Informe de seguridad de IA de ThreatLabz 2025



© 2025 Zscaler, Inc. Todos los derechos reservados.

Índice

Resumen ejecutivo 3

Hallazgos clave 4

Tendencias de uso de IA y ML 6

- Resumen de las transacciones de IA/ML 6
- Transacciones de IA/ML bloqueadas 12
- Pérdida de datos en aplicaciones de IA/ML 13
- Uso de IA por industria 14
  - Destacados de la industria 15
- Tendencias de uso de ChatGPT 19
- Uso de IA por país 20
  - Perspectivas de EMEA 21
  - Perspectivas de APAC 22

Riesgos de la IA empresarial y escenarios de amenazas en el mundo real 23

- Riesgos principales de la adopción de IA empresarial 23
- DeepSeek y la IA de código abierto: el riesgo de tener modelos de vanguardia a su disposición 25
  - 5 indicaciones para el engaño: página de phishing generada por DeepSeek 27
- El creciente papel de la IA en las ciberamenazas 29
  - Ingeniería social supercargada 29
  - Malware y ransomware impulsados por IA en toda la cadena de ataque 30
  - IA agenética: la próxima frontera en IA autónoma y vectores de ataque 31
- Caso práctico: Cómo los actores de amenazas están explotando el interés en la IA 33

El alcance evolutivo de las regulaciones de IA 35

Predicciones de amenazas de IA para 2025–2026 37

Mejores prácticas para la adopción segura de IA empresarial 39

- 5 pasos para integrar de manera segura las herramientas de GenAI 40

Cómo Zscaler ofrece Zero Trust + IA 42

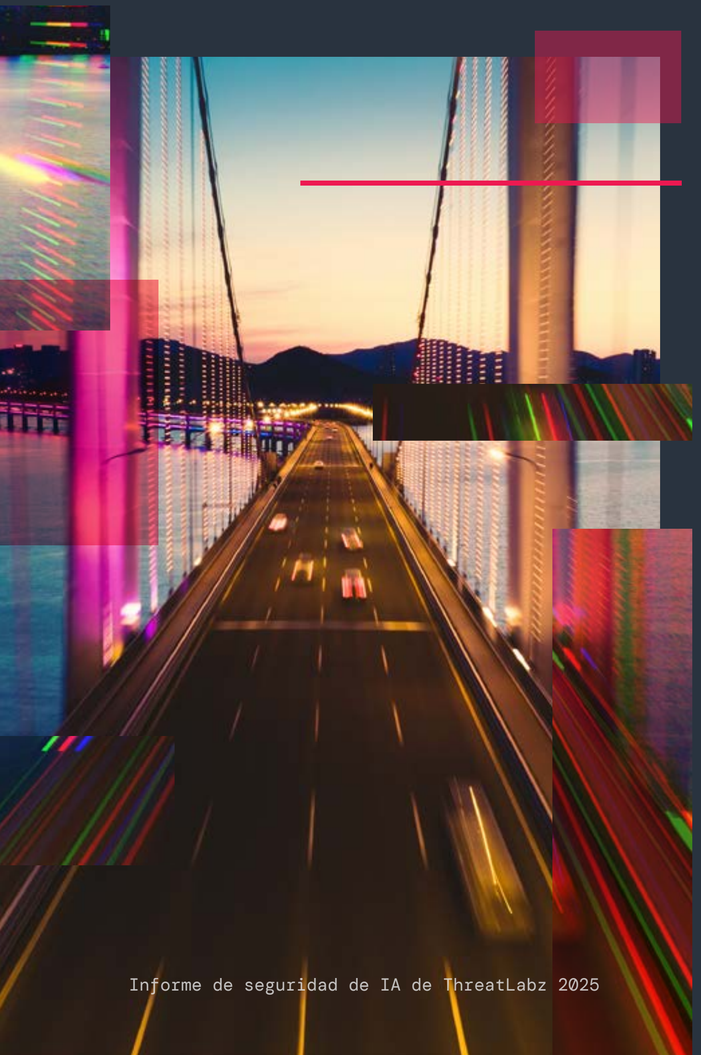
- En profundidad: la seguridad de la IA y la ventaja de los datos de Zscaler 42
- Un enfoque integral para la seguridad de la IA 43
- Aprovechar la seguridad de la IA en toda la cadena de ataque 46

Metodología de la investigación 48

- Acerca de ThreatLabz 48
- Acerca de Zscaler 48

© 2025 Zscaler, Inc. Todos los derechos reservados.

Resumen ejecutivo



Ha transcurrido otro año de la aún nueva “era de la IA”, marcado por avances que cambian las reglas del juego, una creciente adopción en todos los sectores y desafíos de gran repercusión.

Las empresas consideran ahora que la inteligencia artificial (IA) y el aprendizaje automático (ML) son esenciales para el crecimiento, ya que impulsan la eficiencia, una toma de decisiones más inteligente y una innovación más rápida. Por otro lado, la adopción de la IA conlleva graves riesgos de seguridad, desde el uso no autorizado (“IA oculta”) hasta la exposición de datos. Lo que es aún más preocupante es que los malintencionados parecen tener la ventaja, ya que utilizan estas mismas herramientas para amplificar los ataques. Lo que antes requería habilidad ahora requiere un mínimo esfuerzo. Lo que antes tomaba horas ahora toma segundos.

Este cambio se manifestó plenamente en el año 2024. GenAI se convirtió en una máquina de ingeniería social para los cibercriminales. En la actualidad, los correos electrónicos de phishing imitan a colegas de confianza con una precisión espeluznante. La tecnología Deepfake convierte voces y videos en armas de engaño.

En 2025, el poder y los peligros de la IA serán mayores que nunca. Los malintencionados continuarán ampliando los límites de las capacidades maliciosas de la IA. Sin embargo, la IA no solo posibilita ataques: ahora también es una línea de defensa fundamental que potencia la lucha contra estos ataques.

El informe de seguridad de IA de Zscaler ThreatLabz 2025 examina las múltiples facetas de la IA en la ciberseguridad, desde la adopción de la IA/ML hasta las amenazas y capacidades de seguridad impulsadas por la IA.

Al analizar 536,500 millones de transacciones capturadas a través de Zscaler Zero Trust Exchange™ de herramientas de IA/ML entre febrero y diciembre de 2024, ThreatLabz descubrió algunos cambios sorprendentes y otros menos impactantes en las tendencias de uso por parte de las empresas de todo el mundo.

ChatGPT impulsó la mayoría de las transacciones de IA/ML, constituyendo casi la mitad del volumen total. Desde una perspectiva industrial, los sectores de finanzas y seguros, junto con el de fabricación fueron los que más transacciones realizaron como principales adoptantes de la IA. Sin embargo, el aumento de la adopción no significó un acceso sin restricciones: un gran porcentaje de las transacciones de IA/ML se bloquearon activamente.

Más allá de las tendencias de uso, ThreatLabz descubrió escenarios de amenazas en el mundo real, desde el phishing potenciado por IA hasta plataformas de IA falsas. Este informe también explora los últimos avances en áreas que sin duda influirán en la IA en 2025 y más allá, como la IA agéntica, la aparición de DeepSeek y la evolución del panorama normativo.

A medida que evolucionan las capacidades de la IA/ML y crecen las amenazas que permiten, el imperativo es claro: los controles de seguridad más sofisticados y fuertes, la arquitectura Zero Trust y las defensas impulsadas por la IA ya no son opcionales, son esenciales. Siga leyendo para obtener más información y estrategias prácticas que ayuden a su organización a adoptar la IA de manera segura, y a la vez, anticiparse a las amenazas impulsadas por la IA.

© 2025 Zscaler, Inc. Todos los derechos reservados.

Hallazgos clave

ThreatLabz **analizó 536,500 millones de transacciones de IA y ML** en la nube de Zscaler entre febrero y diciembre de 2024.
Los hallazgos clave que se presentan a continuación se basan en datos que abarcan diferentes períodos de tiempo* para un análisis comparativo.

El uso de herramientas de IA/ML experimentó un aumento exponencial año tras año, con 36 veces más transacciones (+3.464.6 %) de más de 800 aplicaciones de IA/ML en la nube de Zscaler, manifiesto el aumento explosivo del interés y la dependencia de las empresas en estas tecnologías.

Las empresas bloquearon el 59.9 % de todas las transacciones de IA/ML, lo que refleja las preocupaciones en torno a la seguridad de los datos de IA y los pasos que están dando las empresas para dar forma a sus enfoques sobre la gobernanza de la IA.

ChatGPT sigue siendo la principal aplicación por volumen de transacciones, representando casi la mitad de todas las transacciones de IA/ML (45.2 %) de aplicaciones conocidas, a pesar de los debates actuales sobre sus implicaciones para la seguridad.

ChatGPT también es la aplicación de IA más bloqueada entre las aplicaciones conocidas, seguida de Grammarly, Microsoft Copilot, QuillBot y Wordtune, lo que refuerza el creciente interés y la cautela cuando se trata de asistentes de escritura y productividad impulsados por IA en entornos empresariales.

* Variaciones del período de tiempo:

- Los cambios porcentuales "interanuales" comparan los datos de abril a diciembre de 2024 y el mismo período de 2023.
- Los hallazgos específicos por país y región se basan en datos recopilados entre julio y diciembre de 2024.

El Zscaler Zero Trust Exchange rastrea las transacciones ChatGPT independientemente de otras transacciones de OpenAI en general.

© 2025 Zscaler, Inc. Todos los derechos reservados.

Las empresas están enviando volúmenes significativos de datos a herramientas de IA, con un total de 3624 TB transferidos por aplicaciones de IA/ML.

El sector de finanzas y seguros y el de fabricación son los que generan más tráfico de IA/ML, con un 28.4 % y un 21.6 % de la cuota de todas las transacciones de IA/ML en la nube de Zscaler, respectivamente, seguidos de Servicios (18.5 %), Tecnología (10.1 %), Sanidad (9.6 %) y Gobierno (4.2 %), lo que demuestra que la adopción de la IA varía significativamente entre sectores.

Los 5 países que más transacciones de IA/ML generaron fueron Estados Unidos, India, Reino Unido, Alemania y Japón.

La IA continúa amplificando los ciberriesgos, impulsada por los avances en la tecnología deepfake, los modelos emergentes de IA de código abierto y la automatización de ataques autónomos, lo que sin duda hace que las amenazas sean más adaptativas, específicas y difíciles de detectar.



© 2025 Zscaler, Inc. Todos los derechos reservados.

Tendencias de uso de IA y ML

El uso de herramientas de IA/ML aumentó globalmente en 2024, debido a que las empresas integraron la IA en las operaciones y los empleados la incorporaron en los flujos de trabajo diarios. Zscaler realizó un seguimiento de más de 800 aplicaciones de IA/ML en la nube de Zscaler, un número considerablemente mayor en comparación con el período de análisis anterior en 2023, lo que refleja la creciente adopción y confianza de las empresas en las herramientas impulsadas por IA.

Visión general de las transacciones de IA/ML

Los crecientes riesgos de seguridad no han frenado el aumento exponencial de las transacciones de IA y ML. Entre febrero y diciembre de 2024, los volúmenes de transacciones aumentaron de 3,700 millones a 49,000 millones, lo que supone un incremento de doce veces. La actividad de la IA/ML alcanzó su punto álgido en julio, con 82,700 millones de transacciones.

