ThreatLabz 2026: Informe sobre el grado de preparación para la IA de vanguardia



Tabla de contenido



Resumen ejecutivo 3

En cifras 5

Cohorte de evaluación 6

Perspectiva de amenazas 7

Metodología 9

Hallazgos clave 10

La nueva cadena de ataque:
Cómo la IA de vanguardia encadena de manera autónoma pequeñas brechas en una única secuencia de explotación 15

Pruebas de caso:
Tres organizaciones en progreso 16

Realismo en la implementación:
Lo que realmente requiere el sprint de 90 días 18

Recomendaciones 19

©2026 Zscaler, Inc. Todos los derechos reservados.

Resumen ejecutivo

El año pasado recibimos una llamada de una de las empresas de IA más conocidas. Nos acababan de notificar sobre el lanzamiento de un modelo de IA de vanguardia optimizado para realizar pruebas de seguridad autónomas. Esa llamada nos impulsó a emprender un camino para ayudar a las organizaciones a prepararse, examinando su grado de preparación ante estos ataques autónomos. Nos sorprendieron tanto los resultados como sus capacidades.

Cabe destacar que no aplicamos la IA de vanguardia a los entornos de las organizaciones; sino que, primero lo probamos en nuestra propia infraestructura, utilizando esos conocimientos para desarrollar el programa de evaluación que ahora ofrecemos.

El objetivo no es solo señalar las deficiencias, sino también ayudar a ilustrar dónde se puede reforzar la seguridad a medida que las organizaciones se preparan para una nueva oleada de amenazas. Todos estamos aprendiendo juntos, y el objetivo de estos hallazgos es servir no solo como advertencia, sino como una lección compartida sobre cómo podemos mejorar la resiliencia en todos los ámbitos.

Dado que Zscaler opera con una arquitectura nativa de la nube, a diferencia de los entornos híbridos y heredados que utilizan la mayoría de las empresas, empleamos los mismos métodos de pruebas de vanguardia que se emplean con modelos reales y los replicamos con modelos y tácticas, técnicas y procedimientos (TTP) igualmente eficaces para emular la manera en que estos adversarios operan en los entornos de las organizaciones. Esto nos permitió diseñar y calibrar el programa de evaluación para las realidades operativas de la infraestructura empresarial, en lugar de para arquitecturas como la nuestra.

Nuestro hallazgo principal: la respuesta ganadora a la IA no es simplemente más IA, sino una arquitectura que, desde su diseño, cierre las vías de explotación.

Las empresas globales no están preparadas para los ciberataques impulsados por la IA de vanguardia, pero no por la razón que la mayoría de la gente supone. El peligro no reside en que la IA descubra vulnerabilidades desconocidas y nunca antes vistas. La clave reside en que los modelos de vanguardia pueden tomar las vulnerabilidades más obvias, las brechas que toda organización ya sabe que existen, y encadenarlas en rutas de ataque completas a la velocidad de las máquinas. La mayoría de las arquitecturas nunca fueron diseñadas para eso, e intentar superar a la IA, ya sea con humanos o con herramientas, es una batalla perdida desde el principio. La arquitectura es lo único que gana. La buena noticia es que, con la arquitectura Zero Trust adecuada, la mayoría de las empresas están más cerca de la resiliencia de lo que creen.

El Programa de evaluación del grado de preparación para la IA de vanguardia mide la resiliencia de la organización frente a un tipo específico de adversario: un modelo de IA de vanguardia que opera a la velocidad de las máquinas y es capaz de realizar reconocimiento autónomo, encadenamiento de vulnerabilidades y movimiento lateral. Utilizamos el término “IA de vanguardia” para describir esta clase de adversario porque puede razonar a través de superficies de ataque de maneras que los escáneres tradicionales e incluso los analistas humanos no pueden replicar a una velocidad comparable. En pruebas paralelas realizadas en entornos de producción, los modelos de vanguardia demostraron un razonamiento con una alta relación señal–ruido y un contexto sólido, actuando eficazmente como multiplicadores de fuerza para el descubrimiento de cadenas de ataque. Estas capacidades no son teóricas; ya han sido demostradas.

©2026 Zscaler, Inc. Todos los derechos reservados.

En cientos de organizaciones empresariales evaluadas en el primer semestre de 2026, incluidas empresas de la lista Fortune 500 de los sectores de servicios financieros, sanitarios, fabricación, energía y ocho sectores adicionales, la puntuación media del grado de preparación para la IA de vanguardia fue de 37 sobre 100. Casi todas las organizaciones (91 %) tenían infraestructura de IA accesible públicamente y ninguna de ellas (100 %) contaba con controles de acceso basados en la identidad. Además, ninguna organización inspeccionó suficiente tráfico cifrado como para que su sistema de seguridad funcione según lo previsto. La empresa promedio desconocía prácticamente el 64 % de los datos que circulaban por su propia red, y ninguna había logrado cerrar la brecha entre poseer capacidades de defensa y ponerlas en práctica.

La brecha emergente más urgente agrava el hallazgo estructural anterior: la gobernanza de la IA se encuentra en su punto cero en todas las empresas evaluadas. Las superficies de IA, incluidos los copilotos, las API de inferencia, los puntos finales de MCP y las puertas de enlace de la IA oculta, se están multiplicando más rápido de lo que los programas de seguridad pueden absorberlas y gestionarlas.

Esto representa una oportunidad significativa para mejorar el grado de preparación frente a las amenazas de la IA de vanguardia. Descubrimos que casi un tercio de las organizaciones podrían mejorar drásticamente la protección extendiendo el engaño a puntos críticos de acceso. Asimismo, llevar la prevención de intrusiones al modo de aplicación total, integrar la detección de anomalías en la capa de aplicación y refinar las políticas de segmentación representan pasos concretos y de gran impacto para mejorar la fidelidad de las alertas del SOC y la postura defensiva general.

Los responsables de seguridad deberían leer este informe como un catalizador para el debate a nivel de la junta directiva. Es posible pasar de una media de 37 en la cohorte a una puntuación objetivo del grado de preparación para la IA de vanguardia de entre 73 y 98, el umbral en el que los cinco pilares defensivos alcanzan la eficacia operativa contra los ataques de clase IA de vanguardia, en un plazo de 90 días. Requiere tres acciones ejecutivas: (1) establecer la resiliencia de IA de vanguardia como un mandato empresarial con informes a la junta directiva, (2) asegurar la alineación ejecutiva interna y la capacidad dedicada de sprint, y (3) ejecutar un sprint de doble vía: activar los controles latentes y bloquear cada superficie de IA simultáneamente.

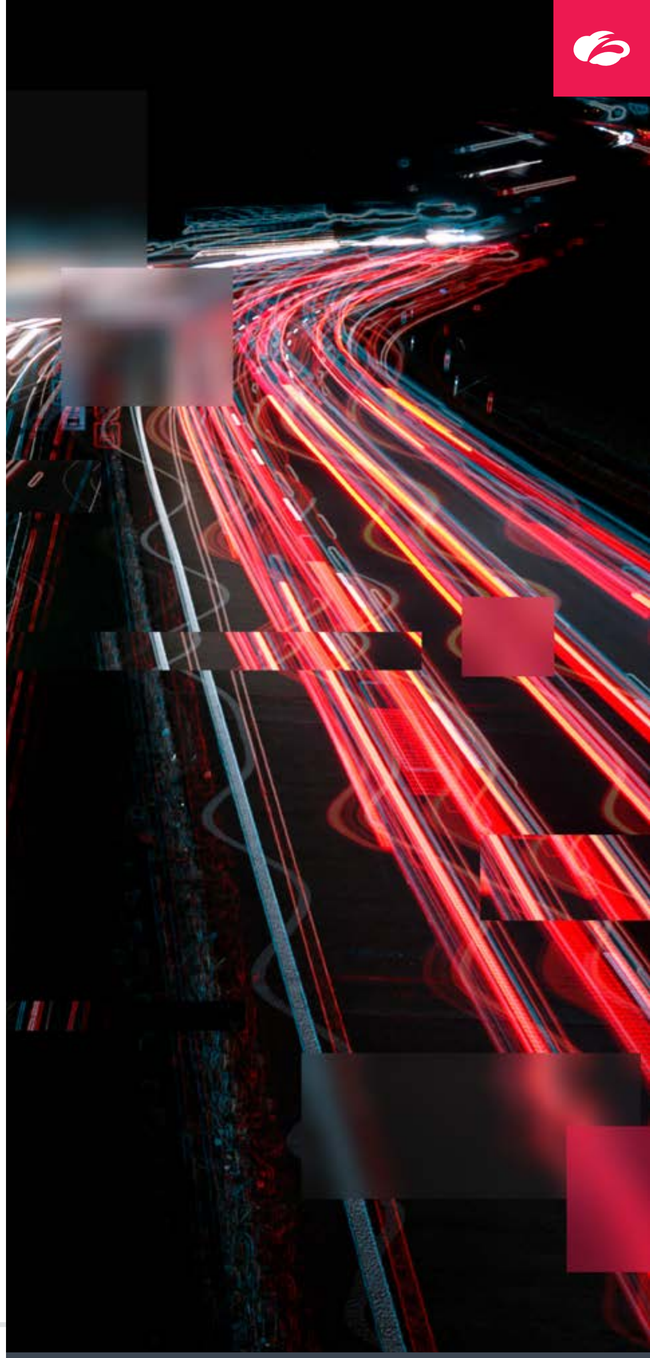
El punto de partida operativo es el mismo para todas las organizaciones: empezar con lo que ya se tiene. El punto de partida estratégico es la gobernanza: ninguna organización puede posponer la seguridad del acceso a la IA sin aceptar una exposición vulnerable a la explotación en la actualidad.



DEEPEN DESAI,
vicepresidente ejecutivo de ciberseguridad de Zscaler



GEORGE MOSER,
vicepresidente global y director de CXO residente de Zscaler



©2026 Zscaler, Inc. Todos los derechos reservados.

Según las cifras

El 100 % de las organizaciones evaluadas carecen de autenticación basada en la identidad para su IA.

El 91 % de las organizaciones con infraestructura de IA accesible públicamente la dejan completamente expuesta a ataques externos.

37 sobre 100. El mínimo era de 17 y el máximo de 60.

