobernar el MCP. El 92 % restante no lo controla o ni siquiera sabe que existe. El 36 % no tiene ninguna visibilidad del tráfico de IA M2M y otro 28 % depende completamente de la seguridad de las plataformas de proveedores sin ninguna verificación independiente. Solo el 14 % inspecciona el tráfico de API con el mismo rigor que aplica al tráfico de los usuarios.

La identidad no humana queda fuera del alcance

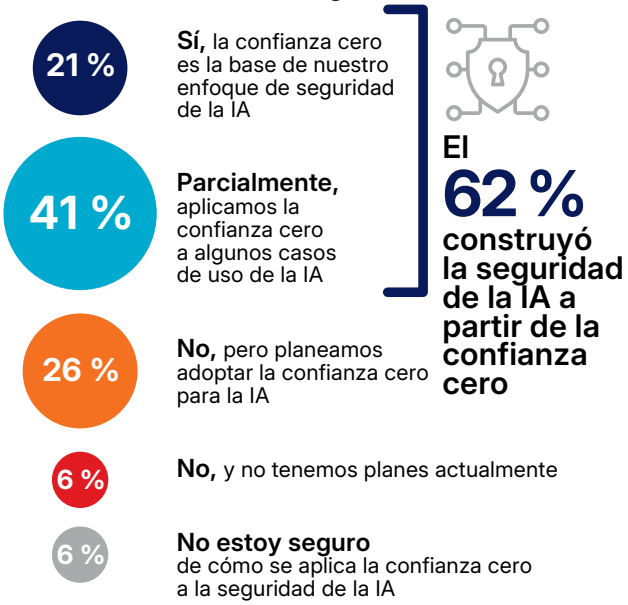
¿Permiten a su organización sus controles de acceso de confianza cero actuales proteger las identidades no humanas?

La mayoría de los programas de confianza cero no abarcan las identidades no humanas



La confianza cero sigue siendo la estrategia

¿Se basa su estrategia de seguridad de la IA en los principios de la confianza cero (verificar cada solicitud, controlar cada flujo de datos y conceder acceso en función del riesgo dinámico)?



Para cerrar esta brecha, es necesario que la organización integre las capas de protocolo e identidad, de modo que las credenciales, los alcances y los permisos de los agentes se evalúen con el mismo rigor que se aplica a las identidades humanas.

La mayor parte de la seguridad de la IA se basa en la confianza

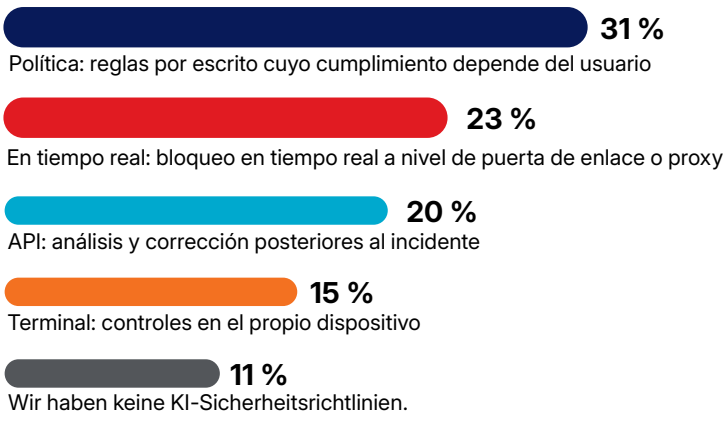
La confianza cero se aplica a las identidades humanas, pero deja a los agentes de IA y a las NHI sin apenas gobernanza. La pregunta que plantea esta brecha es de carácter práctico: cuando una herramienta de IA incumple una política o un agente lleva a cabo una acción arriesgada, ¿qué mecanismo de control interviene realmente? La encuesta abarca todo el espectro de la aplicación de la normativa. Un 31 % garantiza la seguridad de la IA mediante políticas por escrito y su cumplimiento por parte de los empleados. Otro 20 % se basa en el análisis de API posterior al evento, detectando las infracciones una vez finalizada la acción. Los controles basados en terminales representan un 15 %, mientras que la aplicación en tiempo real en línea supone un 23 %. El 11 % restante no cuenta con ninguna política de seguridad relacionada con la IA. La categoría más importante en cuanto a cumplimiento es el sistema de confianza. La siguiente categoría es revisar lo que ya ha ocurrido, después del hecho.