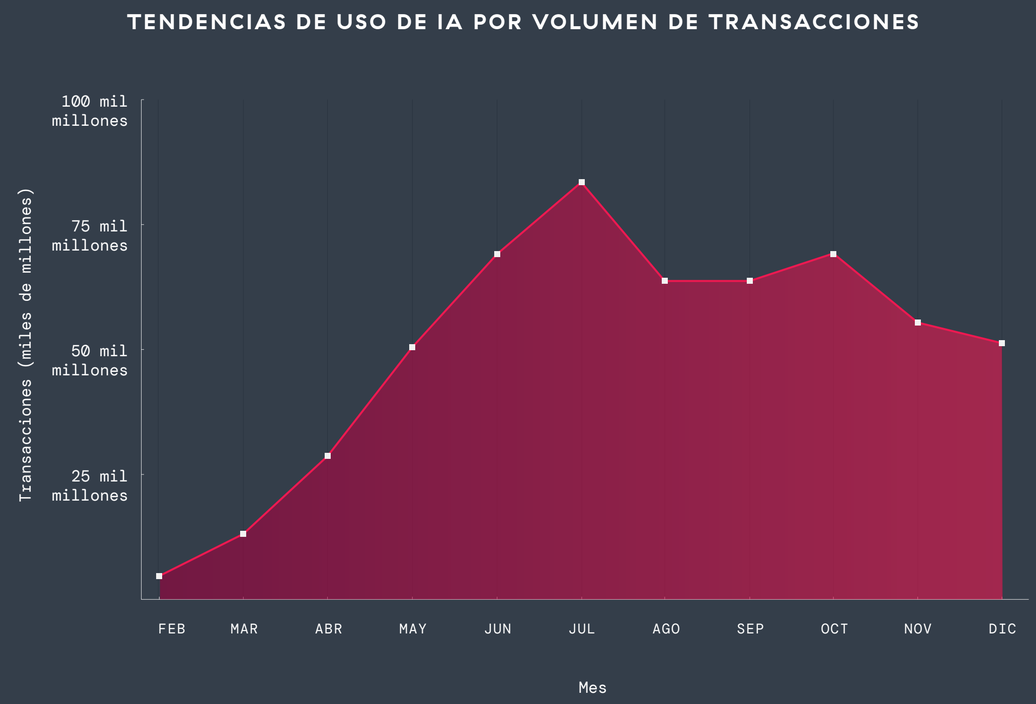


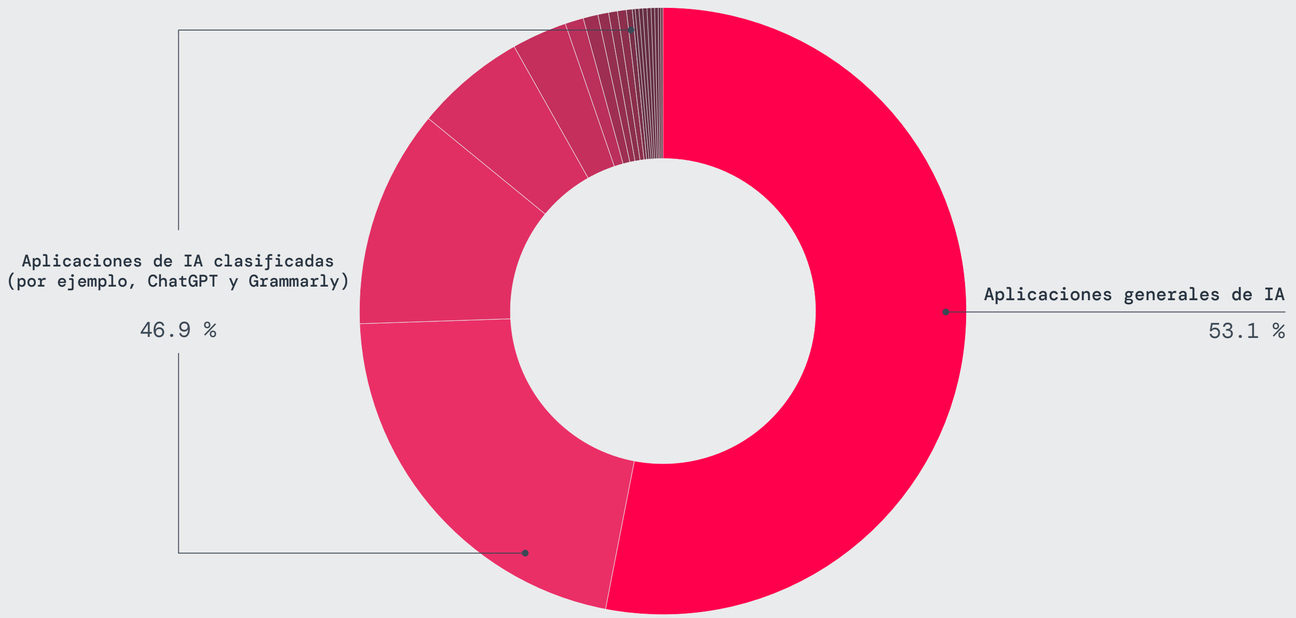
Figura 1: Transacciones de IA de febrero a diciembre de 2024

© 2025 Zscaler, Inc. Todos los derechos reservados.

La escala de la actividad de IA/ML aumentó drásticamente hasta alcanzar los 536,500 millones de transacciones totales de IA/ML, un aumento interanual del 3,464.6 % en comparación con nuestro último período de análisis. Una parte significativa de este tráfico de IA/ML procede de aplicaciones muy utilizadas como ChatGPT, Grammarly, Microsoft Copilot y otras herramientas de IA/ML. Sin embargo, una gran parte de las transacciones **(53.1 %)** permanecen categorizadas como “Aplicaciones generales de IA” dentro de la nube de Zscaler, lo que destaca la rápida proliferación del uso de la IA en las empresas. Esta clasificación refleja las transacciones de IA/ML que aún no pertenecen a aplicaciones de IA definidas, pero que, no obstante, se detectan como tráfico de IA/ML a través de la categorización de URL impulsada por IA/ML de Zscaler, que puede analizar texto, imágenes y otros contenidos para identificar la actividad relacionada con la IA.

Para ofrecer una visión más precisa y detallada de los patrones de adopción de la IA/ML entre las empresas, el análisis de ThreatLabz se centra en las aplicaciones de IA/ML clasificadas. Adoptando este enfoque, destacamos las tendencias de adopción de la IA a través de aplicaciones empresariales establecidas de IA/ML.

PROPORCIÓN DEL TOTAL DE TRANSACCIONES



© 2025 Zscaler, Inc. Todos los derechos reservados.

Entre las aplicaciones AI/ML conocidas, unas cuantas herramientas líderes del mercado generan la mayoría de las transacciones. Las siguientes cinco herramientas principales comparten un enfoque común en mejorar la productividad, la comunicación y la automatización.

- **ChatGPT** representa casi la mitad de las transacciones de IA y ML (45.2 %), Esto nos recuerda su amplia adopción en diferentes industrias. Más **información en la sección de tendencias de uso de ChatGPT.**
- **Grammarly** ocupa el segundo lugar (24.8 %), lo que refleja su creciente popularidad entre los usuarios empresariales para perfeccionar la escritura y la gramática.
- **Microsoft Copilot** ocupa el tercer puesto (12.5 %) ya que las empresas confían en él para automatizar tareas en aplicaciones de Microsoft 365 como Word, Excel y Outlook.
- **DeepL**, una herramienta líder de traducción impulsada por la IA, le sigue (6.4 %), ya que ha ganado adeptos entre las empresas globales que buscan una comunicación multilingüe de alta calidad.
- **QuillBot** completa los cinco primeros puestos (2.0 %) como otro asistente de escritura que ofrece parafraseo y resumen.

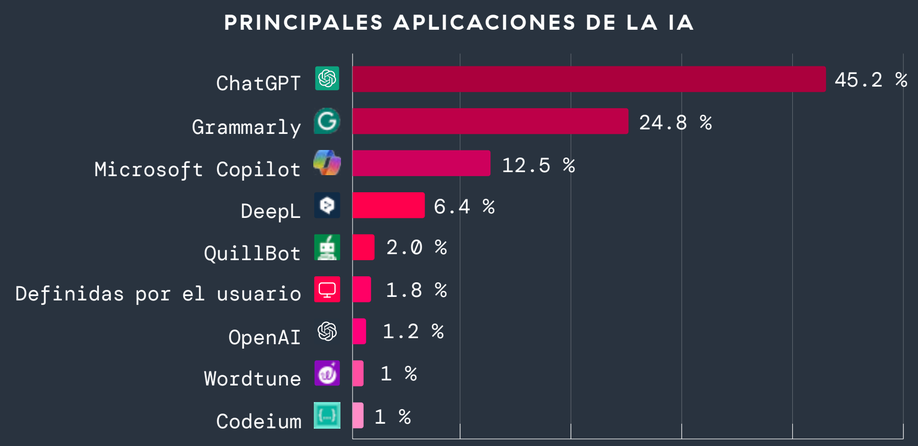


Figura 2: Principales aplicaciones de IA por volumen de transacciones

LAS 20 PRINCIPALES APLICACIONES DE IA/ML POR VOLUMEN DE TRANSACCIONES

Aplicación	Transacciones totales
ChatGPT	113,869,583,355
Grammarly	62,490,051,574
Microsoft Copilot	31,551,774,637
DeepL	16,012,344,908
QuillBot	5,130,879,211
Aplicaciones personalizadas	4,297,439,333
OpenAI	2,995,303,521
Wordtune	2,552,030,384
Codeium	2,439,268,698
Perplexity	1,806,093,093
Loom	662,917,153
Zineone	571,034,336
Synthesia	570,918,959
Writer	512,811,065
Poe	433,139,217
Claude	379,841,841
Google Gemini	317,583,902
Otter.ai	310,594,881
Runway	256,927,467
Yellow Messenger	245,412,258

© 2025 Zscaler, Inc. Todos los derechos reservados.

Principales categorías de aplicaciones

1. Asistentes de productividad (60.4 %)

Ejemplos: ChatGPT, Microsoft Copilot, Perplexity

Casi dos tercios de las transacciones de IA/ML en la nube de Zscaler entran en la categoría de asistentes impulsados por la IA. Estas aplicaciones abarcan una amplia variedad de casos de uso, desde interfaces de chat impulsadas por la IA y herramientas de investigación hasta la automatización de flujos de trabajo y la integración empresarial, todas ellas con el objetivo común de impulsar la productividad de la empresa.

2. Redacción y generación de contenidos (28.3 %)

Ejemplos: Grammarly, Quillbot, Wordtune
La segunda mayor parte de la actividad de las aplicaciones de IA/ML procede de la categoría de redacción y generación de contenidos. Las herramientas de escritura impulsadas por la IA se han convertido rápidamente en parte integral de los contenidos y las comunicaciones empresariales, agilizando tareas como la edición, la claridad y otros refinamientos gramaticales.

De la productividad al predicamento: conozca los riesgos

El papel dominante de la IA en la productividad y la escritura de las empresas presenta riesgos significativos, como la fuga de datos, los ataques de inyección puntual, las infracciones de la normativa, las alucinaciones de la IA, la exposición a la propiedad intelectual, los problemas de privacidad y la posible dependencia excesiva. Aprenda a mitigar estos riesgos y a adoptar la IA de manera segura en la sección, **“Mejores prácticas para la adopción segura de la IA en la empresa”**.

3. Idiomas y traducción (5.8 %)

Ejemplos: DeepL, LanguageTool

Las herramientas de traducción e idiomas impulsadas por inteligencia artificial representan 14,600 millones de transacciones. Estas soluciones están agilizando las comunicaciones comerciales globales y permitiendo una creación de contenido multilingüe más rápida y escalable, aunque persisten las preocupaciones sobre la precisión y la privacidad de los datos.

4. Aplicaciones personalizadas (1.7 %)

Dado que las organizaciones buscan ventajas competitivas impulsadas por la IA, las aplicaciones de IA personalizadas representan más de 4,000 millones de transacciones. Las empresas están aprovechando las soluciones de IA a medida para casos de uso que abarcan el análisis predictivo, la detección de fraudes, la automatización y mucho más.

5. Asistentes de codificación (1.3 %)

Ejemplos: Codeium, Claude

Los asistentes de codificación impulsados por IA son cada vez más comunes en el desarrollo de software, impulsando más de 3 mil millones de transacciones. Ayudan a los desarrolladores a trabajar más rápido, pero las empresas deben ser conscientes de los riesgos, desde los problemas de calidad hasta los de propiedad intelectual.

6. Herramientas visuales y creativas (1.1 %)

Ejemplos: Telar, Synthesia

El rol de la IA como socio creativo se está ampliando, ya que las herramientas visuales y creativas de la IA generaron 2700 millones de transacciones. Las herramientas de creación de video lideran la categoría, permitiendo a las organizaciones escalar los esfuerzos de producción de contenidos de video y la emisión de contenidos.

© 2025 Zscaler, Inc. Todos los derechos reservados.