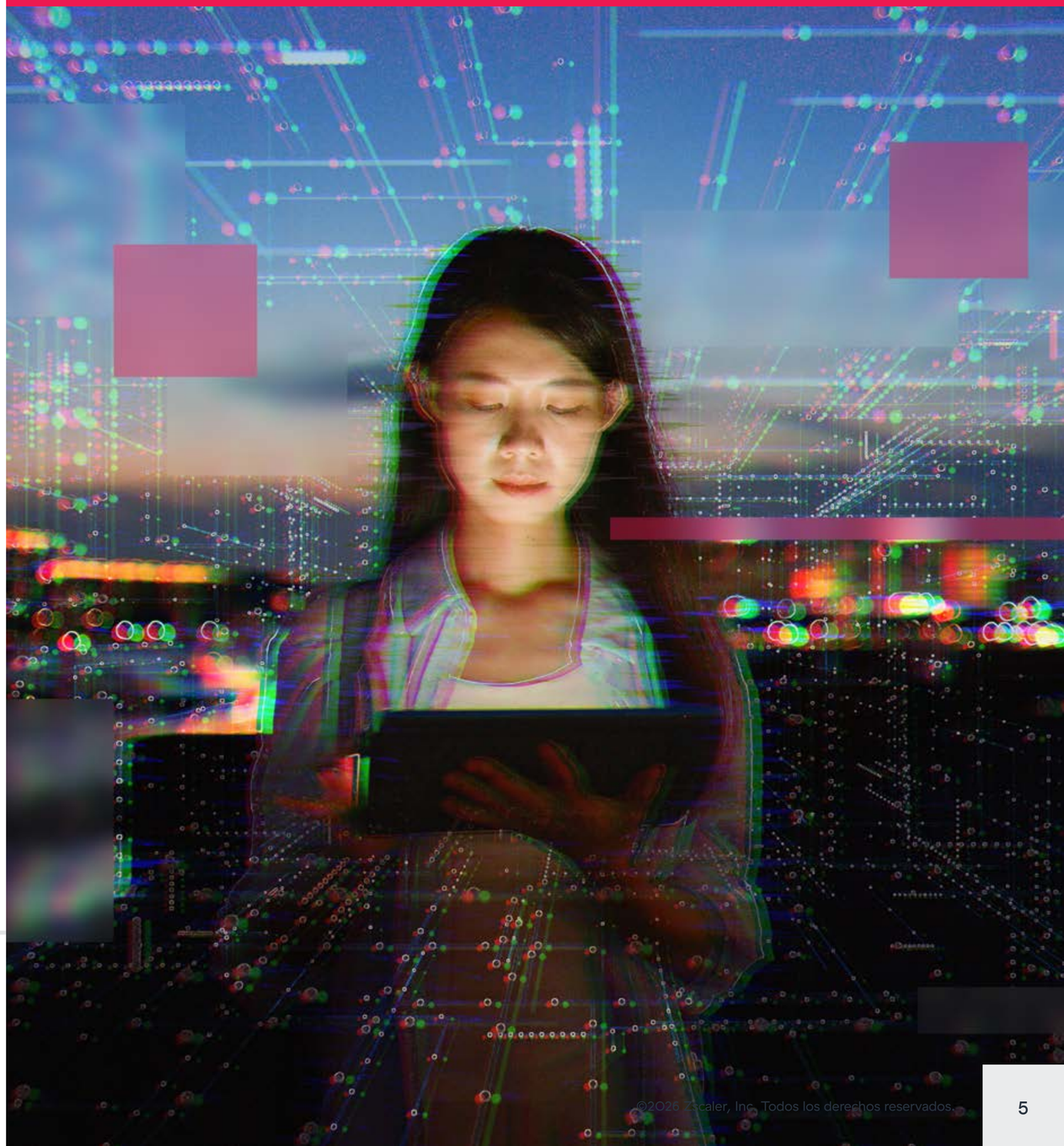
Las tasas de inspección SSL oscilan entre **el 0.4 % y el 68.1 %, con un promedio de aproximadamente el 36 %**. Esto se sitúa muy por debajo del nivel de inspección del 70 % que se considera necesario para una cobertura de seguridad aceptable con controles de seguridad en línea (DLP, IPS, sandboxing, salida de IA).

En el peor escenario posible, la identidad con mayores privilegios podía acceder a **483,250 solicitudes**

En un caso representativo específico de comodines, solo 195 dominios comodín expusieron más de **108,033** aplicaciones. (Un comodín es un símbolo como "*" o un punto inicial "." utilizado en las políticas para representar múltiples dominios o rutas a la vez, lo que introduce eficiencia administrativa, pero amplía drásticamente la superficie de ataque cuando no se controla adecuadamente.

Nueve de cada diez organizaciones tienen dispositivos VPN expuestos. Cada uno de ellos representa una superficie de ataque previa a la autenticación que un modelo de IA de vanguardia puede explorar a la velocidad de las máquinas.

Más de 750 activos presentan vulnerabilidades explotadas conocidas por CISA en infraestructuras conectadas a Internet, confirmadas mediante sondeos activos, y los exploits se estaban utilizando activamente en la práctica en el momento de la evaluación.

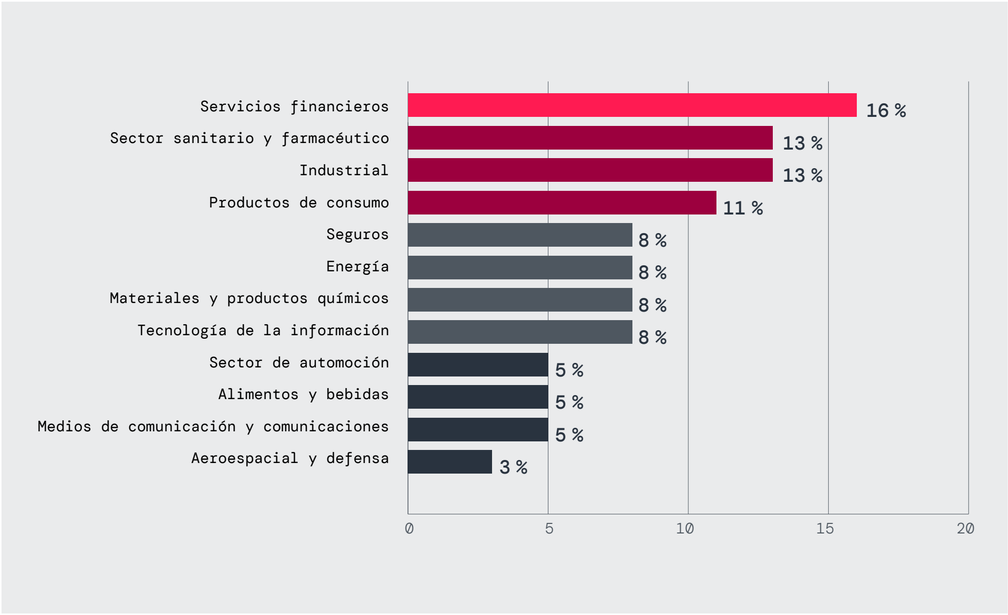


©2026 Zscaler, Inc. Todos los derechos reservados.



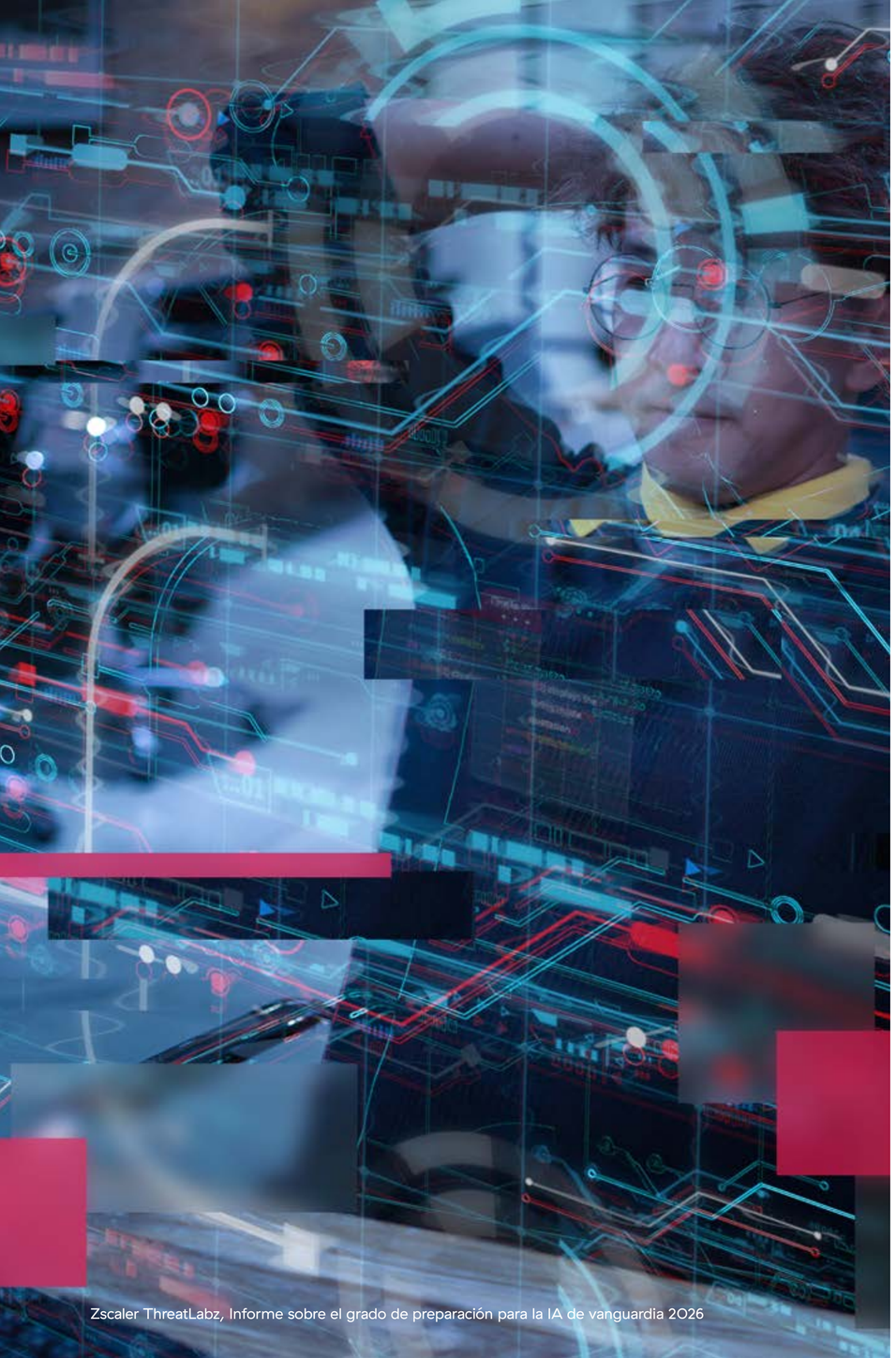
Cohorte de evaluación

Las conclusiones de este informe se basan en las evaluaciones del grado de preparación para la IA de vanguardia realizadas en cientos de organizaciones empresariales durante el primer semestre de 2026. El 32 % son empresas incluidas en la lista Fortune 500. El grupo abarca 12 sectores compuestos de la siguiente manera:



Las organizaciones evaluadas eran principalmente grandes empresas con decenas de miles de activos localizables en Internet. Los activos detectables a través de Internet incluyen cualquier sistema, aplicación o punto final accesible desde Internet pública sin autenticación.