Un análisis más detallado de los datos sugiere que incluso ese 23 % de las organizaciones que afirma aplicar medidas de cumplimiento en tiempo real podría estar utilizando instrumentos poco precisos. El 42 % controla las aplicaciones de IA mediante un sistema binario de bloqueo o permiso para plataformas completas, sin posibilidad de permitir únicamente las cuentas corporativas y bloquear las personales, ni de permitir una consulta de investigación y bloquear la carga de un modelo financiero. Solo el 19 % cuenta con controles granulares a nivel de actividad que diferencian las acciones en una aplicación permitida. Cuando existen controles en tiempo real, estos se centran en las acciones humanas: el bloqueo de la subida de archivos (48 %) y la detección de actividades de pegado (37 %) ocupan los primeros puestos, mientras que la publicación de contenido (29 %) y los controles de descarga (25 %) se quedan rezagados. Las llamadas a las API iniciadas por agentes, los intercambios de tokens OAuth y los flujos de datos M2M pasan prácticamente desapercibidos.

Esta brecha en la ejecución es un fallo de coherencia. Cuando el CASB, el motor de la DLP y la política de acceso solo ven por separado una parte del panorama, ninguno de ellos puede aplicar la estrategia en su totalidad. Una prueba de diagnóstico sencilla: si una sola decisión de política requiere datos procedentes de más de dos consolas, es en esa fragmentación donde falla la aplicación. Cerrar la brecha implica unificar esas capas, de modo que una única evaluación tenga en cuenta la clasificación del contenido, la identidad del usuario y el tipo de instancia de IA antes de que se complete una acción.

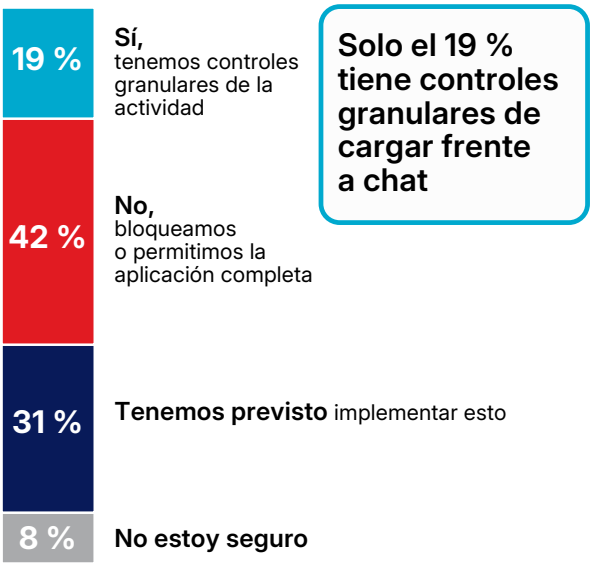
La aplicación de normativas ocurre principalmente tras la acción

¿Cómo se aplican principalmente sus políticas de seguridad de la IA?



Los controles siguen siendo poco precisos

¿Aplican políticas diferentes para «cargar» frente a «chat»?



Solo el 19 % tiene controles granulares de cargar frente a chat

La seguridad de la IA no puede integrarse en los modelos perimetrales tradicionales. Requiere una aplicación en tiempo real y nativa de la nube que evalúe la identidad, el contenido y el contexto en un único punto de decisión antes de que se produzca la ejecución.

¿Qué nivel de madurez tiene su seguridad de IA?

Cada una de las brechas analizadas hasta ahora amplifica las demás: la escasa visibilidad compromete la DLP, los agentes no controlados omiten los controles de acceso y la aplicación fragmentada de las normas deja expuestas todas las capas. El modelo de madurez de capacidades que se muestra a continuación clasifica seis dominios fundamentales de la seguridad de la IA en tres niveles de preparación. Cada celda describe las capacidades en esa etapa. Busque en cada fila la descripción que se ajuste a su organización. El dominio con el nivel de madurez más bajo es su eslabón más débil y el punto más probable de fallo. Ahí es donde debería dirigirse la inversión en primer lugar.

DOMINIOS DE SEGURIDAD DE LA IA	REACTIVO	GESTIONADO	ADAPTATIVO
Alineación de gobernanza y riesgos	Políticas sobre el papel. Aplicación incoherente.	Políticas aplicadas a herramientas de IA gestionadas. IA en la sombra sin gobernanza.	Política integrada en controles técnicos, en tiempo real.
Visibilidad y conciencia situacional	Parcial o sin visibilidad. No puede distinguir entre instancias personales y corporativas.	Supervisión del nivel de actividad en las soluciones SaaS y las API gestionadas. Distinción entre instancias personales y corporativas.	Visibilidad en tiempo real de todos los flujos de trabajo de IA, incluidos entre agentes y máquina-máquina.
Protección de datos y activos	Los controles basados en patrones fallan ante la transformación impulsada por la IA.	DLP ampliada al tráfico de IA, incluidos *chats* y *prompts*. Detección semántica en fase piloto.	Inspecciones semánticas en todos los flujos de datos de IA.
Control de acceso y ejecución	Aplicación tras la ejecución. Sistema basado en la confianza para la aplicación de políticas.	Aplicación en tiempo real para la IA iniciada por humanos. Acciones de agentes registradas, no bloqueadas.	Aplicación dinámica antes de la ejecución, humana y no humana.
Detección y respuesta	Control basado en registros. No hay lógica de detección específica para el comportamiento impulsado por la IA.	Reglas de detección para patrones conocidos de uso indebido de la IA. Contención manual.	Control continuo con contención automática en todas las actividades de IA.
Integración arquitectónica y resiliencia operativa	Fragmentación. Visibilidad, protección y aplicación en silos.	Controles integrados para SaaS gestionado. Brechas en el tráfico de API y agentes.	Marco unificado duradero ante la automatización y la escalabilidad.