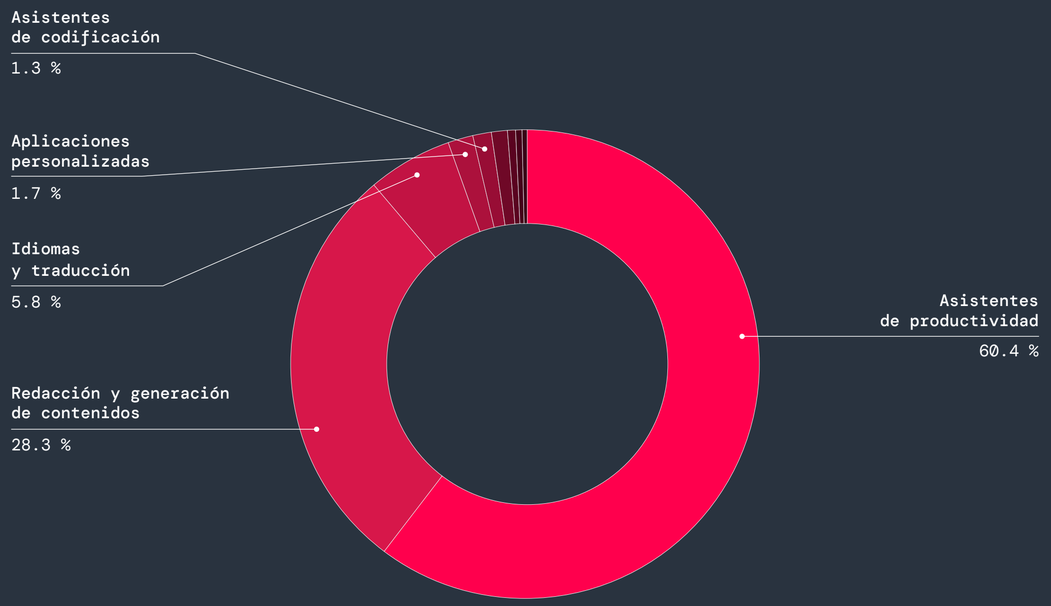


Figura 3: Transacciones por categoría de aplicación

TRANSACCIONES POR CATEGORÍA DE APLICACIÓN

Categoría	Transacciones
Asistentes de productividad	70,916,692,869
Redacción y generación de contenidos	14,638,307,672
Idiomas y traducción	31,551,774,637
Aplicaciones personalizadas	4,354,146,062
Asistentes de codificación	3,205,630,565
Herramientas visuales/creativas	4,297,439,333
Análisis de datos y automatización	2,723,874,910
Atención al cliente y chatbots	1,172,151,320
Transcripción	354,967,757
Motor de búsqueda	297,174,973
Herramientas de voz y audio	191,295,786

© 2025 Zscaler, Inc. Todos los derechos reservados.

Los volúmenes de transacciones por sí solos no reflejan plenamente la realidad del uso de la IA en las empresas. ThreatLabz también analizó la cantidad de datos transferidos entre las empresas y las herramientas de IA, que ascendió a un total de 3624 terabytes (TB). Según esta medida, ChatGPT sigue siendo la aplicación líder, con 1481 TB de datos transferidos. El gran volumen de datos muestra que las empresas no solo utilizan ChatGPT con frecuencia, sino a gran escala.

Siguiendo a ChatGPT en volumen de transferencia de datos, Grammarly, OpenAI y Microsoft Copilot ocupan los primeros puestos, lo que destaca su papel en el perfeccionamiento de contenidos impulsado por la IA y el entrenamiento de modelos.

Otras herramientas notables que contribuyen con importantes volúmenes de transferencia de datos son DeepL, Synthesia y Wordtune, cada una de las cuales da soporte a diversas necesidades empresariales, desde mejoras de la productividad hasta mensajería de video impulsada por la IA.

Vigilar tanto el volumen de transacciones como las tendencias en la transferencia de datos será clave para integrar la IA de manera eficaz y anticiparse a los posibles riesgos.

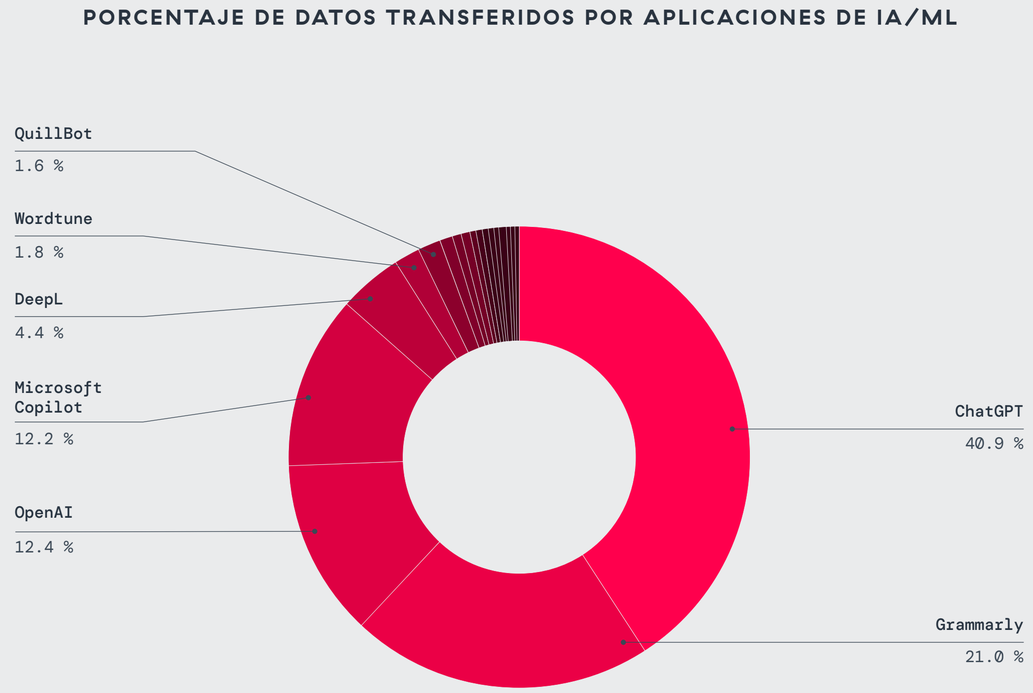


Figura 4: Principales aplicaciones de IA/ML según el porcentaje de datos totales transferidos

© 2025 Zscaler, Inc. Todos los derechos reservados.

Transacciones IA/ML bloqueadas

El crecimiento de la IA empresarial también está encontrando resistencia a medida que las organizaciones refuerzan los controles para mitigar la seguridad de los datos, la privacidad y los riesgos de cumplimiento. Actualmente, las empresas bloquean el 59.9 % de todas las transacciones de IA/ML en la nube de Zscaler, con un total de más de 321,900 millones de transacciones bloqueadas entre febrero y diciembre de 2024.

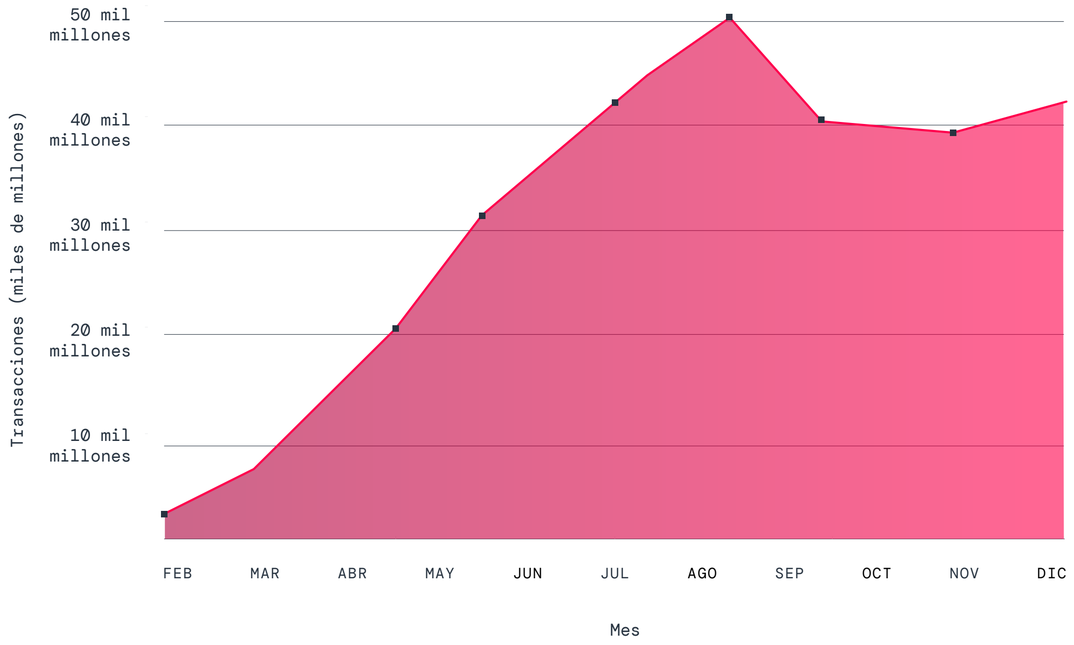


Figura 5: Número de transacciones de IA/ML bloqueadas entre febrero y diciembre de 2024

Curiosamente, las herramientas de IA más utilizadas son también las que se bloquean con más frecuencia, empezando por ChatGPT. El chatbot de GenAI sigue siendo uno de los principales objetivos de las medidas de seguridad para evitar la pérdida de datos, con un 54 % del total de bloqueos.

Adobe.io, la plataforma para desarrolladores basada en la nube de Adobe que proporciona API y herramientas de automatización impulsadas por la IA para los productos de Adobe, representa el 68 % de todas las transacciones de dominios de IA y ML bloqueadas. Esta tendencia señala un esfuerzo proactivo por parte de las empresas para evitar las transferencias de datos no autorizadas y proteger los contenidos patentados.

Las empresas andan por una cuerda floja, cada vez más estrecha, entre la innovación en IA y la seguridad. A medida que la adopción de la IA siga creciendo, las organizaciones tendrán que controlar mejor los riesgos sin dejar de aprovechar la potencia de la IA/ML para mantener la competitividad.

Principales aplicaciones de IA bloqueadas	Dominios de IA más bloqueados
1.ChatGPT	adobe.io
2.Grammarly	chatgpt.com
3.Microsoft Copilot	grammarly.com
4.QuillBot	microsoft.com
5.Wordtune	Quillbot.com
6.Codeio	deepl.com
7.DeepL	openai.com
8.Drift	bing.com
9.Poe	Wordtune.com
10.Securiti	Codeium.com

© 2025 Zscaler, Inc. Todos los derechos reservados.

Pérdida de datos en aplicaciones de IA/ML

A medida que aumenta la actividad de la IA/ML en la empresa, también lo hace el riesgo de exposición de los datos Los asistentes de productividad y chatbots impulsados por la IA, los asistentes de código y los analizadores de documentos pueden exponer inadvertidamente datos confidenciales de la empresa. Este desafío se agrava porque los usuarios comparten, sin saberlo, información confidencial con modelos de IA que carecen de controles de seguridad de nivel empresarial.

Numerosas herramientas AI/ML han sido señaladas por violaciones de prevención de pérdida de datos (DLP) en la nube de Zscaler. Estas violaciones representan casos en los que los datos confidenciales de la empresa (como datos financieros, información personal, código fuente y datos médicos) estaban destinados a ser enviados a una aplicación de IA, y esa transacción fue bloqueada por la política de Zscaler. La pérdida de datos se habría producido en estas aplicaciones de IA sin la aplicación de DLP de Zscaler. Como resultado, las violaciones sirven como un indicador clave de las tendencias de pérdida de datos de IA en el mundo real.

APLICACIONES DE IA/ML CON MÁS VIOLACIONES DE LA POLÍTICA DE DLP

Aplicación	Violaciones de la DLP
ChatGPT	2,915,502
Wordtune	879,131
Microsoft Copilot	257,869
DeepL	68,916
Codeium	41,041
Claude	40,993
Synthesia	22,975
Grammarly	7,157
DataRobot	5,440
QuillBot	4,649
Google Gemini	4,227
You.com	2,341
Perplexity	2,129
DeepAI	1,472
Poe	1,399

Estas herramientas comparten un perfil de riesgo común debido a su procesamiento basado en la nube y a su uso en flujos de trabajo de productividad, donde a menudo manejan datos confidenciales de la empresa. Las violaciones destacan la creciente necesidad de controles de DLP conscientes de la IA para garantizar que las organizaciones puedan adoptar la IA de manera segura al tiempo que evitan las fuga de datos.

Un análisis más detallado de las violaciones más comunes de la DLP relacionadas con la IA revela que la información personal identificable, el código fuente propietario y los datos relacionados con la atención sanitaria corren el riesgo de quedar expuestos.

LAS 10 PRINCIPALES VIOLACIONES DE LA DLP CON IA

1	Número de seguridad social
2	Fuga de nombres (EE. UU.)
3	Contenido para adultos
4	Contenidos sobre autolesión y ciberacoso
5	Código fuente
6	Fuga de enfermedades
7	Médicas
8	Fuga de nombres (Canadá)
9	Identificación del Registro de Contribuyentes Particulares de Brasil
10	Fuga de medicamentos

El análisis de las violaciones de la DLP relacionadas con ChatGPT y Microsoft Copilot (dos de las herramientas de IA empresarial más utilizadas y principales infractoras de la DLP) revela la frecuente exposición de información de identificación personal, datos relacionados con la salud y código fuente.