©2026 Zscaler, Inc. Todos los derechos reservados.



Perspectivas de amenazas

Los modelos de IA de vanguardia han demostrado la capacidad de identificar y encadenar vulnerabilidades empresariales a la velocidad de las máquinas.

Los hallazgos anteriores describen la situación actual: condiciones observadas en el grupo durante el primer semestre de 2026. El análisis de amenazas que se presenta a continuación explica la importancia del cronograma de activación. Las capacidades de los adversarios que explotan estas condiciones están madurando más rápido que los ciclos operativos empresariales habituales. Cada tendencia que se describe a continuación refuerza directamente uno o más hallazgos de la evaluación.

En el programa de pruebas de modelos de vanguardia de Zscaler, los modelos de la clase de IA de vanguardia se evaluaron en tres modos: caja negra (sin conocimiento interno), caja blanca (con código fuente, binarios compilados y archivos estáticos) y caja gris (con contexto del sistema parcial o completo). Estas pruebas confirmaron de manera independiente que, al proporcionarles un contexto sólido, los modelos de IA de vanguardia pueden razonar en entornos empresariales con una alta relación señal–ruido e identificar de constantemente las cadenas de explotación de varios pasos que los analistas humanos y los escáneres tradicionales omitieron. En arquitecturas de producción, el contexto sólido demostró ser un factor multiplicador para el descubrimiento de vías de ataque.

Esta validación independiente confirma que las deficiencias de configuración identificadas en los hallazgos 1 a 6 pueden ser explotadas por un modelo de IA de vanguardia que opere de manera autónoma, y que los modelos de vanguardia actuales pueden aprovecharse de estas condiciones específicas, tal como se observó en la cohorte evaluada.

Es muy probable que el descubrimiento de vulnerabilidades mediante inteligencia artificial acelere la disponibilidad de nuevos exploits en 2027.

El tiempo medio de explotación se ha vuelto negativo: actualmente, las vulnerabilidades se aprovechan antes de que se puedan implementar los parches en los entornos empresariales, incluso cuando se dispone de un sistema de aplicación automática de parches. El desafío es de carácter estructural, ya que los plazos de explotación se han acortado más rápido de lo que la gestión del cambio en las empresas puede adaptarse.

En el caso de los activos con acceso a Internet que presentan valores KEV y puntuaciones EPSS iguales o cercanas a 1.0, tal como se confirmó en toda la cohorte evaluada, el lapso entre la divulgación y la explotación activa se ha reducido a horas. Las organizaciones que confían en la aplicación de parches durante los periodos de mantenimiento estándar casi con toda seguridad se enfrentarán a la explotación de vulnerabilidades conocidas durante el ciclo de remediación.

©2026 Zscaler, Inc. Todos los derechos reservados.

Es probable que la expansión de la superficie de ataque de la IA supere a los programas de gobernanza durante los próximos 12 a 18 meses.

La ausencia generalizada de gobernanza de la IA refleja un desfasaje estructural entre la velocidad de implementación de la IA y la adaptación de los programas de seguridad. Los equipos de desarrollo, los proveedores de SaaS y las unidades de negocio están implementando copilotos, puntos finales de inferencia e interfaces MCP sin una revisión de seguridad. Según la tasa de proliferación observada, es probable que la superficie de IA no regulada se duplique antes de que la gobernanza alcance una cobertura operativa. Las organizaciones que se demoren probablemente se enfrentarán a la explotación durante el período de evaluación.

Es muy probable que el movimiento lateral mediante el compromiso de identidad se convierta en la técnica dominante posterior a la explotación contra entornos Zero Trust.

Los resultados obtenidos mediante el método comodín indican que la suplantación de identidad es la vía de menor resistencia. Es muy probable que un adversario de la clase de IA de vanguardia priorice el robo de credenciales sobre la explotación tradicional de la red. La recompensa es desproporcionada: miles de solicitudes a través de una sola identidad. Las organizaciones con sesiones que nunca caducan y sin validación del estado del dispositivo son casi con toda seguridad los primeros objetivos.

Es probable que el punto ciego del tráfico cifrado se convierta en el principal canal de exfiltración para el robo de datos impulsado por la IA.

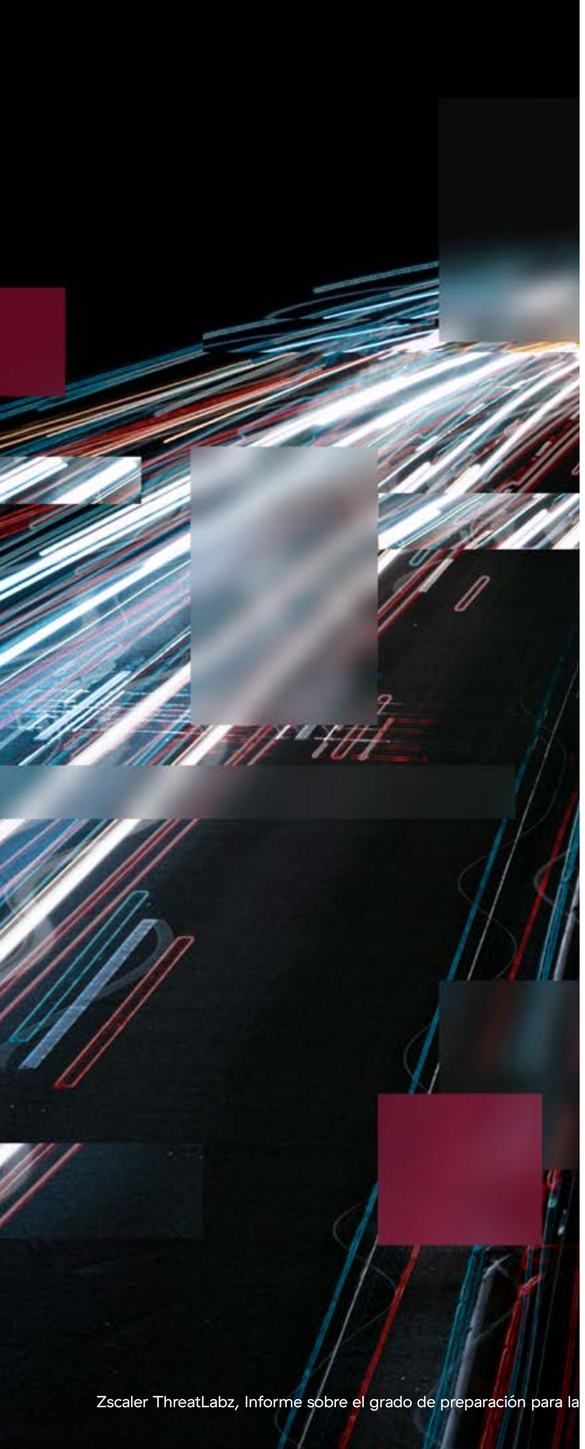
La IA de vanguardia puede codificar datos confidenciales dentro de llamadas API cifradas, interacciones de modelos o tráfico en la nube. Las herramientas de IA que generan patrones de apariencia legítima, combinadas con sistemas de seguridad incapaces de inspeccionarlos, crean un canal de exfiltración invisible y de amplio ancho de banda. Es probable que esto ya se esté explotando y casi con toda seguridad se intensificará.

El encadenamiento autónomo de exploits probablemente reducirá el tiempo desde el acceso inicial hasta la vulneración total a minutos.

La convergencia de dispositivos VPN expuestos, vulnerabilidades KEV, segmentación con comodines y detección latente crea un entorno en el que un adversario de IA probablemente pueda pasar del reconocimiento al acceso interno generalizado sin activar ninguna respuesta. Los datos sugieren que hoy en día, esta cadena completa es factible contra la mayor parte de la cohorte. Es probable que en un plazo de 12 meses los operadores más sofisticados dispongan de herramientas autónomas que ejecuten procesos de extremo a extremo.



©2026 Zscaler, Inc. Todos los derechos reservados.



Metodología

Validación de amenazas de la IA de vanguardia

Utilizamos OpenAI GPT-5.5 y Anthropic Claude Opus 4.8 en condiciones controladas y con un contexto empresarial real. Utilizamos la denominación de "clase de IA de vanguardia" (Frontier-AI-class) porque el modelo específico importa menos que el umbral de capacidad: razonamiento autónomo de múltiples pasos, uso de herramientas, recuperación de contexto fundamentada y secuenciación de explotación autodirigida. Cualquier modelo que alcance o supere ese umbral, ya sea actual o futuro, representa la clase de adversario que mide esta evaluación.

Paralelamente, Zscaler llevó a cabo pruebas internas de modelos de vanguardia utilizando tres modos: caja negra (sin conocimiento interno), caja blanca (código fuente, binarios compilados, archivos estáticos) y caja gris (contexto del sistema parcial o completo). Estas pruebas confirmaron que los modelos de IA de vanguardia logran un razonamiento con una alta relación señal-ruido en entornos empresariales cuando operan con un contexto real.

Marco de puntuación

Cada organización fue calificada contra el modelo de amenazas de IA de vanguardia utilizando un compuesto ponderado de 100 puntos en cinco pilares defensivos: Ocultarse del ataque (25 puntos, 25 %), Limitar el radio de explosión (25 puntos, 25 %), Gobernar el uso de la IA (20 puntos, 20 %), Proteger los datos (20 puntos, 20 %), y Engaño y Señuelos (10 puntos, 10 %). La puntuación evaluó tanto la propiedad del control como el estado de activación operativa.

Fuentes de datos:

Las evaluaciones se basaron en múltiples categorías de información, incluyendo la gestión de la superficie de ataque externa (EASM), la priorización de vulnerabilidades impulsada por la IA, el descubrimiento de superficies específicas de IA, las revisiones de configuración de acceso privado y, cuando fue posible, la telemetría en línea que abarca la cobertura de inspección, la aplicación de políticas y la postura de autenticación.

Modelo de evaluación

Aprovechando la amplia visibilidad de EASM, el motor Opus validó y dio prioridad a los hallazgos para identificar problemas explotables confirmados que requerían medidas, refinando el conjunto de candidatos en un 84 % de media y en un 92 % en el caso más complejo. También reevaluó las vulnerabilidades CVE según el retorno de la inversión de remediación, contrastó los hallazgos con CISA KEV e identificó vías de explotación de varios pasos en toda la superficie de ataque validada.