Al preguntar de que se arrepentían en relación con la adopción de la IA, el 38 % afirma que hubiera preferido haber establecido un marco de gobernanza antes de implantar la IA a gran escala y el 25 % hubiera preferido haber invertido antes en controles de visibilidad. Solo el 7 % se muestra satisfecho con su enfoque actual (el indicador de confianza más bajo de toda la encuesta).

El cambio viene impulsado por la presión

Un 52 %
dice que la normativa motiva el cambio

Un 47 %
dice que una infracción motiva el cambio

Reducción de la brecha en la ejecución

El modelo de madurez de capacidades muestra dónde están las brechas. A continuación, se resumen las acciones más útiles para mejorar cada uno de los vectores de riesgo, comenzando con la visibilidad, la base sobre la que se apoya el resto de los controles.

1. **Cerrar las brechas de visibilidad de la IA:** el 94 % afirma que hay brechas; el 88 % no puede distinguir las cuentas personales de las corporativas.
   Ampliar el control a nivel de actividad en el tráfico de SaaS, API y M2M, comenzando con la distinción a nivel de cuenta entre las cuentas de IA personales y corporativas: el requisito previo para una DLP fiable, controles de acceso y pistas de auditoría.

2. **Convertir las políticas en medidas de protección aplicables:** el 68 % gobierna de forma reactiva; solo el 7 % lo hace en tiempo real.
   Identificar los tres casos de uso de la IA que suponen el mayor riesgo dentro del entorno, integrar políticas aplicables para ellos en los controles técnicos y asignar un responsable a cada uno antes de ampliar la cobertura al resto de casos de uso de IA.

3. **Implementar la protección de datos semántica:** el 46 % no supera la prueba de transformación de contenido.
   Realizar la prueba de transformación de contenido con el sistema de DLP propio: tomar un documento clasificado, pedir a una herramienta de IA que lo vuelva a formular y comprobar si los controles detectan el resultado. Ese resultado es la base para realizar una inspección basada en el contenido que evalúe el significado en el momento de la transferencia.

4. **Aplicación antes de la ejecución:** el 23 % aplica medidas en tiempo real; el 9 % puede detener con antelación una acción peligrosa de un agente.
   Comprobar qué agentes tienen acceso de escritura en la actualidad y establecer puntos de control de autorización para cualquier acción que cree cuentas, modifique permisos o transfiera datos al exterior.

5. **Modernizar la detección y la contención:** el 67 % se basa en los registros o no tiene visibilidad de las acciones de los agentes; el 37 % ya ha sufrido problemas operativos causados por la IA.
   Definir qué se considera una anomalía en las acciones de los agentes dentro del entorno y crear reglas de detección para esos patrones.
   Establecer procedimientos de contención que permitan intervenir en la capa de solicitud antes de que se complete una acción, en lugar de generar una incidencia cuando el daño ya está hecho.

6. **Reducir la fragmentación de controles:** el 42 % utiliza controles binarios de bloqueo o permiso sin diferenciar según el nivel de actividad; solo el 14 % inspecciona el tráfico de las API con el mismo rigor que se aplica al tráfico de los usuarios.
   Unificar CASB, DLP y las políticas de acceso para que una única evaluación tenga en cuenta la clasificación de contenidos, la identidad del usuario y el tipo de instancia de IA. Si una decisión sobre políticas requiere actualmente datos procedentes de más de dos consolas, es precisamente en esa fragmentación donde falla la aplicación de las políticas.

La seguridad de la IA ya es una disciplina operativa. Se han definido las dimensiones de madurez, la secuencia de dependencias está clara y las acciones son concretas. Lo que queda es la decisión de construir.