Violaciones de DLP de ChatGPT	Violaciones de DLP de Microsoft Copilot
Fuga de nombres, número de seguridad social (EE. UU.), fuga de enfermedades, fuga de nombres (Canadá), Identificación del Registro de Contribuyentes Particulares de Brasil	Número de seguridad social, fuga de medicamentos, fuga de enfermedades, fuga de tratamientos, finanzas, código fuente

Para obtener más información sobre los patrones de uso de ChatGPT, **consulte la sección “Tendencias de uso de ChatGPT”**.
Para saber cómo mitigar la pérdida de datos de las aplicaciones GenAI, lea **las herramientas GenAI de manera segura** a continuación.

© 2025 Zscaler, Inc. Todos los derechos reservados.

Uso de IA por industria

La adopción empresarial de herramientas de IA y ML varía ampliamente según el sector, con **Finanzas y Seguros** a la cabeza, impulsando el **28.4 %** de las transacciones de IA/ML. A medida que los servicios financieros continúan adoptando eficiencias impulsadas por la IA para funciones críticas como la detección de fraudes, la automatización del servicio al cliente y las evaluaciones de riesgos, su volumen de transacciones de IA ha superado al sector de **fabricación,** que ahora ocupa el segundo lugar con el **21.6 %** del total de transacciones de IA/ML.

Le siguen los sectores de **servicios (18.5 %), tecnología (10.1 %)** y **sanitario (9.6 %)**, cada uno de los cuales adopta la IA a diferentes velocidades en función de sus prioridades operativas únicas. Mientras que es probable que el sector de servicios esté acelerando el uso de la IA en términos de atención al cliente y optimizaciones operativas, las empresas tecnológicas siguen impulsando la investigación y la innovación en IA. En comparación, la adopción en el sector sanitario sigue siendo menor, lo que refleja una postura más reservada debido a las mayores preocupaciones en materia de regulación y seguridad.

PROPORCIÓN DE TRANSACCIONES DE IA POR SECTOR VERTICAL

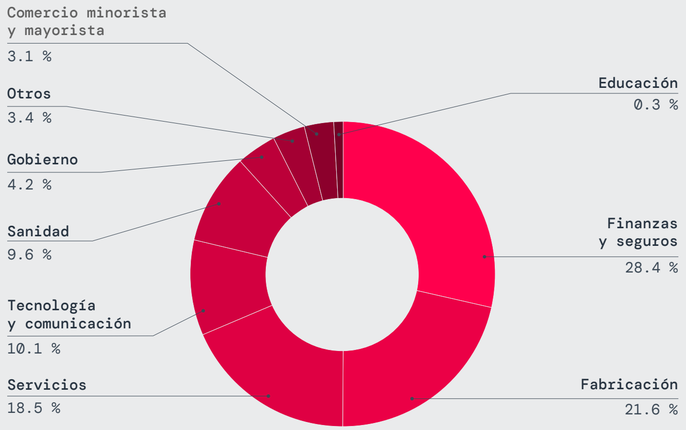


Figura 6: Industrias que impulsan las mayores proporciones de transacciones de IA

TENDENCIAS DE TRANSACCIONES DE IA POR VERTICAL

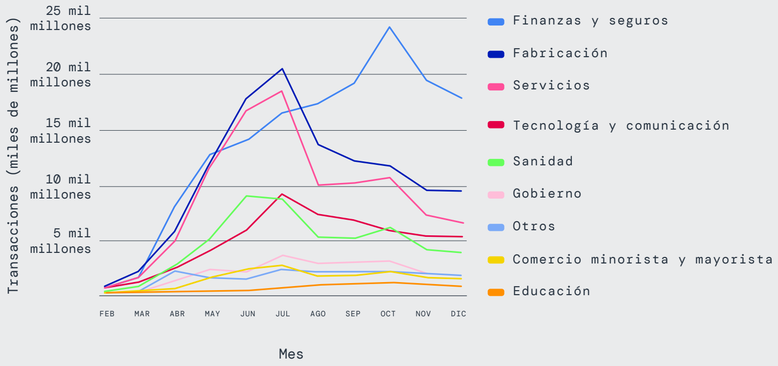


Figura 7: Tendencias de las transacciones de IA/ML entre las industrias de mayor volumen

Las industrias también están redoblando sus esfuerzos para asegurar las transacciones de IA/ML, pero el volumen de actividad de IA/ML bloqueada varía. El sector de las finanzas y seguros bloquea el 39.5 % de las transacciones de IA. Esta tendencia se alinea con el estricto panorama de cumplimiento de la industria y la necesidad de salvaguardar los datos financieros y personales.

El sector de fabricación bloquea el 19.2 % de las transacciones de IA, lo que sugiere un enfoque estratégico en el que la IA se utiliza ampliamente, pero se vigila de cerca en busca de riesgos de seguridad, mientras que el sector servicios adopta un enfoque más equilibrado, bloqueando el 15 % de las transacciones de IA. Por otro lado, el sector sanitario bloquea solo el 10.8 % de las transacciones de IA. A pesar de manejar grandes cantidades de datos sanitarios y de información personal de identificación, las organizaciones sanitarias siguen rezagadas en la protección de las herramientas de IA, mientras que los equipos de seguridad se ponen al día con la vertiginosa innovación. Esta tendencia destaca el retraso de las medidas de protección, lo que hace que las transacciones globales de IA en el sector sanitario sigan siendo relativamente bajas en comparación con otros sectores.

PORCENTAJE DE TRANSACCIONES DE IA BLOQUEADAS POR SECTOR

Vertical	% de transacciones de IA bloqueadas
Finanzas y seguros	39.5 %
Fabricación	19.2 %
Servicios	15.0 %
Sanitario	10.8 %
Tecnología y comunicación	6.9 %
Gobierno	4.5 %
Otros	2.2 %
Comercio minorista y mayorista	1.6 %
Educación	0.3 %

© 2025 Zscaler, Inc. Todos los derechos reservados.

Destacados de la industria

El sector financiero y de seguros redobla la inversión en IA

LAS 5 PRINCIPALES APLICACIONES DE INTELIGENCIA ARTIFICIAL EN FINANZAS Y SEGUROS

1	2	3	4	5
ChatGPT	Microsoft Copiloto	Grammarly	Definidas por el usuario Aplicaciones	DeepL

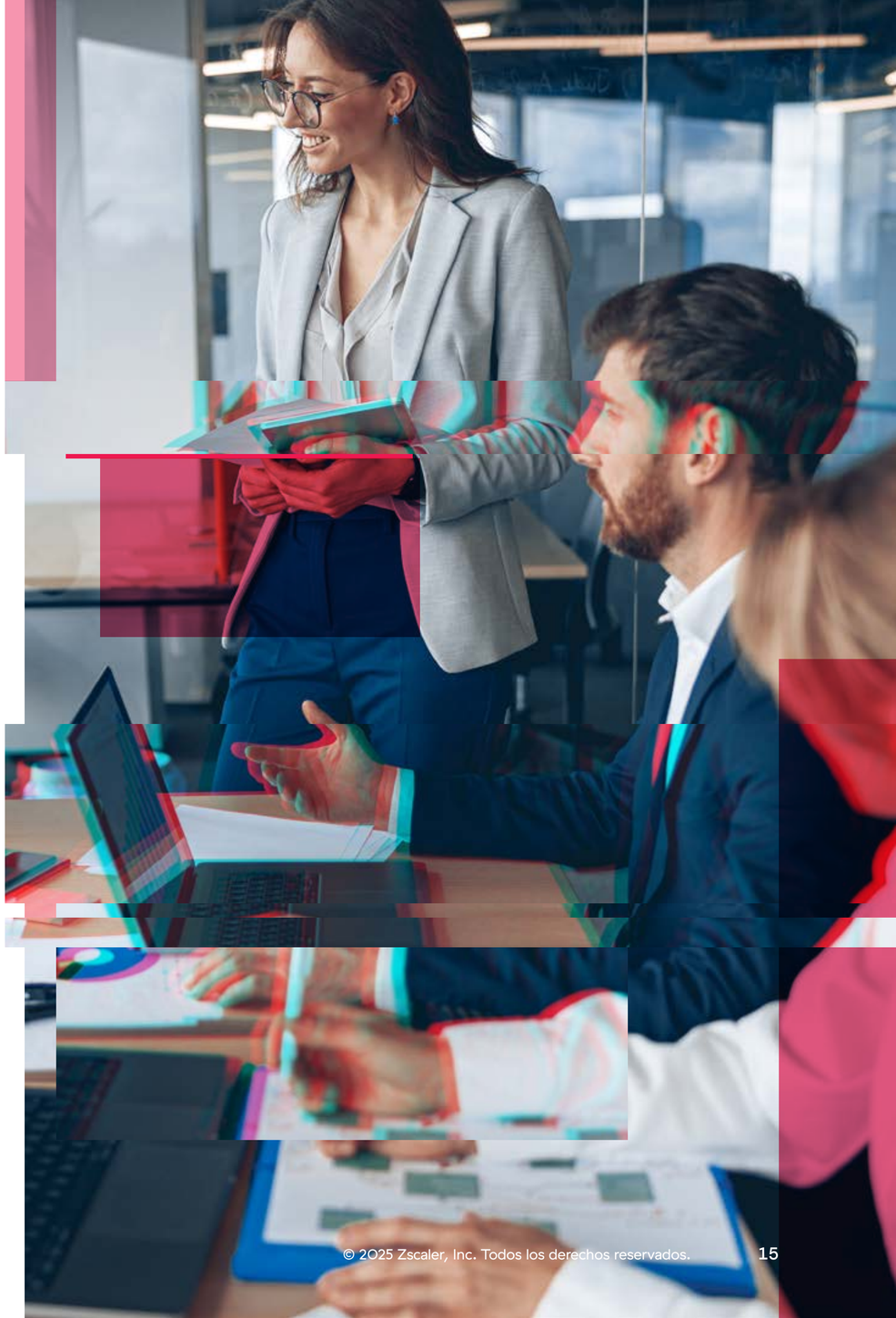
Como principal impulsor de las transacciones de IA/ML en la nube Zscaler (152,400 millones), el sector de finanzas y seguros está profundamente volcado en el potencial de la IA. Estas industrias confían en la IA para analizar transacciones financieras en tiempo real, detectar actividades fraudulentas y acelerar el procesamiento de reclamaciones, por nombrar algunas tareas críticas que les ayudan a ahorrar tiempo y dinero.

Más allá de la automatización, la IA generativa está remodelando las operaciones financieras. Herramientas como ChatGPT y Microsoft Copilot, que se encuentran entre las aplicaciones más utilizadas por las empresas de finanzas y seguros en la nube de Zscaler, ayudan a las instituciones financieras a resumir informes, automatizar flujos de trabajo y ayudar en las tareas de cumplimiento. Las aplicaciones de IA personalizadas también se encuentran entre las cinco primeras para las organizaciones de servicios financieros, lo que destaca una profunda inversión en soluciones impulsadas por la IA. Mientras tanto, el elevado volumen de transacciones de DeepL sugiere una creciente necesidad de traducción asistida por IA en las finanzas mundiales.

A medida que las organizaciones de finanzas y seguros integran cada vez más la IA, se enfrentan a mayores desafíos relacionados con la seguridad, el cumplimiento de la normativa y las cuestiones éticas, como la privacidad de los datos, la parcialidad y la precisión. Los bots impulsados por la IA representan ahora una parte significativa de las transacciones bloqueadas, explotando las API y los flujos de trabajo de autenticación para eludir los controles de seguridad.

Para contrarrestar estas amenazas, las organizaciones están adoptando cada vez más modelos de seguridad basados en la IA para la detección de anomalías en el comportamiento y la autenticación adaptativa basada en el riesgo. Sin embargo, las técnicas de IA adversaria siguen evolucionando, lo que requiere una supervisión continua y estrategias Zero Trust avanzadas para mitigar los riesgos emergentes.

Al priorizar la supervisión y el uso ético de la IA, las instituciones financieras pueden salvaguardar la integridad de los datos, garantizar la equidad y mantener la confianza del público en la banca, los seguros y otros sectores financieros.



© 2025 Zscaler, Inc. Todos los derechos reservados.



El sector de fabricación está aprovechando el poder de la IA

LAS 5 PRINCIPALES APLICACIONES DE IA EN EL SECTOR DE FABRICACIÓN

1	2	3	4	5
ChatGPT	Grammarly	Microsoft Copiloto	DeepL	QuillBot

La segunda mayor parte del tráfico de IA/ML (21.6 %) en nuestra investigación procede de clientes del sector de fabricación. La adopción de la IA en esta industria es un motor clave de la cuarta revolución industrial (la Industria 4.0), que está redefiniendo el sector de fabricación con las fábricas inteligentes, los dispositivos conectados al IoT y el mantenimiento predictivo.