Qué abarca (y qué no abarca) esta evaluación

La evaluación analiza la exposición de la superficie de ataque externa y el estado de configuración de los controles de seguridad. No evalúa la eficacia de la detección del SOC a nivel de reglas SIEM, el grado de preparación para la respuesta a incidentes, la postura de seguridad del proveedor de identidad ni la cobertura de detección de puntos finales. La puntuación refleja la probabilidad de que un adversario que utilice IA pueda penetrar, atravesar y exfiltrar sin encontrar un control efectivo. No representa una puntuación total de la postura de seguridad.

Fundamentalmente, los ataques de IA de vanguardia pueden comprimir la cadena de ataque, desde el reconocimiento hasta la exfiltración, en cuestión de minutos, por debajo del umbral de respuesta incluso de los SOC maduros y con la suficiente rapidez como para explotar las deficiencias de configuración antes de la intervención humana.

©2026 Zscaler, Inc. Todos los derechos reservados.

Hallazgos clave

Hallazgo 1

Las empresas pueden maximizar sus capacidades de defensa de IA existentes con unos pocos pasos.

Las organizaciones pueden mejorar significativamente sus defensas con prácticas de ciberhigiene.

Las oportunidades de activación son amplias. En el 64 % de las organizaciones, las plataformas de engaño con licencias están preparadas para su implementación con el fin de agregar una mayor profundidad de detección. El 15 % de las organizaciones tienen sistemas de prevención de intrusiones listos para pasar del modo de supervisión al modo de aplicación total de las normas. El 18 % cuenta con capacidades de autenticación que pueden extenderse a sitios adicionales. Mediante la optimización de la configuración, las organizaciones pueden alcanzar puntuaciones objetivo de entre 73 y 98 sobre 100. Partiendo de una puntuación media inicial de 37 para todo el grupo, esto representa una mejora individual considerable.

Por qué esto importa frente a las amenazas de clase de IA: las herramientas de seguridad que operan por debajo del nivel de aplicación total ofrecen una resistencia reducida a un adversario que se mueve a la velocidad de las máquinas.

Hallazgo 2

La gobernanza de la IA se encuentra en su punto de partida en todas las empresas evaluadas, independientemente del tamaño de la organización o del alcance del control.

Este hallazgo es universal. El 100 % de las organizaciones evaluadas carecen de autenticación basada en la identidad para su IA, y el 91 % tienen puntos finales de IA expuestos públicamente. La puntuación media del pilar de gobernanza de la IA fue de 2.1 sobre 20 (10.5 % del máximo). La puntuación más alta del grupo fue de 3 sobre 20.

La exposición es considerable: los puntos finales de IA pública eran comunes en toda la cohorte, incluidos los puntos finales de MCP expuestos a Internet y los chatbots o copilotos no autenticados. Entre los casos más graves observados se encontraban las puertas de enlace de GenAI de gravedad crítica expuestas a Internet pública.

Esta brecha existe porque las interfaces de IA son implementadas por equipos de desarrollo, unidades de negocio y proveedores externos a un ritmo que supera la capacidad de adaptación de los programas de seguridad. La gobernanza de la IA no es una iniciativa para el futuro. Se trata de una brecha actual y explotable en todas las organizaciones evaluadas, y la categoría de superficie de ataque no gestionada de más rápido crecimiento.

©2026 Zscaler, Inc. Reservados todos los derechos.

Hallazgo 3

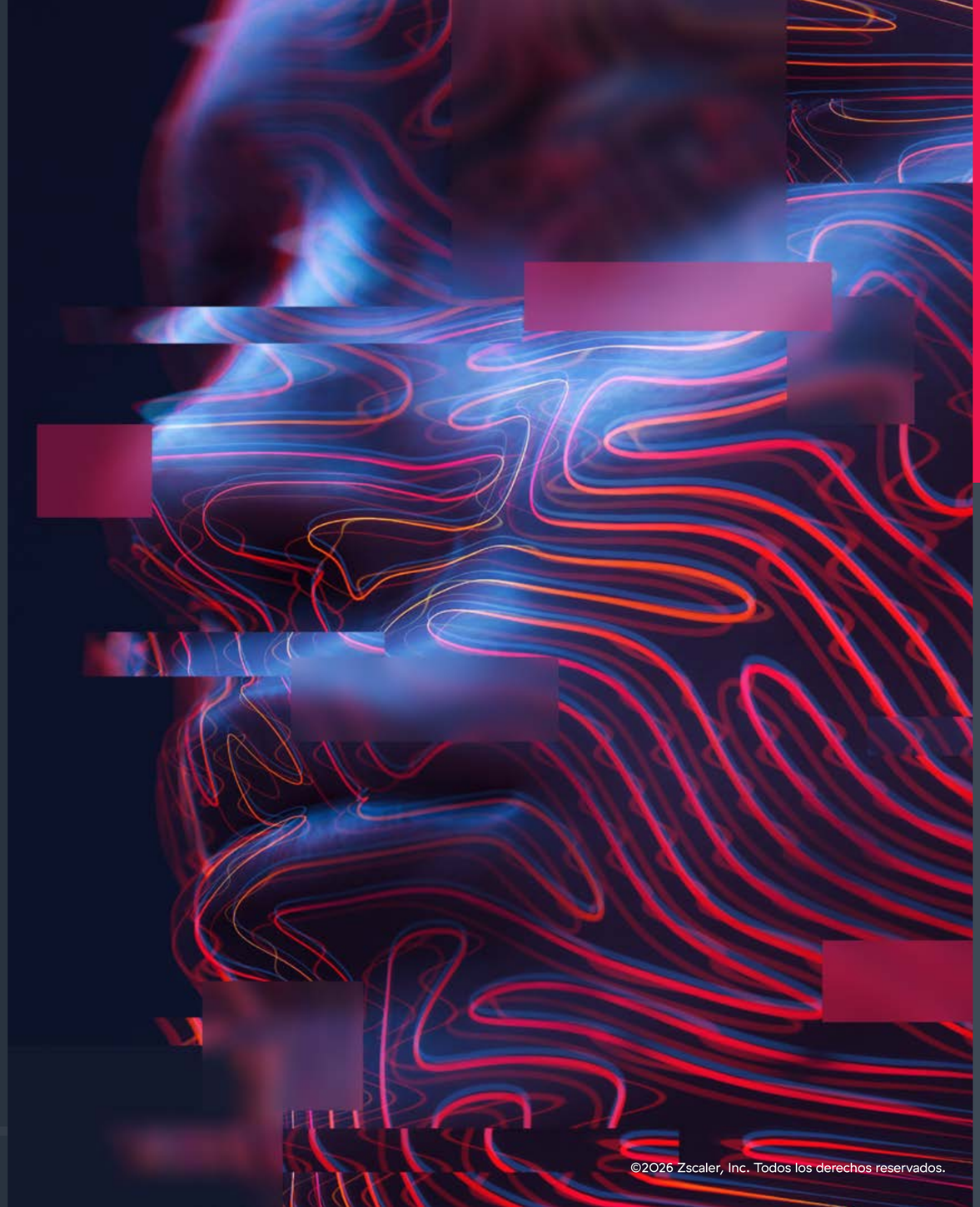
La inspección limitada del tráfico cifrado reduce la eficacia de los controles de seguridad implementados en la mayor parte del tráfico.

La inspección del tráfico cifrado ha sido durante mucho tiempo una prioridad en las operaciones de seguridad. En el contexto de las amenazas impulsadas por la IA, una cobertura de inspección completa se vuelve fundamental. Es casi seguro que los modelos de IA de vanguardia que realizan exfiltración utilizarán canales cifrados, en particular, las llamadas a la API de IA, las cargas a servicios de almacenamiento en la nube y las comunicaciones entre modelos que se mezclan con los patrones de tráfico legítimos. Un adversario que opera dentro del tráfico no inspeccionado es invisible para los controles en línea. Esto eleva la inspección del tráfico cifrado de una buena práctica operativa a un requisito fundamental contra las amenazas de clase de IA.

La empresa promedio inspecciona aproximadamente el 36 % de su tráfico cifrado. Ninguna organización cumple con el umbral de inspección calificada del 70 % definido en esta evaluación como la proporción de tráfico cifrado que pasa a través de controles de seguridad en línea activos (DLP, IPS, sandboxing, salida de IA) suficientes para que dichos controles logren una cobertura de seguridad óptima. En el peor de los casos, el 99.6 % de todo el tráfico cifrado pasa por la infraestructura de seguridad sin ser inspeccionado en absoluto.

Cuando el tráfico cifrado no se inspecciona, los controles de seguridad integrados, como DLP, IPS, sandboxing y la gobernanza de salida mediante IA, se vuelven ineficaces, lo que deja a las organizaciones a ciegas ante amenazas críticas como la exfiltración de datos y la distribución de malware. Dado que más del 86 % de las ciberamenazas y todas las interacciones de los modelos de IA utilizan canales cifrados, la inspección exhaustiva del tráfico es un requisito indispensable para defenderse de los ataques de clase IA.

Las causas fundamentales son operativas, no técnicas. La gran mayoría de las organizaciones utilizan un modelo de inspección con consentimiento explícito (predeterminado: “no inspeccionar”). En todo el grupo de organizaciones estudiadas, es frecuente que estas recurran a reglas SSL de “Omitir otras políticas” o “No inspeccionar”, que desactivan la inspección para categorías completas de tráfico. Entre las organizaciones afectadas, el número de estas reglas de omisión osciló entre 2 y 51.



©2026 Zscaler, Inc. Todos los derechos reservados.