La gobernanza está rezagada

Un 68 % está elaborando políticas de gobernanza y seguridad para las implementaciones de IA y los datos, o reaccionando ante ellas

La visibilidad es incompleta

Un 94 % carece de una visibilidad completa del uso de la IA

Los controles de datos no superan la reescritura

Solo el 8 % confirma una detección semántica de intención/contenido

La interceptación apenas tiene lugar

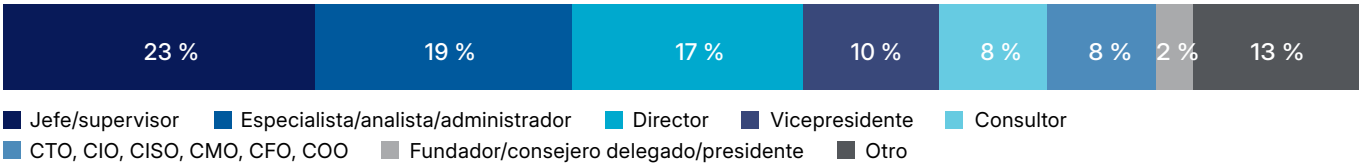
Solo el 9 % cuenta con prevención en tiempo real para las acciones de API y agentes

Metodología y datos demográficos

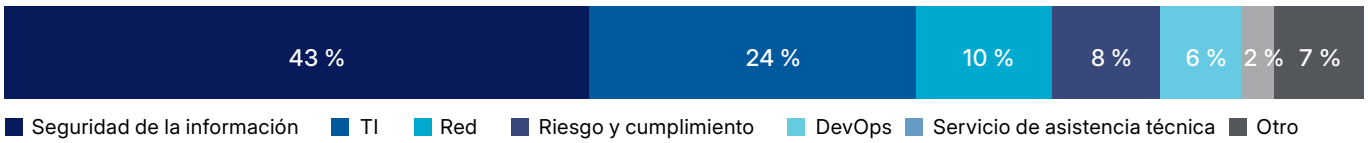
Este informe se basa en una encuesta realizada a principios de 2026 a 1253 profesionales de ciberseguridad y TI. Los encuestados son profesionales de la seguridad, arquitectos y responsables tecnológicos encargados de proteger la infraestructura empresarial, los entornos de la nube y las aplicaciones basadas en IA en una amplia variedad de sectores y organizaciones de distintos tamaños.

El estudio analiza cómo las organizaciones protegen las implementaciones de IA, centrándose en la madurez de la gobernanza, la visibilidad de la actividad de la IA, la protección de los datos, la gestión de identidades no humanas y el control de los agentes autónomos. Mediante un método de muestreo estratificado, la encuesta alcanzó un nivel de confianza del 95 % con un margen de error de +/- 2,8 %.

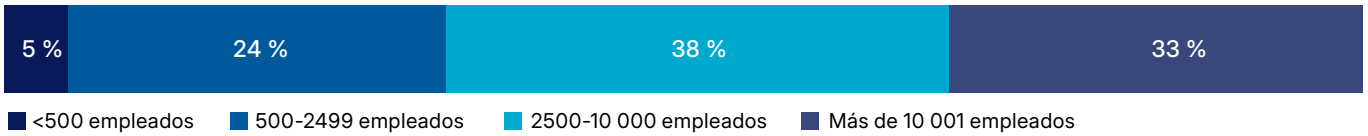
NIVEL PROFESIONAL



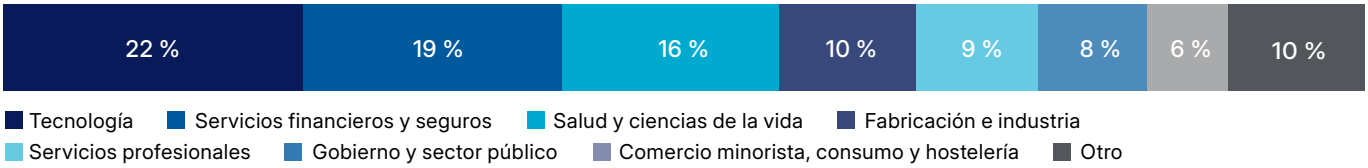
DEPARTAMENTO



TAMAÑO DE LA COMPAÑÍA



SECTOR



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.

Para más información, visite

netskope.com/es/products/ai-security

Cybersecurity
INSIDERS

EVALÚE EL NIVEL DE MADUREZ DE SU SEGURIDAD

Un estudio independiente sobre ciberseguridad que pone de manifiesto las brechas existentes que determinan la estrategia de ciberseguridad

Cybersecurity Insiders elabora estudios independientes basados en encuestas realizadas a líderes y profesionales del sector de la ciberseguridad de todo el mundo. Nuestros informes revelan dónde fallan las estrategias de seguridad en la práctica, lo que ayuda a las organizaciones a evaluar su nivel de madurez, identificar las brechas en sus capacidades y priorizar las acciones necesarias para cerrarlas.

Para obtener más información, visite

cybersecurity-insiders.com

Abril de 2026 // Versión 1.7