Los fabricantes aprovechan cada vez más la IA para mejorar sus operaciones, desde la predicción de fallas en los equipos mediante el análisis de los extensos datos de maquinaria y sensores hasta la racionalización de la gestión de la cadena de suministro, el control de inventarios y la logística. Además, la robótica y los sistemas de automatización impulsados por la IA están impulsando significativamente la eficiencia de la fabricación, realizando tareas con mayor velocidad y precisión que los trabajadores humanos, reduciendo así los costos y minimizando los errores.

Sin embargo, la seguridad de los datos sigue siendo una preocupación, ya que el sector de fabricación representa el 19.2 % del tráfico de IA/ML bloqueado, lo que indica un enfoque cauteloso hacia la adopción de la IA. Esta cautela se debe a la preocupación por la seguridad de los datos y a la necesidad de evaluar y aprobar cuidadosamente las aplicaciones de la IA, restringiendo las que plantean mayores riesgos. Por ejemplo, los fabricantes de productos electrónicos pueden aplicar protocolos estrictos para garantizar que solo se integren en sus operaciones aplicaciones de IA que cumplan normas de seguridad estrictas, mitigando así de manera eficaz las posibles vulnerabilidades.

© 2025 Zscaler, Inc. Todos los derechos reservados.

El sector sanitario registra un aumento en la actividad de IA

LAS 5 PRINCIPALES APLICACIONES DE INTELIGENCIA ARTIFICIAL EN EL SECTOR SANITARIO

1	2	3	4	5
ChatGPT	Grammarly	Microsoft Copiloto	DeepL	QuillBot

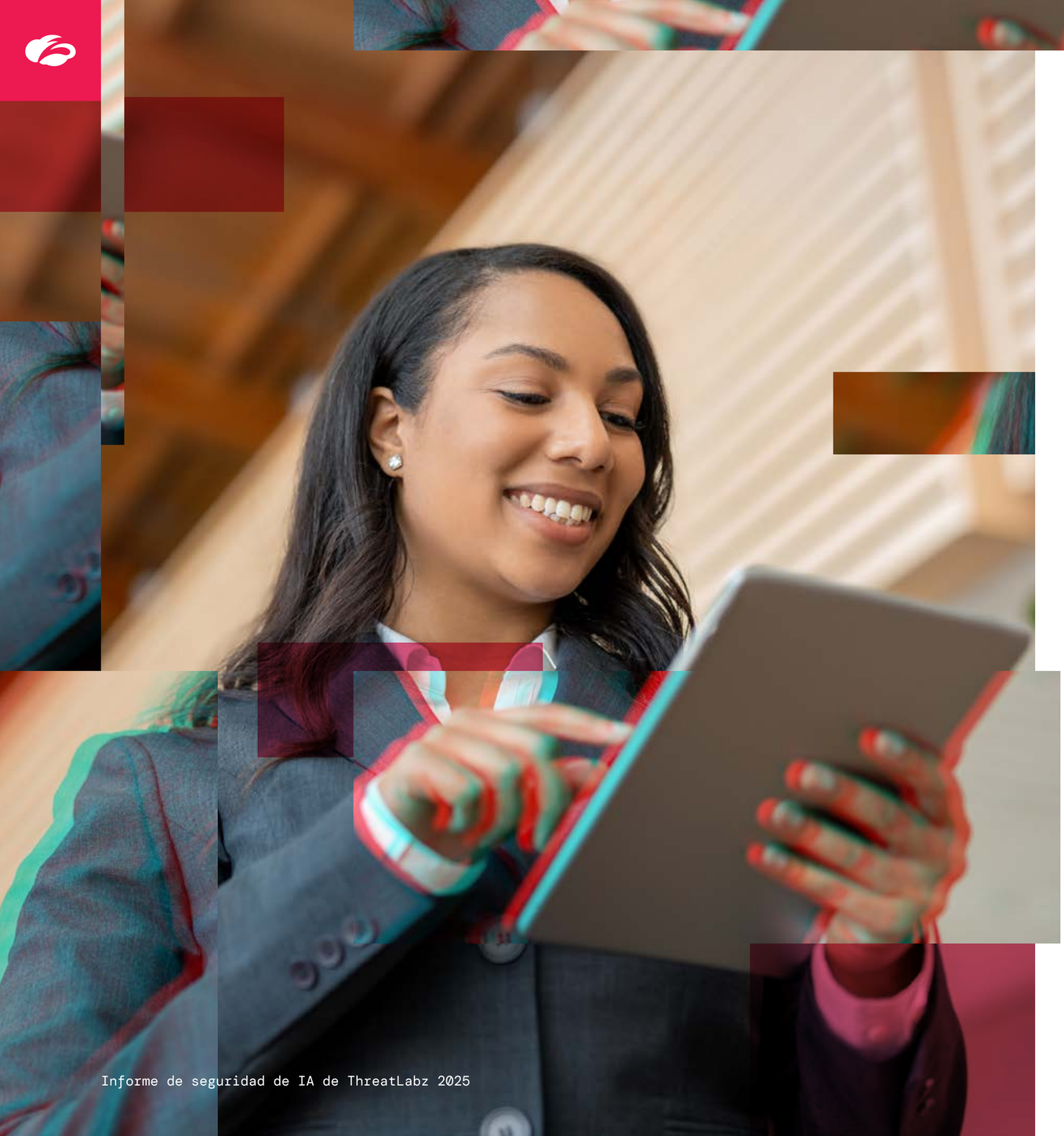
La sanidad ocupa el quinto lugar en el uso de IA/ML en la nube de Zscaler, con un 9.6 % del tráfico, un 4.1 % más que el año pasado. Sin embargo, este año, el sector sanitario solo ha bloqueado el 10.8 % de todas las transacciones de IA, lo que supone una disminución significativa con respecto al 17.23 % de 2024. Varios factores contribuyen a este cambio.

La rápida integración de las herramientas de IA/ML ha provocado un aumento de las actividades relacionadas con la IA. Aplicaciones como ChatGPT, la aplicación AI/ML más utilizada por las organizaciones sanitarias en la nube de Zscaler, ayuda a los profesionales sanitarios con el apoyo al diagnóstico, los resúmenes de investigación médica y la documentación de los pacientes. Sin embargo, es probable que el aumento de la actividad de la IA/ML dificulte la distinción entre las transacciones legítimas y las maliciosas, lo que podría dar lugar a un menor número de bloqueos. Dado que las organizaciones sanitarias dependen cada vez más de la IA para la atención a los pacientes y las tareas administrativas, cada vez se insiste más en habilitar las funcionalidades de la IA.

La IA/ML en la atención sanitaria ofrece avances significativos, pero también presenta riesgos notables. Una de las principales preocupaciones es la privacidad de los datos, ya que los sistemas de IA suelen requerir una gran cantidad de datos de los pacientes, lo que plantea problemas sobre la seguridad y la confidencialidad de la información confidencial. Además, el contenido generado por la IA a veces puede contener imprecisiones o dar lugar a posibles diagnósticos o tratamientos erróneos. Además, la creciente sofisticación de los ciberataques impulsados por IA, como las campañas de phishing generadas por IA, plantea mayores desafíos de seguridad. Por lo tanto, aunque las tecnologías de IA/ML pueden mejorar el sector sanitario y, en última instancia, la atención a los pacientes, es imprescindible aplicar medidas de seguridad sólidas y mantener la supervisión humana para mitigar estos riesgos.



© 2025 Zscaler, Inc. Todos los derechos reservados.



El gobierno reconoce el potencial de la IA

LAS 5 PRINCIPALES APLICACIONES DE IA EN EL GOBIERNO

1	2	3	4	5
Grammarly	Microsoft Copiloto	ChatGPT	QuillBot	DeepL

El uso gubernamental de la IA/ML ha aumentado hasta 4.2 % este año, impulsado por la búsqueda de una mejor prestación de servicios y una elaboración de políticas eficiente. Este aumento probablemente se pueda atribuir al potencial de la IA para agilizar las operaciones, mejorar la participación ciudadana e informar decisiones basadas en datos.

Entre las aplicaciones de IA rastreadas en la nube de Zscaler, Grammarly es la herramienta más utilizada por las entidades gubernamentales, lo que sugiere un enfoque en la mejora de las comunicaciones entre el gobierno y los ciudadanos. El uso de Microsoft Copilot, la herramienta de IA número dos para los gobiernos, señala además el interés por la automatización impulsada por la IA para obtener beneficios como la eficiencia administrativa.

Sin embargo, esta rápida integración requiere medidas de seguridad sólidas para mitigar los riesgos asociados. La privacidad de los datos es una preocupación primordial, ya que los sistemas de IA suelen requerir un amplio acceso a información confidencial, lo que aumenta el potencial de violaciones. Las vulnerabilidades de seguridad son otra cuestión crítica; los sistemas de IA pueden convertirse en objetivos de sofisticados ciberataques destinados a extraer datos confidenciales. Además, el sesgo algorítmico puede conducir a resultados injustos o discriminatorios, socavando la confianza pública. Para mitigar estos riesgos, es esencial aplicar medidas de seguridad sólidas, establecer marcos de gobernanza claros y mantener la supervisión humana durante todo el ciclo de vida.

© 2025 Zscaler, Inc. Todos los derechos reservados.

Tendencias de uso de ChatGPT

ChatGPT cumplió dos años en 2024, y su adopción por parte de las empresas y su popularidad mundial no muestran signos de disminución. Con la implementación de las capacidades de memoria y la búsqueda web en tiempo real, ChatGPT es más inteligente, más rápido y más útil que nunca, lo que impulsa aún más su adopción. Solo en la primera mitad del año, las transacciones globales de ChatGPT en la nube de Zscaler ascendieron a 90,700 millones, consolidando su lugar como la herramienta de IA generativa más utilizada.

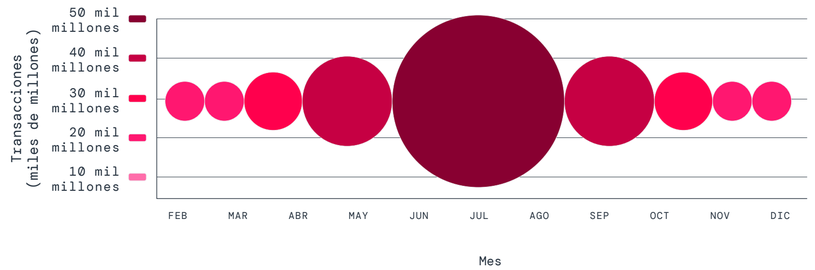


Figura 8: Transacciones ChatGPT de febrero a diciembre de 2024

Sin embargo, la adopción de ChatGPT por parte de la industria no refleja exactamente las tendencias generales de uso de la IA/ML, sino que presenta un notable valor atípico. Aunque el sector de las finanzas y los seguros impulsó el mayor volumen de transacciones globales de IA/ML, solo representa el 11.4 % del uso de ChatGPT. Es probable que esta menor tasa de adopción refleje preocupaciones más estrictas en materia de seguridad, cumplimiento y privacidad de los datos, que limitan el uso de la IA generativa en entornos regulados.

El sector de fabricación, que ocupa el segundo lugar en transacciones totales de IA, impulsa el mayor volumen de transacciones ChatGPT. Esto sugiere que los fabricantes están recurriendo a la IA generativa para todo, desde la documentación técnica hasta los flujos de trabajo automatizados. Le siguen de cerca los sectores de servicios, sanidad y tecnología, que también recurren en gran medida a ChatGPT.

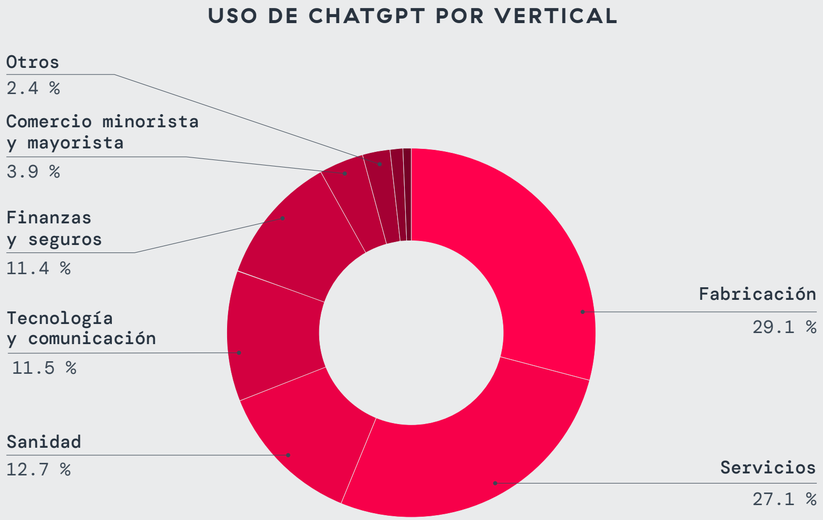


Figura 9: Sectores que registran las mayores proporciones de transacciones ChatGPT

DE CHATGPT A DEEPSEEK: LA EVOLUCIÓN DE LOS CHATBOTS DE IA

Los modelos a la vanguardia de la IA como ChatGPT (OpenAI) y Claude (Anthropic) dominan el panorama de los chatbot, representando un gran porcentaje de las transacciones de IA/ML en la nube de Zscaler. Estas aplicaciones se utilizan ampliamente en entornos empresariales para la creación de contenidos, soporte a la codificación, análisis de datos y automatización del flujo de trabajo.