Hallazgo 4

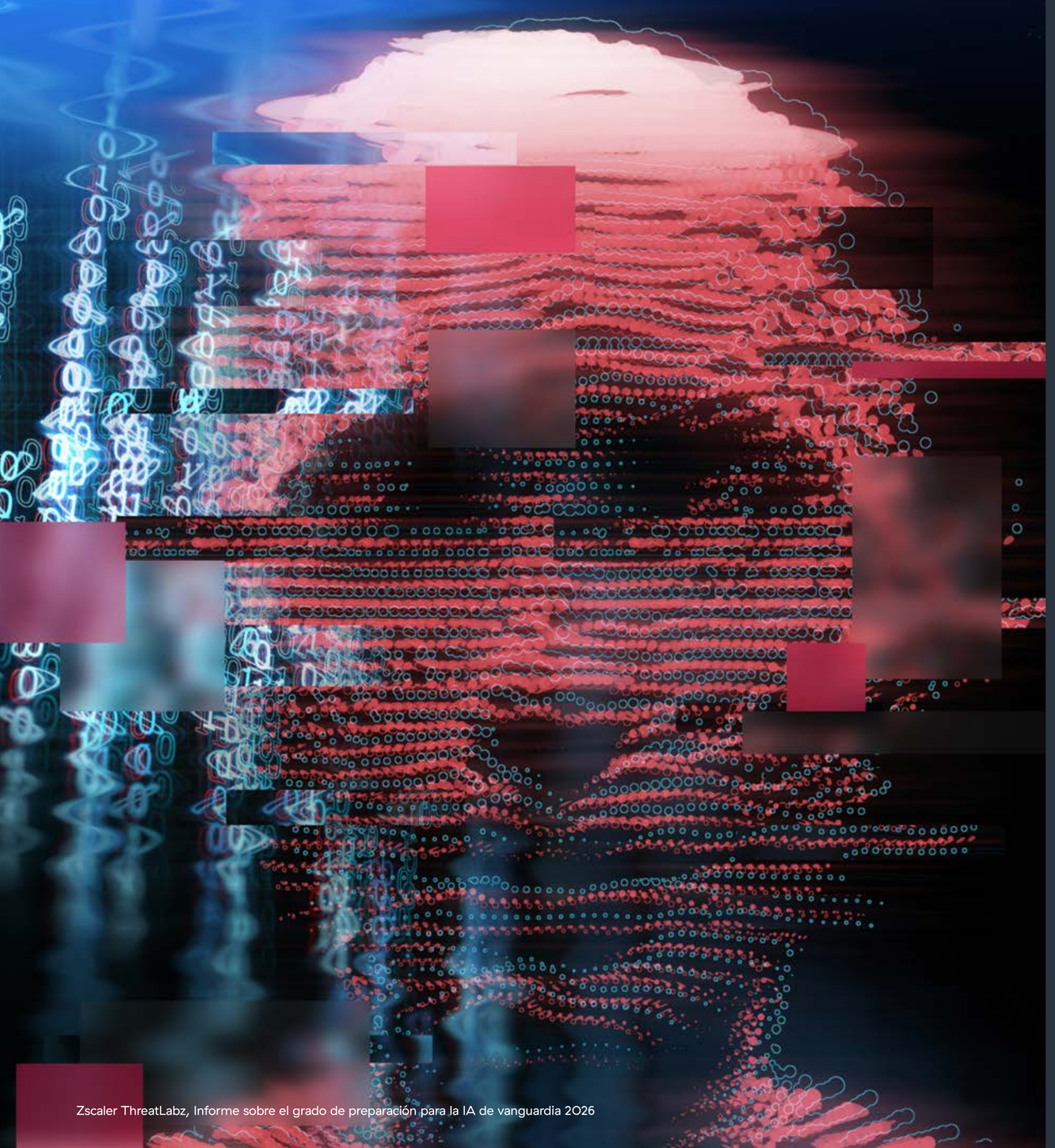
Los entornos evaluados aún no están configurados para detectar o prevenir ataques impulsados por la IA a la velocidad requerida.

El hallazgo 4 destaca una importante deficiencia en la detección, donde los controles de seguridad habilitados fallan porque sus configuraciones (que abarcan técnicas de engaño, sistemas de prevención de intrusiones, detección de anomalías y supervisión de sesiones) no están optimizadas para ataques de clase de IA, lo que deja a las organizaciones expuestas a amenazas que están capacitadas para bloquear, pero que no pueden identificar. En consecuencia, un ataque de la clase de IA de vanguardia permanece invisible para el SOC, sin desencadenantes de engaño, IPS, anomalías de comportamiento o tiempos de espera de sesión, lo que significa que la primera indicación de compromiso probablemente sea una notificación externa en lugar de una detección interna.

- **El engaño (el sistema de alerta temprana):** El 64 % de las organizaciones cuentan con plataformas de engaño cuya cobertura aún no se extiende a los segmentos críticos. Un atacante que se mueva lateralmente no encontraría trampas en la configuración actual. El engaño proporciona una detección eficaz e independiente de firmas contra ataques impulsados por la IA mediante el uso de señuelos de alta fidelidad para activar la contención inmediata y automatizada del movimiento lateral.
- **Prevención de intrusiones (bloqueo en línea):** Para el 15 % de las organizaciones, existe una clara oportunidad de reforzar sus defensas mediante la transición de sus sistemas de prevención de intrusiones al modo de aplicación total. Cuando esta opción está habilitada, las reglas de “Omitir otras políticas” implican que la mayor parte del tráfico nunca se evalúa. Un exploit conocido que transita por canales cifrados pasaría sin ser inspeccionado y sin ser bloqueado.
- **Detección de anomalías en la capa de aplicación (la señal de comportamiento):** Varias organizaciones aún no habían configurado reglas en línea. No se detectó ninguna anomalía en el volumen de acceso mediante comodines. Probablemente, una sola identidad que acceda a miles de aplicaciones en secuencia no generará ninguna alerta.
- **Supervisión de sesiones e identidades (el desencadenante de la contención):** Los tiempos de espera de autenticación configurados sin caducidad en varias organizaciones implican que las sesiones comprometidas podrían conservar su validez indefinidamente. En una organización, ninguna de las 18 reglas de acceso incluía comprobaciones de la postura del dispositivo. Una credencial robada podría funcionar durante semanas sin activar la reautenticación.
- **Lo que probablemente vea el SOC ante un ataque de la clase de AI de vanguardia es lo siguiente:** El SOC no ve nada. No hay activación de engaño. No hay alerta de IPS. No hay anomalía de comportamiento. No hay tiempo de espera de sesión. La primera indicación de compromiso probablemente sería una notificación externa en lugar de una detección interna.



©2026 Zscaler, Inc. Todos los derechos reservados.



Hallazgo 5

La amplia segmentación ha convertido a las plataformas de acceso privado en autopistas de circulación lateral.

En nuestro análisis, el 64 % de las organizaciones que utilizan ZTNA tenían configuraciones con muchos caracteres comodín que podían exponer más aplicaciones de las previstas. En un caso, tan solo 195 dominios comodín permitieron acceder a más de 108,000 aplicaciones, lo que potencialmente posibilitó que una sola identidad accediera a una parte muy amplia del entorno. Si bien estas configuraciones suelen introducirse durante la implementación o la migración para simplificar la incorporación, pueden persistir a medida que los entornos crecen, ampliando el acceso más allá de la intención original. Sin una revisión y un perfeccionamiento periódicos, las políticas basadas en comodines pueden socavar los objetivos de privilegios mínimos y aumentar el impacto de una identidad comprometida, por ejemplo, a través del phishing o el robo de sesión.

Varias organizaciones agravan esto con tiempos de espera de autenticación configurados en “nunca”. Una sesión comprometida permanece válida indefinidamente. Incluso la organización con la puntuación más alta en el límite del radio de la explosión (14.4 sobre 25) seguía utilizando configuraciones comodín. Las organizaciones que obtuvieron puntuaciones más bajas prácticamente no realizaban ninguna segmentación.

Este patrón significa que el entorno de acceso privado aún no refleja la segmentación de privilegios mínimos que las arquitecturas Zero Trust están diseñadas para aplicar. Frente a un adversario de la clase de IA de vanguardia capaz de comprometer la identidad a gran escala, las configuraciones dominadas por comodines representan la vía de movimiento lateral de mayor impacto en los entornos evaluados.

©2026 Zscaler, Inc. Todos los derechos reservados.

Hallazgo 6

Más de un tercio de las empresas mantienen vulnerabilidades conocidas y explotadas en su infraestructura conectada a Internet.

Estas vulnerabilidades a menudo persisten no porque sean desconocidas, sino porque la remediación debe competir con realidades operativas como los períodos de mantenimiento, los requisitos de control de cambios y las dependencias de la infraestructura. Más de la mitad de la cohorte (53 %) recibieron recomendaciones de remediación para las vulnerabilidades explotadas conocidas (KEV, por sus siglas en inglés) de CISA en activos expuestos a Internet. Durante las investigaciones activas, más de un tercio de las empresas mantuvieron estos KEV que podían ser explotados activamente en su perímetro.

El resultado es una creciente discrepancia entre la rapidez con la que los defensores pueden solucionar los problemas y la rapidez con la que los atacantes pueden identificar y explotar las vulnerabilidades. A medida que los adversarios utilizan cada vez más la IA para acelerar el descubrimiento y la explotación de vulnerabilidades, es probable que ese período de exposición se reduzca aún más.

Se detectaron tres CVE en la mayor parte de la cohorte. Todas figuran en el catálogo de vulnerabilidades explotadas conocidas de CISA. Todos los resultados fueron confirmados mediante sondeo activo.

- **CVE-2024-38475 (Apache mod_rewrite):** puntuación EPSS 0.9988 a 0.9997
- **CVE-2021-40438 (Apache mod_proxy SSRF):** puntuación EPSS 0.9999 a 1.0000
- **CVE-2023-44487 (HTTP/2 reseteo rápido):** puntuación EPSS 0.999

Su persistencia, incluso en organizaciones sofisticadas, refleja una realidad operativa: la infraestructura a nivel web, en particular Apache y F5/NGINX, a menudo queda fuera de la automatización estándar de parches y requiere coordinación manual entre los periodos de cambio.



Si bien la exposición actual refleja las vulnerabilidades conocidas, las organizaciones que mantienen KEV en su perímetro también se enfrentan al riesgo de que la explotación de vulnerabilidades aún no reveladas en los mismos sistemas pueda surgir antes de que se complete la remediación existente.

El hallazgo 5 amplifica aún más este riesgo, ya que cada activo expuesto a Internet sin parchear representa un punto de entrada previo a la autenticación y, en entornos donde una sola identidad puede llegar a cientos de miles de aplicaciones, una vulnerabilidad CVE explotada puede extender el acceso mucho más allá del servidor afectado.

Las acciones prioritarias consisten en mantener una visibilidad continua de los activos mediante la gestión de la superficie de ataque externa, reducir la superficie de ataque detectable mediante el acceso intermediado, acortar los ciclos de detección, clasificación y aplicación de parches para los activos expuestos justificados por el negocio mediante la priorización basada en la posibilidad de explotación, y utilizar la aplicación autónoma de parches virtuales durante los periodos de remediación.

©2026 Zscaler, Inc. Todos los derechos reservados.

La nueva cadena de ataque: Cómo la IA de vanguardia encadena de manera autónoma las brechas pequeñas en una única secuencia de explotación

Estos hallazgos no son casos aislados. Están reforzando los vínculos de una cadena de explotación más amplia. Individualmente, cada una representa una brecha que se puede abordar; en conjunto, generan una exposición considerablemente mayor de lo que sugiere cualquiera de los hallazgos individualmente.

Un modelo de IA de vanguardia con contexto bien definido puede identificar y secuenciar de manera autónoma condiciones como una vulnerabilidad CVE sin parchear, un sistema IPS que aún no se aplica, una configuración comodín, un punto final de IA no controlado y una ruta de salida no inspeccionada hacia una ruta de explotación continua a la velocidad de las máquinas, sin intervención humana. Los escáneres tradicionales y las pruebas de penetración no detectan esta trayectoria compuesta a una velocidad comparable. La siguiente secuencia fue validada en los entornos evaluados:

Paso 1: Enumerar un dispositivo VPN expuesto (uno de más de 500) mediante escaneo de Internet. Identificar CVE-2024-38475 (Apache mod_rewrite, EPSS 0,9997) en la capa web.

Paso 2: Explotar la vulnerabilidad de preautenticación. Obtener acceso inicial sin credenciales. El sistema IPS está desactivado. No se genera ninguna alerta.

Paso 3: Extraer credenciales de la infraestructura VPN comprometida. Obtener una única identidad empresarial válida.

Paso 4: Autenticarse en la plataforma de acceso privado. La configuración comodín otorga acceso a más de 17,000 aplicaciones. El tiempo de espera está configurado en “nunca”. No se realiza ninguna comprobación del estado del dispositivo.

Paso 5: Las aplicaciones se procesan a velocidad de las máquinas. No hay señuelos de engaño en la ruta. El SOC no detecta ninguna alerta.

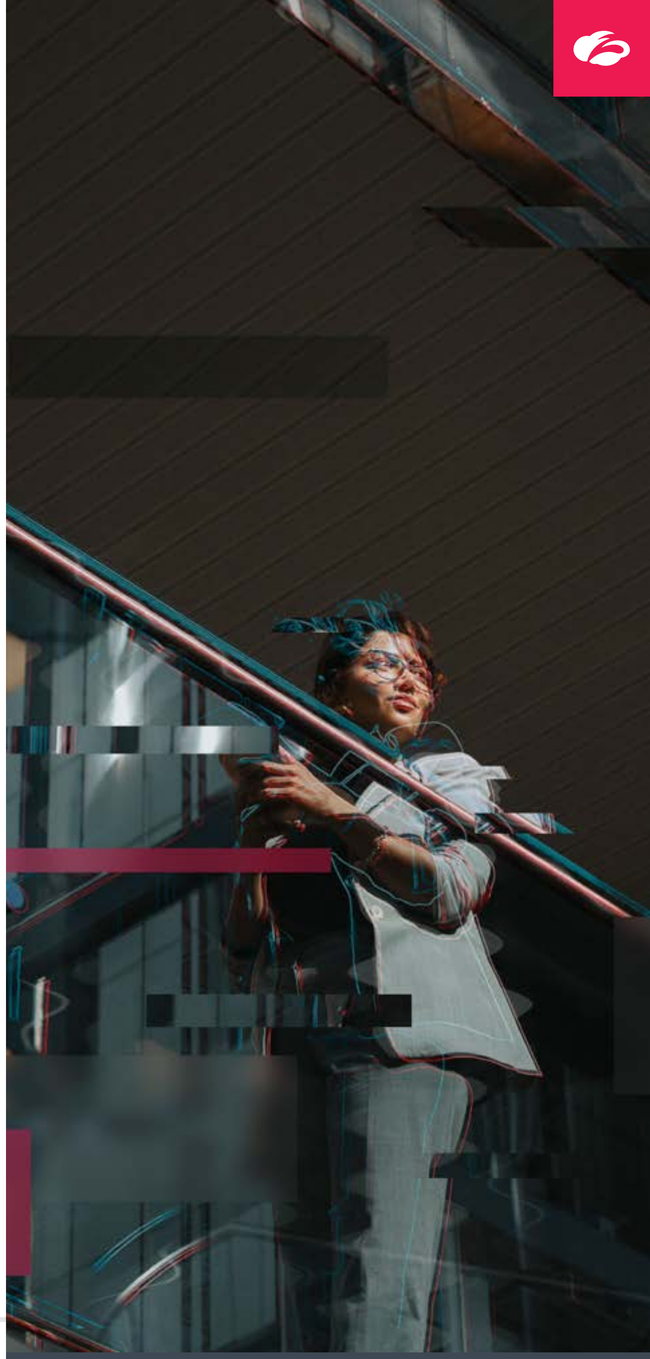
Paso 6: Llegar a un punto final de inferencia de IA no gobernado (uno de más de 300) accesible sin autenticación adicional. Consultar los datos exclusivos a través de la ventana de contexto del modelo.

Paso 7: Exfiltración de datos a través de un canal cifrado. Dos tercios del tráfico sin inspeccionar. El sistema DLP no puede detectar la salida. No se genera ninguna alerta.

Tiempo total desde el escaneo inicial hasta la exfiltración: Durante las pruebas internas contra adversarios, los sistemas de IA empresariales alcanzaron un tiempo medio hasta la primera falla crítica de tan solo 16 minutos.

En el programa de prueba de modelos de frontera de Zscaler (clase de AI de vanguardia y GPT 5.5 en tres modos: sin conocimiento interno, acceso al código fuente y contexto del sistema parcial/completo), se validó esta capacidad de razonamiento en cadena. El sistema de contexto basado en datos permite que los modelos de vanguardia identifiquen vías de explotación de múltiples pasos que ni los escáneres tradicionales ni los analistas humanos detectarían a la misma velocidad.

Todos los componentes fueron observados en los entornos de las organizaciones durante el primer semestre de 2026. La cadena representa el modelo adversario contra el cual se calibra la puntuación del grado de preparación para la IA de vanguardia.



©2026 Zscaler, Inc. Todos los derechos reservados.

Pruebas de caso: Tres organizaciones en desarrollo

Los siguientes casos prácticos anonimizados representan a organizaciones en diferentes etapas de su proceso de remediación. Ilustran las condiciones iniciales y la metodología operativa utilizada para subsanar las deficiencias detectadas durante la evaluación.

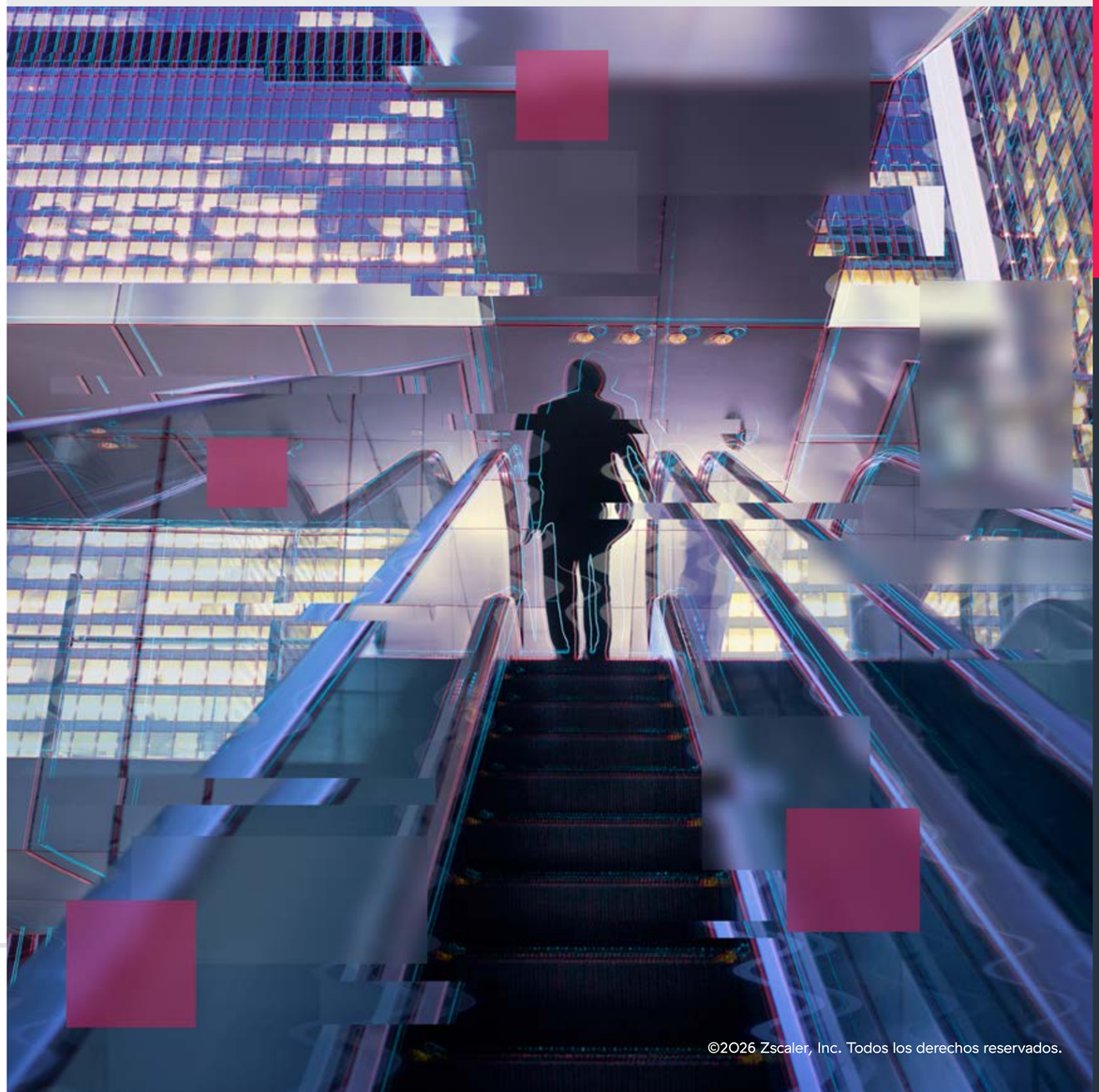
CASO 1

Organización global de servicios financieros

Estado inicial: más de 10,000 activos con acceso a Internet. Más de 4000 CVE validados (más de 900 únicos, 165 críticos). Tres de los CVE más frecuentes del grupo (CVE-2024-38475, CVE-2021-40438, CVE-2023-44487) confirmados en infraestructura expuesta.

Resultados de la evaluación: La extensa infraestructura de la organización expuesta a Internet presentaba vulnerabilidades críticas sin parchear, acumuladas a lo largo de sucesivas generaciones de infraestructura. La inspección SSL aún no se había extendido para cubrir la mayor parte del tráfico. Si bien se contaba con capacidades de engaño, estas no se habían implementado. Las interfaces de IA permanecían fuera del perímetro de gobernanza.