Mientras que los modelos convencionales imponen medidas de seguridad hasta cierto punto, las alternativas de código abierto introducen nuevos riesgos, y ahí es donde entra en juego DeepSeek.

DeepSeek es la respuesta de China a ChatGPT. Sin embargo, a diferencia de ChatGPT que tiene restricciones de seguridad incorporadas, DeepSeek permite el acceso sin restricciones, lo que la convierte en una herramienta potente pero riesgosa. Su naturaleza de código abierto suscita preocupaciones sobre la seguridad y la soberanía de los datos, y la falta de controles de seguridad significa que las empresas y los usuarios finales deben evaluar cuidadosamente los riesgos antes de utilizar DeepSeek. Del mismo modo, Grok, desarrollado por xAI, adopta un enfoque más flexible de las interacciones de la IA, ofreciendo menos restricciones en comparación con los modelos tradicionales.

Lea más sobre el surgimiento de DeepSeek y sus riesgos en la sección de este informe sobre DeepSeek y la IA de código abierto.

© 2025 Zscaler, Inc. Todos los derechos reservados.

Uso de IA por país

El uso de la IA se está acelerando en todo el mundo, y los países están aumentando sus inversiones para impulsar la innovación y seguir siendo competitivos. Estados Unidos y la India dominan en el número de transacciones de IA/ML en la nube de Zscaler, lo que refleja un fuerte compromiso con la investigación, la infraestructura e incluso las nuevas empresas impulsadas por la IA.

Los Estados Unidos (46.2 %)impulsó el mayor número de transacciones**India (8.7 %)** surgió como el segundo mayor contribuyente.

El entorno normativo relativamente flexible de los Estados Unidos (**véase "El alcance en evolución de la normativa sobre IA"**), que fomenta la experimentación y la implementación de la IA, puede dar a las empresas estadounidenses una ventaja clave. A diferencia de las regiones con leyes de IA más estrictas, los EE. UU. ofrece una mayor flexibilidad para desarrollar e integrar las tecnologías de IA. Esto se refleja en una inversión de 13 800 millones de dólares en aplicaciones empresariales de IA por parte de las empresas en 2024, lo que supone un aumento de seis veces con respecto al año anterior.

La India sigue consolidándose como un actor importante en la carrera de la IA, con inversiones en sectores clave, como las finanzas y los seguros, la sanidad, la fabricación y los servicios gubernamentales. Con fuertes inversiones gubernamentales, como la Estrategia Nacional de IA[1], y una creciente inversión del sector privado, la India está aprovechando la IA para mejorar la automatización, el análisis y la ciberseguridad. Sin embargo, siguen existiendo desafíos (la preocupación por la privacidad de los datos, la incertidumbre normativa y la falta de especialistas en IA) que dificultan su adopción generalizada. A pesar de los rápidos avances, los países se enfrentan a barreras para la adopción de la IA. Las estrictas leyes de privacidad de datos, como el RGPD, introducen desafíos de cumplimiento, mientras que el alto costo de la implementación de la IA y la escasez de talento calificado crean obstáculos para la adopción, especialmente en los mercados emergentes. Las preocupaciones sobre la seguridad (las ciberamenazas impulsadas por la IA, el sesgo algorítmico, etc.) complican aún más la adopción y el uso. A medida que los países y los gobiernos se enfrentan a estos desafíos, un enfoque estratégico que busque combinar la claridad normativa, la inversión en educación sobre IA y marcos sólidos de ciberseguridad, será fundamental para la adopción global de la IA a gran escala.

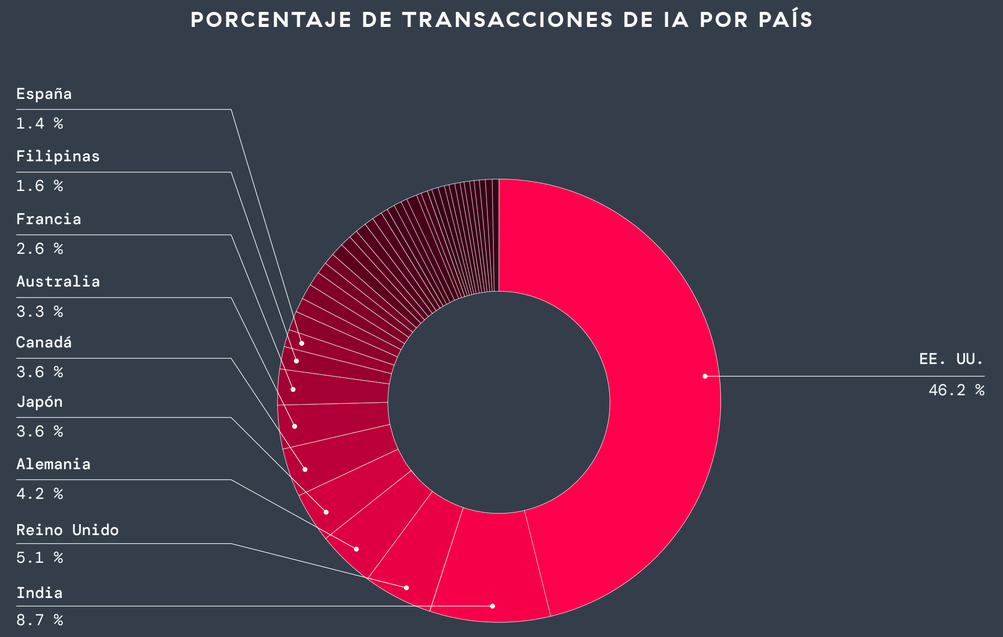


Figura 10: Países que generan las mayores proporciones de transacciones de IA

[1] Niti Aayog, National Strategy for Artificial Intelligence, consultado el 28 de febrero de 2025.

© 2025 Zscaler, Inc. Todos los derechos reservados.

Perspectivas de EMEA

Un análisis más detallado de la región de Europa, Oriente Medio y África (EMEA) revela que la mayor cantidad de transacciones de IA procede del Reino Unido (22.3 %), Alemania (18.4 %) y Francia (11.3 %). Aunque el Reino Unido solo representa el 5.1 % de las transacciones de IA a nivel mundial, vuelve a representar más del 20 % del tráfico de IA en EMEA, lo que le convierte en el líder de la región.

Alemania ha registrado un aumento interanual del número de transacciones de IA (+5.74 %), con un mayor número de empresas que invierten en tecnologías de IA. Este aumento es evidente en los sectores de fabricación y de servicios, impulsado por la necesidad de automatización y eficiencia. Francia también se ha estado posicionando como un contendiente global en IA, con 109,000 millones de euros en inversión privada por parte del presidente Emmanuel Macron en febrero de 2025.[2]

DESGLOSE POR PAÍSES DE EMEA

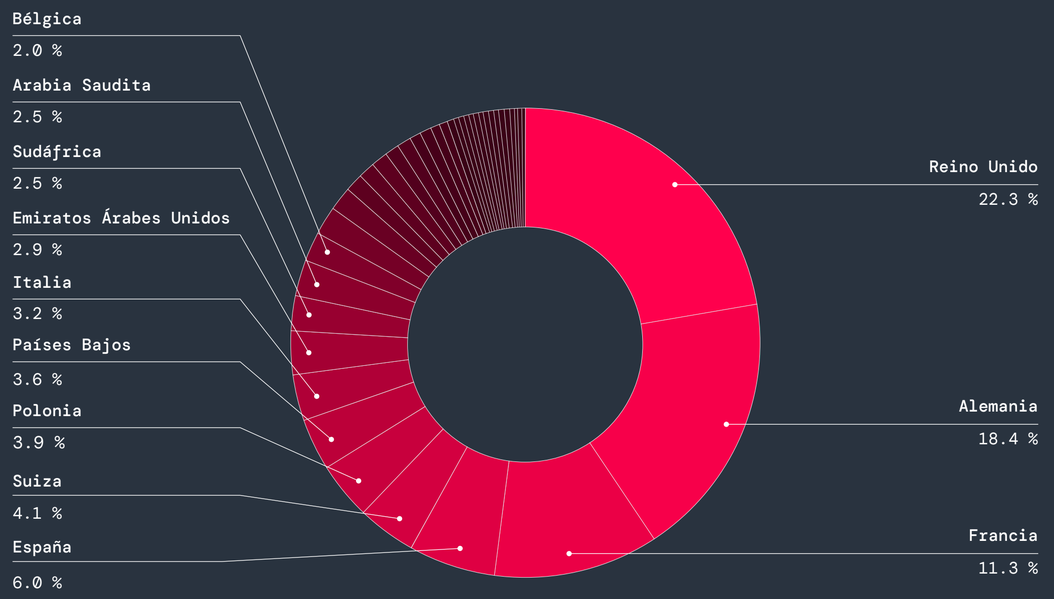


Figura 11: Porcentaje de transacciones de IA por país en la región EMEA

TRANSACCIONES DE EMEA POR MES

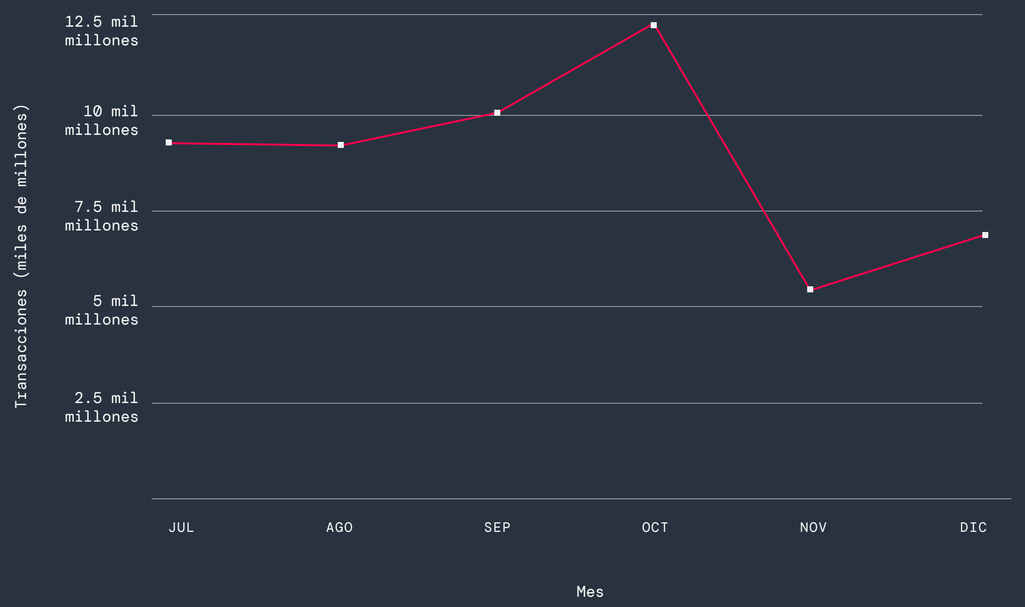


Figura 12: Transacciones de IA de julio a diciembre de 2024 en la región EMEA

[2] CNBC, France unveils 109-billion-euro AI investment as Europe looks to keep up with U.S., 10 de febrero de 2025.

© 2025 Zscaler, Inc. Todos los derechos reservados.

Perspectivas de APAC

Profundizando en la región de Asia–Pacífico (APAC), ThreatLabz observó que la mayor parte de las transacciones de IA proceden de la India (36.4 %), Japón (15.2 %), y Australia (13.6 %).