Se iniciaron medidas correctivas: Se dio prioridad a la aplicación de parches a 165 vulnerabilidades CVE críticas. Se cambió la inspección SSL de opción activada a opción desactivada. Se activó la detección de engaños en los segmentos financieros más importantes. Se inició el control de acceso mediante IA.



©2026 Zscaler, Inc. Todos los derechos reservados.



CASO 2

Productor mundial de alimentos

Estado inicial: 13 dispositivos VPN expuestos. 40 superficies de IA públicas fuera del perímetro de gobernanza (la mayor densidad de exposición de IA en relación al tamaño en la cohorte). 2 CVE validados en infraestructura expuesta.

Resultados de la evaluación: La rápida transformación digital de la organización superó la capacidad del programa de seguridad para incorporar nuevas superficies. Las unidades de negocio habían implementado dispositivos de IA como parte de iniciativas operativas, pero aún no se habían integrado a la arquitectura de seguridad. No se había establecido una política de segmentación para el entorno de acceso privado. La infraestructura VPN permanecía expuesta a Internet, pendiente de migración.

Se iniciaron medidas correctivas: Protección VPN detrás de un acceso privado (inmediato). Control de acceso mediante IA para 40 puntos finales. Creación de políticas de segmentación desde el principio. Implementación de técnicas de engaño en segmentos de producción.

CASO 3

Conglomerado mundial de artículos de lujo

Estado inicial: 487 nombres de dominio públicos incluidos en el ámbito de aplicación. 922 TB de tráfico cifrado no inspeccionado mensualmente. 10,000 aplicaciones definidas.

Resultados de la evaluación: La arquitectura de inspección de la organización se había implementado con un modelo de participación voluntaria conservador, lo que dejaba sin examinar la mayor parte del tráfico cifrado. Con 10,000 aplicaciones y una segmentación prácticamente nula, el entorno había crecido sin una gobernanza de acceso adecuada.

Se iniciaron medidas correctivas: La conversión de la inspección SSL tiene la máxima prioridad. Se reemplazarán los comodines en 10,000 aplicaciones. Se activará la detección de engaños. Se establecerá una base de referencia para la gobernanza de la IA.

©2026 Zscaler, Inc. Todos los derechos reservados.

Realismo en la implementación: Lo que realmente requiere el sprint de 90 días

Técnicamente, es posible alcanzar la resiliencia en un plazo de 90 días, dado que, en la mayoría de los casos, los controles ya están bajo control y las cadenas de dependencia se comprenden bien. Sin embargo, debido a que no es algo que se logre sin esfuerzo, debe ser un mandato ejecutivo. Cerrar la brecha de preparación de la IA de vanguardia es una iniciativa empresarial que se beneficia de la visibilidad a nivel de la junta directiva, la alineación entre el CISO y el CEO, y una capacidad de ejecución dedicada. Se trata de una oportunidad para lograr una alineación compartida en un panorama de amenazas en constante evolución. Los responsables de seguridad deberían:

1. Informar a sus juntas directivas sobre su grado actual de preparación, enmarcándolo dentro del contexto regulatorio de la SEC, DORA y NIS2.

2. Conseguir el patrocinio de la alta dirección para un proyecto intensivo de 90 días con responsables designados e informes periódicos a la alta gerencia.

3. Asignar equipos pequeños y especializados para llevar a cabo la implementación, teniendo en cuenta que las dificultades que se presenten debido a la ampliación de la cobertura de inspección, los ajustes de acceso durante el reemplazo de comodines y situaciones similares, son parte prevista del proceso de activación.

La inspección SSL requiere una alineación interfuncional. Las organizaciones deben coordinar el trabajo con los equipos de TI y seguridad para mapear completamente la cobertura de inspección actual, identificar qué flujos de tráfico requieren inspección y determinar dónde la inspección es operativamente inviable (por ejemplo, aplicaciones con certificados fijos o integraciones heredadas). Cuando no se pueda aplicar la inspección, los controles de mitigación, como el aislamiento del navegador, la prevención de pérdida de datos en el punto final, la inspección de API basada en CASB y el filtrado de la capa DNS, pueden compensar la visibilidad y ofrecer prevención de amenazas sin necesidad de descifrado en línea.

La sustitución de comodines requiere la coordinación de los responsables de las aplicaciones. Reemplazar más de 200 comodines por nombres de dominio completos (FQDN) no es tarea exclusiva del equipo de seguridad. Cada comodín se resuelve en aplicaciones que son propiedad de unidades de negocio, equipos de desarrollo y proveedores, quienes deben validar la continuidad del acceso. El equipo de seguridad define la arquitectura objetivo, pero su ejecución requiere coordinación interfuncional.

El cronograma presupone una capacidad de ejecución dedicada. Las organizaciones que asignaron un equipo de sprint específico (de 2 a 4 ingenieros) avanzaron según lo previsto. Quienes buscaban optimizar sus cargas de trabajo existentes se beneficiaron de apoyo adicional para mantener el ritmo. Su éxito dependió del respaldo de la alta dirección, incluyendo una comunicación regular con los altos directivos, un equipo dedicado con capacidad protegida e informes semanales de hitos en comparación con la puntuación objetivo.

Si bien cualquier actualización de seguridad importante puede ocasionar cambios operativos temporales, estos pasos son esenciales para construir una defensa más sólida a largo plazo. Nuestra evaluación destaca que abordar de manera proactiva las deficiencias, como la puntuación inicial de 37, es un paso fundamental para gestionar las amenazas digitales modernas. Equilibrar estos ajustes a corto plazo con la necesidad de una protección sólida es fundamental para fortalecer la postura de seguridad general de la organización.



©2026 Zscaler, Inc. Todos los derechos reservados.

Recomendaciones

1 Habilitar el modo de aplicación en los controles de seguridad

Todas las organizaciones evaluadas contaban con controles de alto impacto (IPS, autenticación, bloqueo de IP maliciosas, protección contra ataques de dominio) que pueden aplicarse por completo mediante cambios de configuración. Los equipos de operaciones de seguridad deben auditar todas las capacidades propias para comprobar su estado de activación, aprovechando las auditorías de configuración del proveedor y el soporte de activación según corresponda, y habilitarlas en modo de aplicación en un sprint de 14 días.

Reconocemos la preocupación de todos los equipos SOC: cambiar el IPS de supervisión a bloqueo puede generar falsos positivos e interrumpir el tráfico. El enfoque correcto es una implementación por fases: comenzar en segmentos no críticos, ajustar y validar, y luego expandir a producción con criterios claros de reversión.

2 Proteja inmediatamente todas las superficies de IA con acceso basado en la identidad

Todos los puntos finales de IA accesibles desde Internet pública deben estar protegidos por la verificación de identidad empresarial. Esto es una recomendación para el 100 % de las organizaciones evaluadas. Mientras no se implemente el acceso basado en la identidad del usuario, la infraestructura de IA seguirá siendo vulnerable a cualquier adversario que se conecte a Internet.

3 Aumente la inspección SSL/TLS a un mínimo del 70 % de cobertura calificada dentro de los 90 días

Invierta el modelo de suscripción voluntaria a exclusión voluntaria: inspeccione todo el tráfico de manera predeterminada y luego cree exclusiones específicas. Elimine las reglas de “Omitir otras políticas”. Reduzca las excepciones a la regla de “No inspeccionar” al mínimo necesario. Prevea entre 2 y 4 semanas de interrupción en la aplicación. Asigne capacidad de respuesta ante picos de demanda del servicio de asistencia técnica. Sin esta cobertura, DLP, IPS, sandboxing y los controles de salida de IA no pueden funcionar en la mayor parte del tráfico.

La inspección SSL debe basarse en el riesgo, no ser universal. Comience con las aplicaciones más críticas y el tráfico de mayor riesgo, y luego amplíe de manera selectiva. Dado que algunas aplicaciones y servicios requieren excepciones, como la fijación de certificados o debido a restricciones regulatorias, las organizaciones deben definir con antelación los procedimientos de excepción y los procedimientos de respuesta rápida. Cuando la inspección no sea posible, reduzca las brechas de visibilidad con controles compensatorios.

4 Reemplace los dominios comodín con FQDN con nombre en los segmentos críticos.

Audite todas las configuraciones de acceso privado en busca de entradas comodín. Priorice la sustitución en los segmentos más estratégicos: sistemas financieros, repositorios de propiedad intelectual e infraestructura de identidad. Implemente tiempos de espera de autenticación, validación del estado del dispositivo en cada regla de acceso y segmentación a nivel de las aplicaciones que limite cada identidad únicamente a las aplicaciones necesarias. Planifique la coordinación interfuncional con los responsables de las aplicaciones.



©2026 Zscaler, Inc. Todos los derechos reservados.