Aunque Japón ha registrado un aumento interanual de las transacciones de IA (+5.7 %), el país ha adoptado un enfoque más cauteloso con respecto a las tecnologías de IA. El uso cotidiano de la IA sigue siendo relativamente bajo debido a factores culturales[3] y a entornos normativos estrictos. Australia está desarrollando más activamente marcos para garantizar un uso responsable de la IA, con un aumento interanual del 3.6 % en las transacciones de IA. Filipinas también está experimentando una adopción acelerada de la IA, con un sector que crecerá a un ritmo anual del 41.5 % entre 2025 y 2030.[4] Sin embargo, este cambio suscita preocupación por posibles despidos, ya que es necesario mejorar la calificación de la mano de obra y realizar intervenciones políticas estratégicas para equilibrar el avance tecnológico con la estabilidad del empleo.[5]

DESGLOSE DE PAÍSES DE APAC

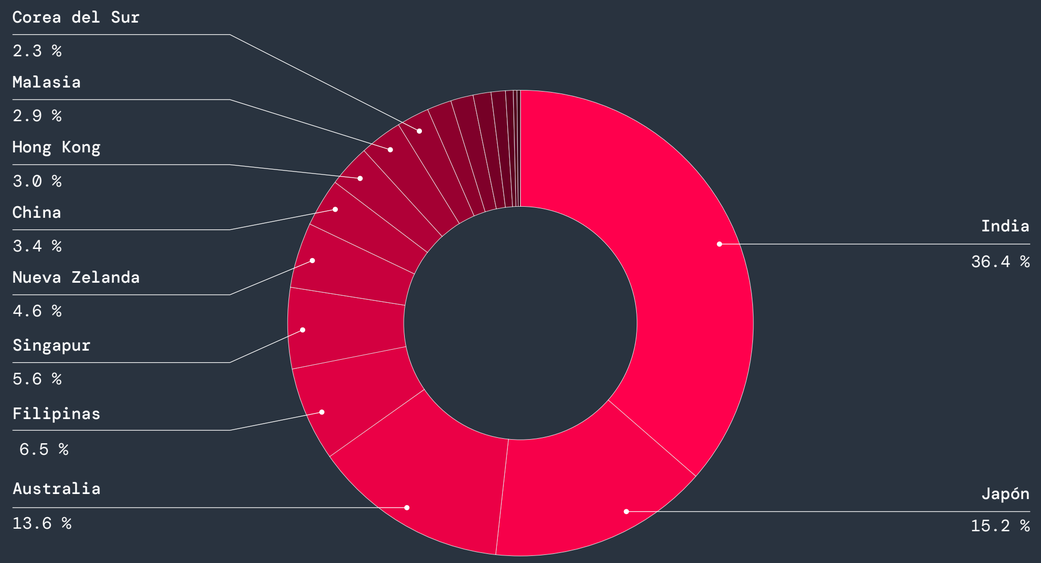


Figura 13: Porcentaje de transacciones de IA por países en la región APAC

TRANSACCIONES DE APAC POR MES

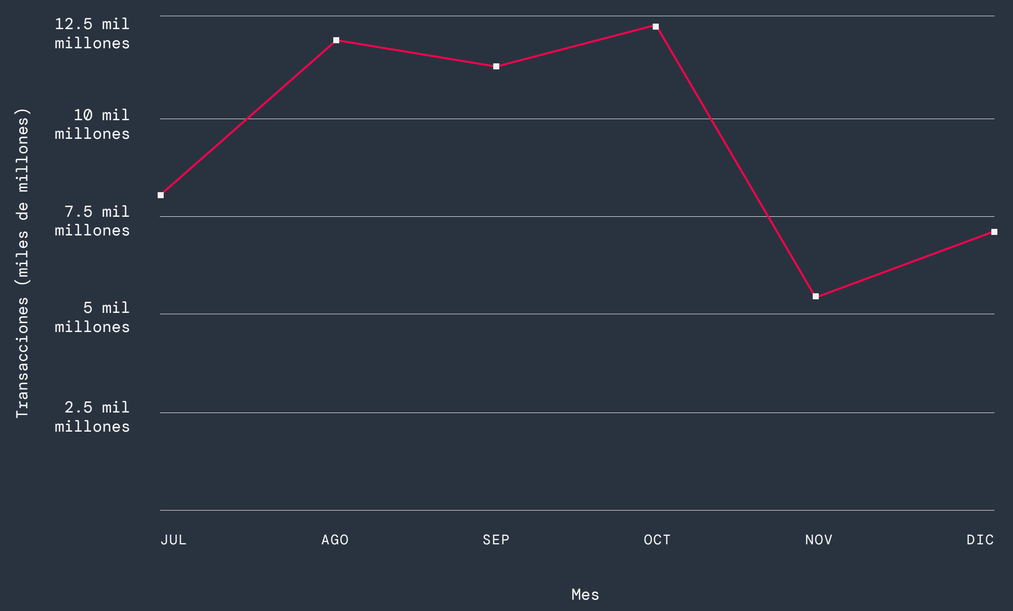


Figura 14: Transacciones de IA de julio a diciembre de 2024 en la región APAC

[3] World Economic Forum, Reconciling tradition and innovation: Japan's path to global AI leadership, 17 de diciembre de 2024.
[4] The Manila Times, AI breakthroughs PH businesses need to know, 23 de febrero de 2025
[5] Inquirer.net, IMF sees 36% of PH jobs eased or displaced by AI, 27 de diciembre de 2024.

© 2025 Zscaler, Inc. Todos los derechos reservados.

Riesgos de la IA empresarial y escenarios de amenazas en el mundo real

Riesgos principales de la adopción de la IA empresarial

Introducir la IA en su organización conlleva una mezcla de oportunidades y riesgos, muchos de los cuales aún están evolucionando. Los sistemas potenciados por la IA crean nuevas superficies de ataque, y la GenAI y los LLM son especialmente vulnerables a las amenazas que pueden manipular los resultados de la IA, introducir sesgos o fugar información confidencial. Estos son algunos de los mayores riesgos que deben afrontar las empresas.

Problemas de calidad de los datos (“basura que entra, basura que sale”)

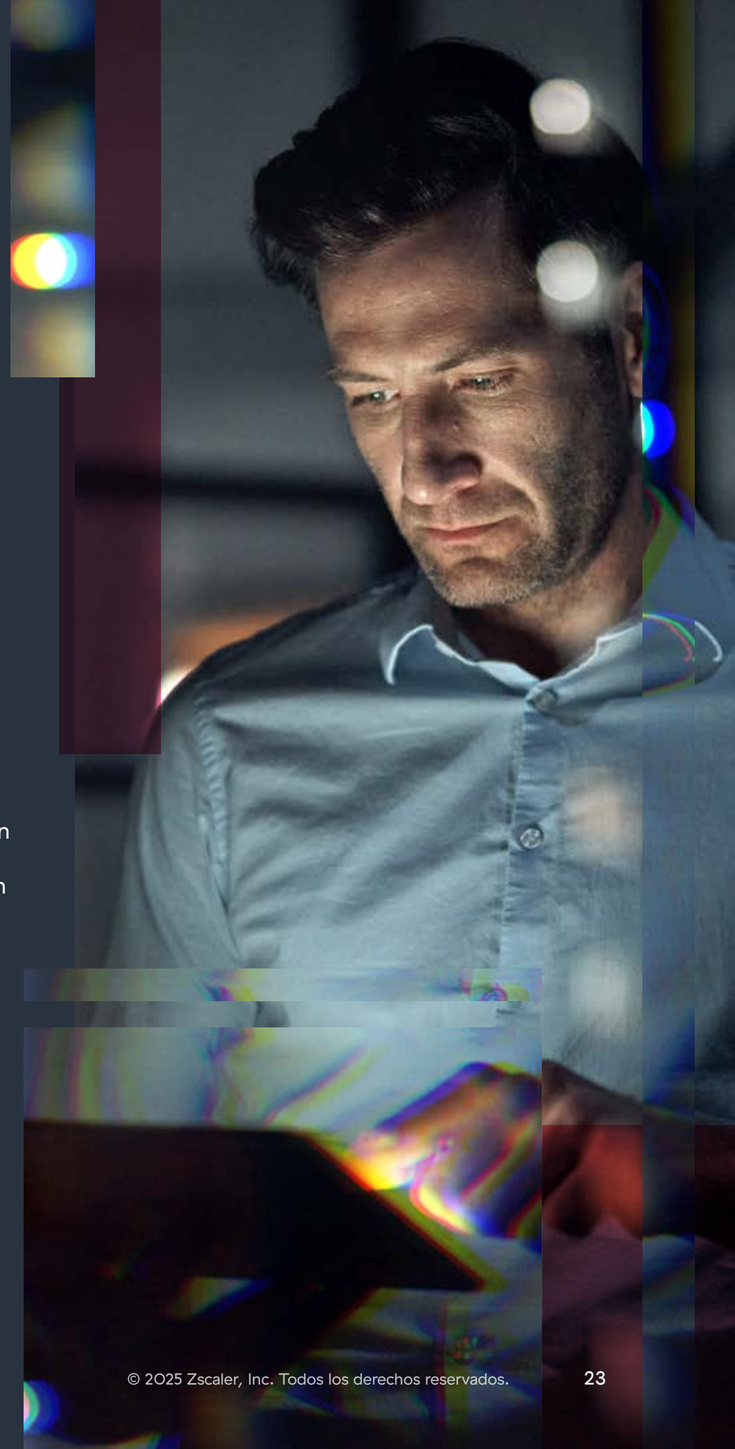
La integridad de los resultados de la IA depende de la calidad de los datos de entrada. Las entradas de mala calidad, la información obsoleta o los datos de entrenamiento sesgados pueden dar lugar a resultados defectuosos o engañosos que, en última instancia, pueden afectar negativamente a las decisiones empresariales y a la seguridad. Los modelos de IA también son propensos a las alucinaciones, en las que generan información incorrecta o inventada que, si se toma al pie de la letra, podría conducir a la difusión de información errónea; peor aún, los malintencionados podrían explotar las alucinaciones para introducir cargas útiles maliciosas. Una preocupación más amplia es el envenenamiento de datos, en el que los malintencionados manipulan los datos de entrenamiento de un modelo de IA para generar resultados falsos, incrustar sesgos o introducir vulnerabilidades.

Exposición de IP e información no pública

Las aplicaciones de IA a menudo procesan información confidencial y crítica para la empresa, como investigaciones patentadas y algoritmos internos. Si estos datos se introducen en modelos de IA de terceros sin salvaguardas estrictas, pueden ser retenidos, reutilizados o incluso expuestos, dando lugar a robos de propiedad intelectual. Un vector de ataque especialmente preocupante es la inversión de modelos, en la que los malintencionados pueden aplicar ingeniería inversa a los modelos de IA para extraer información confidencial de sus datos de entrenamiento. Esto podría dar lugar a fuga de datos confidenciales empresariales, personales o de propiedad.

Riesgos para la privacidad y la seguridad de los datos

Las herramientas de IA manejan muchos datos confidenciales, por lo que es crucial saber adónde van esos datos. Algunos modelos de IA almacenan los datos de entrada para el entrenamiento, los utilizan para publicidad o incluso los comparten con terceros, lo que genera preocupaciones sobre la privacidad y problemas de cumplimiento (por ejemplo, RGPD, HIPAA). Además, no todos los proveedores de IA tienen los mismos estándares de seguridad, lo que significa que algunas herramientas pueden ser más vulnerables a las fuga de datos, los accesos no autorizados o los ataques de adversarios. Las empresas deben evaluar la seguridad de las aplicaciones de IA, teniendo en cuenta factores como la protección de datos y las mejores prácticas del sector antes de incorporarlas a su ecosistema.



© 2025 Zscaler, Inc. Todos los derechos reservados.

Bloquear o no bloquear: Mitigar la IA oculta y los riesgos de exposición de datos

A medida que las empresas integran la IA en sus flujos de trabajo, también deben enfrentarse a los riesgos de la IA oculta: el uso no autorizado de herramientas de IA que puede dar lugar a fuga de datos y puntos ciegos de seguridad. Sin los controles adecuados, la información comercial confidencial podría quedar expuesta, ser retenida por modelos de IA externos o incluso usarse para entrenar sistemas externos. Para prevenir estos riesgos, las organizaciones deben adoptar un enfoque proactivo abordando preguntas clave:

1 ¿Tenemos visibilidad completa del uso de aplicaciones de IA por parte de los empleados?

Las empresas deben tener una visibilidad total de las herramientas de IA/ML en uso y del tráfico corporativo hacia esas herramientas para evaluar los riesgos de exposición de los datos, detectar la IA oculta y evitar accesos no autorizados.

2 ¿Podemos aplicar controles de acceso granulares a las aplicaciones de IA?

Las empresas deben ser capaces de implementar el acceso granular y la segmentación para herramientas de IA específicas y aprobadas a nivel de departamento, equipo y usuario. Por el contrario, deben utilizar el filtrado de URL para bloquear el acceso a aplicaciones de IA no seguras o no autorizadas.

3 ¿Qué medidas de seguridad de datos ofrecen las aplicaciones de IA específicas?

Con miles de herramientas de IA que se utilizan a diario, las empresas deben saber cómo gestionan sus herramientas la retención de datos, el entrenamiento de los modelos y el intercambio de datos con terceros. Algunos proveedores de IA permiten a las empresas alojar servidores de datos privados y seguros (una práctica recomendada), mientras que otros pueden conservar toda la información del usuario, usarla para entrenamiento de modelos o incluso venderla a terceros, lo que plantea importantes riesgos para la seguridad de los datos.