5 Implemente tácticas de engaño en segmentos críticos de la red

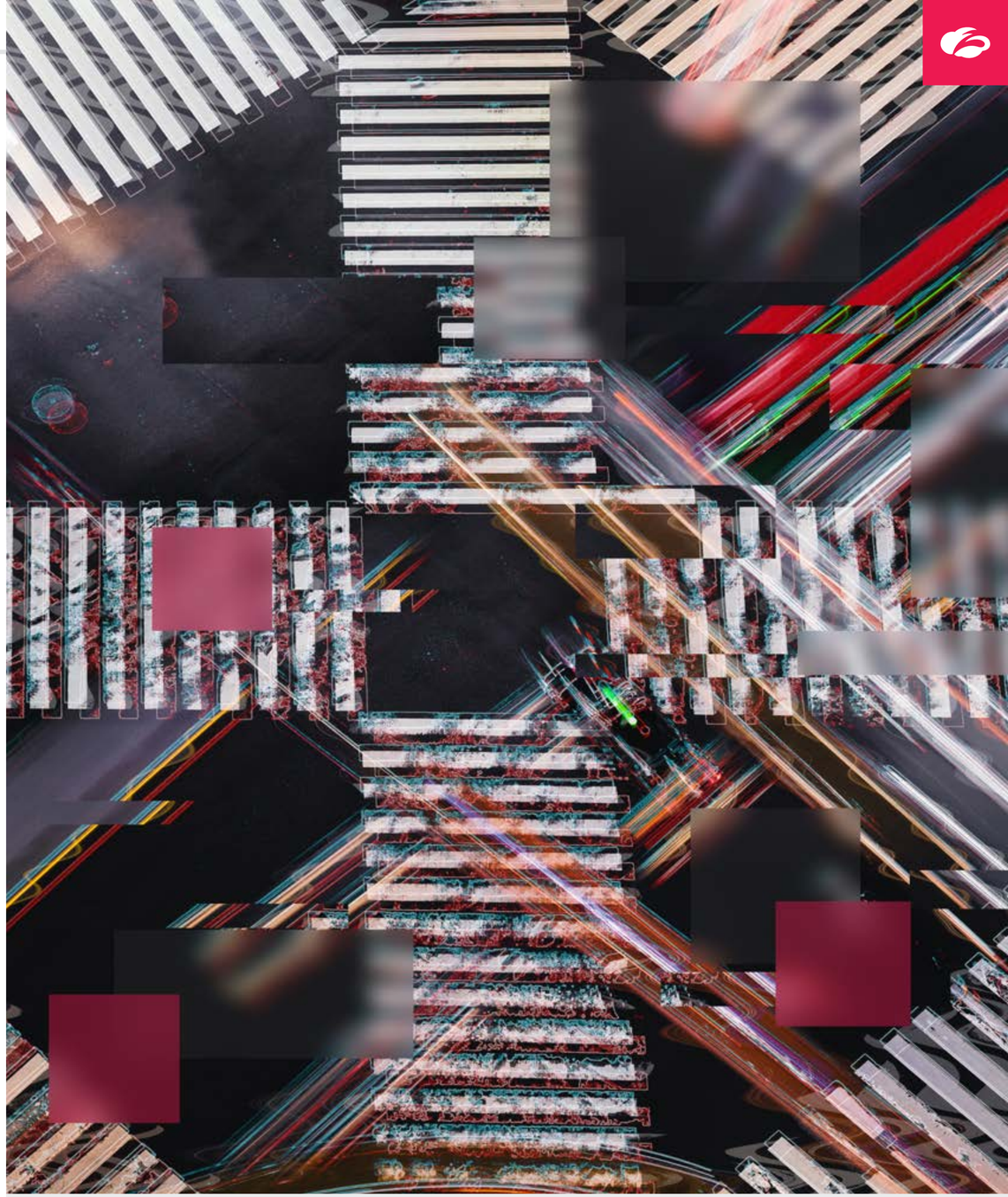
Las organizaciones pueden reforzar sus defensas colocando señuelos cerca de sus aplicaciones más críticas para crear una capa de detección independiente con alertas de alta fidelidad que ayude a los equipos del SOC a responder con mayor rapidez. Cuando se activa un señuelo, Deception revoca automáticamente el acceso del usuario comprometido a las aplicaciones privadas a través de ZPA en todos los dispositivos que hayan iniciado sesión, sin necesidad de que las apruebe un analista. Esto detiene a los atacantes antes de que puedan pivotar, impidiendo el movimiento lateral a la velocidad del ataque, no a la velocidad de la cola del SOC.

6 Traslade los dispositivos VPN expuestos a Internet a un sistema de acceso Zero Trust gestionado.

Cada dispositivo VPN detectable constituye una superficie de ataque antes de la autenticación. Implemente acceso gestionado a toda la infraestructura VPN en un plazo de 30 días. Retire de servicio los dispositivos expuestos en un plazo de 90 días. Esto elimina el primer punto de entrada que exploraría un modelo de IA de vanguardia.

7 Corrija las vulnerabilidades de CISA-KEV en los activos expuestos a Internet dentro de las 72 horas posteriores a la evaluación.

Las vulnerabilidades CVE-2024-38475, CVE-2021-40438 y CVE-2023-44487 tienen puntuaciones EPSS iguales o cercanas a 1.0 y se ha confirmado su explotación activa. Aplique parches o parches virtuales de inmediato. Cuando se requieran periodos de mantenimiento, implemente parches virtuales autónomos para bloquear los intentos de explotación durante ese período.



©2026 Zscaler, Inc. Todos los derechos reservados.

Nota sobre la exposición regulatoria

Las organizaciones que obtienen una puntuación inferior a 40 pueden tener deficiencias en sus controles, lo que aumenta tanto la probabilidad como el impacto potencial de un ciberincidente de importancia. Alentamos a las organizaciones a que revisen y pongan en práctica las recomendaciones de este informe para ayudarles a reducir la exposición al riesgo y fortalecer la resiliencia. En algunos casos, un incidente relacionado con estas deficiencias también podría generar consideraciones de divulgación conforme al punto 1.05 de la SEC (Formulario 8-K).

Para las organizaciones sujetas a DORA o NIS2, la falta de controles fundamentales también puede indicar oportunidades para fortalecer la gestión de riesgos de las TIC y la resiliencia operativa general.

Próximo_ paso

La evaluación del grado de preparación para la IA de vanguardia proporciona a cada organización una puntuación inicial cuantificada, una secuencia de medidas correctivas priorizadas y un objetivo a 90 días que se puede alcanzar principalmente mediante la ampliación o el perfeccionamiento de las capacidades existentes. Los datos de este informe establecen la base de referencia del sector. La pregunta que se plantea todo líder de seguridad es en qué posición se encuentra su organización en relación con esta base de referencia.

Ya ha visto los resultados. Ahora vea su exposición.

Hable con un experto de Zscaler para identificar los puntos donde podría existir riesgo de IA en su entorno y cuales debe abordar primero.

Conozca los riesgos. Actúe rápido. Manténgase seguro.
Obtenga más información en zscaler.com/mx/campaign/zscaler-ai-security-demo

Acerca de ThreatLabz

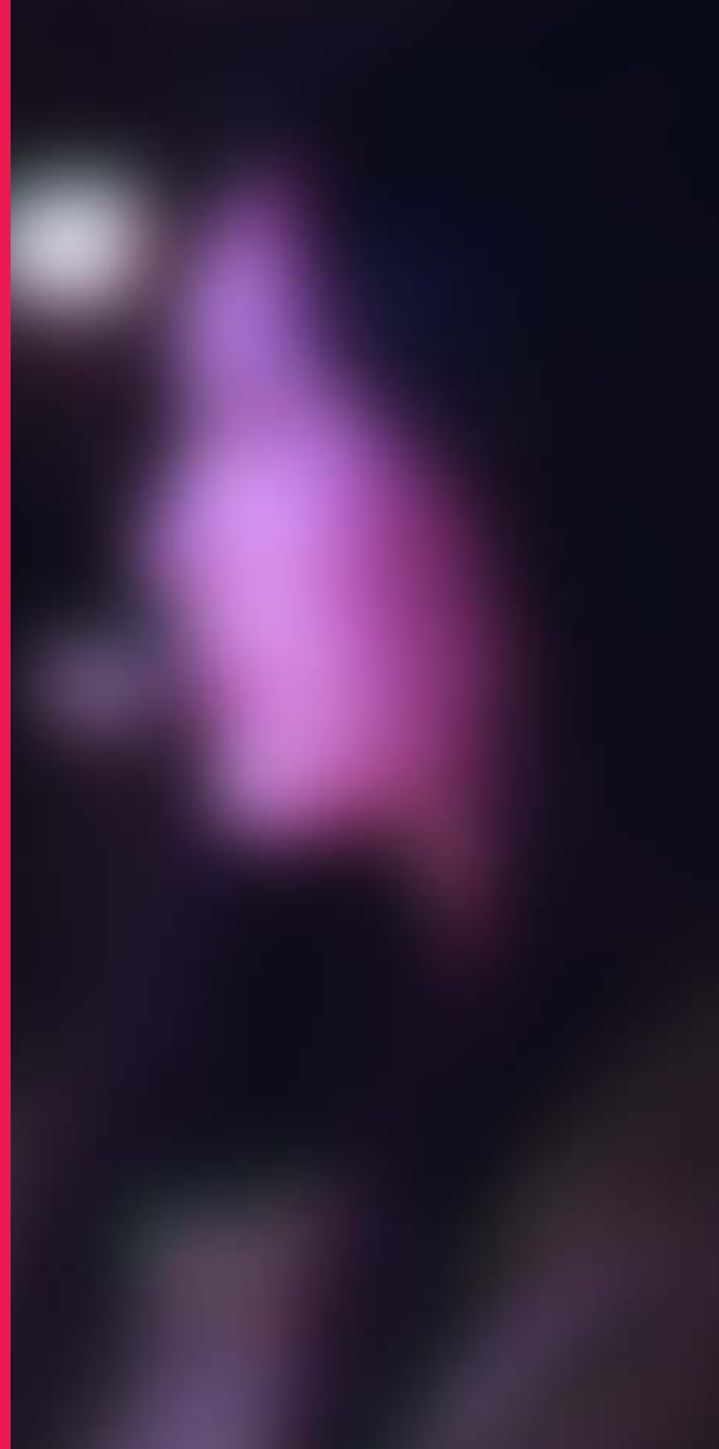
ThreatLabz es la rama de investigación de seguridad de Zscaler. Este equipo de primer nivel es responsable de la búsqueda de nuevas amenazas y de garantizar que las miles de organizaciones que utilizan la plataforma global de Zscaler estén siempre protegidas. Además de la investigación de malware y el análisis conductual, los miembros del equipo están involucrados en la investigación y el desarrollo de nuevos módulos de prototipo para la protección avanzada contra amenazas en la plataforma Zscaler, y realizan auditorías de seguridad internas regularmente para garantizar que los productos e infraestructura de Zscaler cumplan con los estándares de seguridad. ThreatLabz publica regularmente análisis detallados de amenazas nuevas y emergentes en su portal, research.zscaler.com.

Síganos en: X @ThreatLabz | Blog de investigación de seguridad de ThreatLabz

©2026 Zscaler, Inc. Todos los derechos reservados.



zscaler™ | Act Fast. Stay Secure.



Acerca de Zscaler

Zscaler (NASDAQ: ZS) acelera la transformación digital para que los clientes puedan ser más ágiles, eficientes, resistentes y seguros. Zscaler Zero Trust Exchange protege a miles de clientes de ciberataques y pérdida de datos al conectar de manera segura usuarios, dispositivos y aplicaciones en cualquier ubicación. Distribuido en más de 150 centros de datos en todo el mundo, Zero Trust Exchange basado en SSE es la plataforma de seguridad en la nube en línea más grande del mundo. Obtenga más información en **zscaler.com/mx** o síganos en Twitter **@zscaler**.

© 2026 Zscaler, Inc. Todos los derechos reservados. Zscaler™ y otras marcas comerciales listadas en **zscaler.com/mx/legal/trademarks** son (i) marcas comerciales registradas o marcas de servicio o (ii) marcas comerciales o marcas de servicio de Zscaler, Inc. en los Estados Unidos y/u otros países. Cualquier otra marca comercial es propiedad de sus respectivos propietarios.

+1 408.533.0288 Zscaler, Inc. (Oficinas centrales) • 120 Holger Way • San José, CA 95134 zscaler.com/mx