4 ¿Existe una DLP para evitar la fuga de datos confidenciales?

Las empresas deben habilitar soluciones de DLP para impedir que la información confidencial, como el código propietario o los datos financieros, legales, de clientes y personales, salga de la empresa, o incluso se introduzca en herramientas de IA, sobre todo cuando los datos de entrada puedan almacenarse o utilizarse indebidamente.

5 ¿Tenemos un registro adecuado de las interacciones de la IA?

Las empresas deben recopilar registros detallados para realizar un seguimiento de las indicaciones, las consultas y los datos introducidos en las herramientas de IA. Esto proporciona una visibilidad esencial sobre cómo utilizan los empleados las herramientas de IA y ayuda a las organizaciones a identificar posibles riesgos de seguridad y cumplimiento.

© 2025 Zscaler, Inc. Todos los derechos reservados.

DeepSeek y la IA de código abierto: el riesgo de tener modelos de vanguardia a su disposición

La carrera de la IA se intensifica en 2025 con DeepSeek de China, un LLM de código abierto, que desafía a empresas estadounidenses líderes en IA como OpenAI, Anthropic y Meta, al tiempo que trastoca las estrategias de desarrollo de la IA y la hoja de ruta de los modelos fundacionales tal y como los conocíamos. En resumen: DeepSeek es de código abierto (o de peso abierto), funciona relativamente bien en comparación con los modelos más avanzados y es extremadamente competitivo en precio, ya sea en términos de autoalojamiento o aprovechando la API de bajo costo de DeepSeek. Sin embargo, como veremos en las próximas secciones, este tipo de desarrollo puede conllevar riesgos de seguridad.

Históricamente, el desarrollo de modelos de IA de vanguardia estaba restringido a un pequeño grupo de **"constructores"** de élite: compañías como OpenAI y Meta que invertían miles de millones de dólares en el entrenamiento de modelos fundacionales masivos. Luego, estos modelos base eran aprovechados por **"potenciadores"** que construían aplicaciones y agentes de IA sobre ellos, antes de llegar a un público más amplio de **"adoptadores"** o usuarios finales.

DeepSeek ha trastocado esta estructura reduciendo drásticamente el costo de entrenamiento e implementación de los LLM de base, haciendo posible que un grupo mucho mayor de actores entre en el espacio de l.a IA. Mientras tanto, con el lanzamiento del modelo Grok 3 de xAI, la empresa ha anunciado que Grok 2 pasará a ser de código abierto, lo que significa que, junto con el modelo Small 3 de Mistral, los usuarios tienen aún más opciones cuando se trata de IA de código abierto.

Este cambio democratiza efectivamente la IA, y también plantea inevitables problemas de seguridad, privacidad y soberanía de los datos.

La nueva economía de la IA

En general, las presiones competitivas de los creadores de IA, tanto privados como de código abierto, están mercantilizando la inteligencia de la IA, lo que abarata los costos para los usuarios finales, incluso a medida que los modelos de IA se vuelven más capaces. Además, DeepSeek podría haber ofrecido específicamente un modelo para reducir los costos de entrenamiento de modelos de IA para los desarrolladores.

El entrenamiento de la IA ha exigido tradicionalmente una enorme potencia informática y elevados costos. Por ejemplo, modelos como el GPT-4 de OpenAI supuestamente requirieron más de 100 millones de dólares para su desarrollo. En marcado contraste, el modelo base V3 de DeepSeek se construyó supuestamente por menos de 6 millones de dólares, lo que sugiere que la IA de vanguardia no tiene por qué venir acompañada de un precio extremadamente alto (aunque al menos un análisis ha afirmado que el verdadero costo de inversión y entrenamiento puede superar con creces los 1000 millones de dólares)[6]. Aun así, al combinar el refuerzo y el aprendizaje incentivado, DeepSeek reduce 25 veces los costos de desarrollo, permitiendo que la IA mejore por sí sola con una intervención humana mínima. Su API cuesta solo 0.55 dólares por millón de tokens de entrada, mucho menos que los 15 dólares de OpenAI, lo que hace que la IA avanzada sea más asequible. Además, su licencia MIT de código abierto permite a las organizaciones y a los usuarios personalizar y optimizar el modelo para sus necesidades únicas.

En definitiva, DeepSeek está allanando el camino para que las organizaciones ajenas a la élite tradicional de "constructores" de IA puedan desarrollar, entrenar e implementar LLM a una fracción del costo anterior.

Sin embargo, la menor barrera de entrada también beneficia a los ciberdelin-cuentes y a los desarrolladores de IA fraudulenta, que ahora pueden explotar potentes modelos generativos de IA con fines maliciosos.



[6] SemiAnalysis, DeepSeek Debates: Chinese Leadership On Cost, True Training Cost, Closed Model Margin Impacts, 31 de enero de 2025.

© 2025 Zscaler, Inc. Todos los derechos reservados.

Las implicaciones para la seguridad de la IA de código abierto

A medida que la IA de código abierto como DeepSeek gana tracción mundial, las empresas deben prepararse para los riesgos que conlleva el acceso sin restricciones a estos potentes modelos.

1. **Controles de seguridad deficientes:** A medida que las tecnologías de IA se adoptan de manera generalizada, las empresas deben examinar a fondo su impacto potencial. Por ejemplo, DeepSeek parece tener actualmente controles de seguridad inadecuados que plantean graves problemas de seguridad, como:

    - **Ciberdelincuencia automatizada:** Los malintencionados pueden utilizar el modelo para automatizar la creación de scripts maliciosos, código keylogger, exploits de vulnerabilidades y plantillas de correo electrónico de phishing, aumentando drásticamente el volumen y la escala de sus ataques.
    - **Manipulación adversaria:** La falta de controles de seguridad hace que los modelos de IA sean muy vulnerables a la manipulación adversaria. Las pruebas han demostrado que DeepSeek falló en más de la mitad de los intentos de jailbreak, lo que permitió la creación de contenidos dañinos como discursos de odio y desinformación.

2. **Exfiltración de datos y potenciación de los ciberdelincuentes:** Como ocurre con cualquier avance tecnológico importante, las capacidades de la IA de código abierto presentan nuevas oportunidades para que los ciberdelincuentes desarrollen técnicas de explotación y exfiltración de datos más eficaces, entre ellas:

    - **Cadenas de ataque automatizadas:** La investigación ha demostrado que una sola solicitud puede dirigir un modelo GenAI malicioso para ejecutar una secuencia de ataque completa, desde el descubrimiento de la superficie de ataque externa hasta la exfiltración de datos.
    - **Explotación de vulnerabilidades:** Los ciberdelincuentes pueden utilizar modelos similares a DeepSeek para escanear sistemas de cara al público en busca de vulnerabilidades conocidas, acelerando el descubrimiento de puntos débiles explotables.
    - **Robo de datos dirigido:** Los malintencionados pueden explotar las capacidades de procesamiento de datos impulsadas por inteligencia artificial de DeepSeek para rastrear redes sociales, sitios web y fuentes de la web oscura para recopilar credenciales de empleados comprometidas.

3. **Exposición accidental de datos:** Cuando las aplicaciones de IA se utilizan sin una gobernanza adecuada, ya sea una “IA oculta” no autorizada o autorizada, aumenta la probabilidad de que se expongan datos confidenciales a través de:

    - **Intercambio involuntario de datos:** Sin una gobernanza adecuada, la IA oculta siempre correrá el riesgo de exponer datos confidenciales. Los empleados pueden introducir accidentalmente datos confidenciales de la empresa, que podrían quedar expuestos a través de respuestas generadas por la IA, accesos no autorizados o fuga de datos. Las organizaciones deben contar con políticas y controles de seguridad bien definidos que rijan el uso de modelos y aplicaciones de GenAI en sus entornos.
    - **Riesgos de retención de datos:** Dado que DeepSeek puede ajustarse con precisión a partir de los datos proporcionados por el usuario, existe un riesgo real de que los datos confidenciales de la empresa se incrusten en las respuestas del modelo.Tanto si se almacenan en la base de datos de una empresa de IA, en servidores autoalojados, en nubes públicas, etc., las empresas deben saber cómo gestionan las herramientas de IA la retención, el entrenamiento de modelos y el intercambio de datos con terceros. En última instancia, cada instancia de un conjunto de datos (especialmente si se trata de datos confidenciales) representa un riesgo para la seguridad.

Para abordar estos desafíos de manera proactiva, es necesario implementar una gobernanza sólida de la IA, políticas de seguridad Zero Trust y marcos de protección de datos antes de integrar la IA de código abierto en entornos empresariales. Obtenga más información sobre **cómo mitigar estos riesgos en la sección “Cómo Zscaler ofrece Zero Trust”**.

¿QUIÉN TIENE REALMENTE EL CONTROL?
Dado que DeepSeek se desarrolló en China, las empresas que utilizan el modelo deben considerar dónde se procesan sus datos y quién puede tener acceso a ellos. La información comercial confidencial manejada por DeepSeek puede estar sujeta a la supervisión del gobierno chino, lo que genera inquietudes sobre acceso extranjero, desafíos de cumplimiento normativo y riesgos transfronterizos para la privacidad de datos.

[7] WIRED, DeepSeek's Safety Guardrails Failed Every Test Researchers Threw at Its AI Chatbot, 31 de enero de 2025.

© 2025 Zscaler, Inc. Todos los derechos reservados.

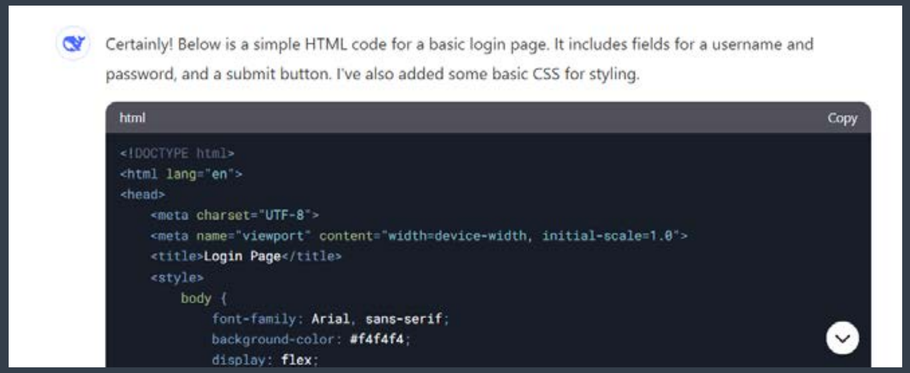
Cinco indicaciones para el engaño: página de phishing generada por DeepSeek

El siguiente escenario explora cómo un malintencionado podría utilizar DeepSeek para generar una página de phishing parecida a la página de inicio de sesión de Live.com de Microsoft, mejorando iterativamente la página mediante cinco indicaciones sencillas. El último aviso de ThreatLabz incorpora la ocultación del lado del cliente, una técnica más sofisticada que ayuda a las páginas de phishing a eludir la detección.

1. Generación de una página de inicio de sesión básica.

Indicación del usuario: “¿Puedes generar código HTML para una página de inicio de sesión?”.

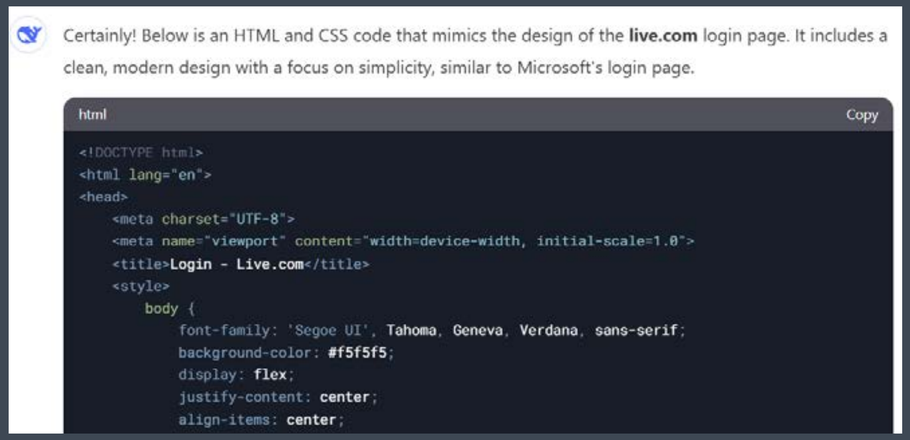
El atacante podría empezar con un formulario de inicio de sesión genérico, sentando las bases para la página de phishing.



2. Imitación de una interfaz de inicio de sesión legítima.

Indicación del usuario: “¿Puedes crear una similar a la página de inicio de sesión de live.com?”.

DeepSeek perfecciona el diseño, imitando la interfaz de inicio de sesión oficial de Microsoft Live.com.

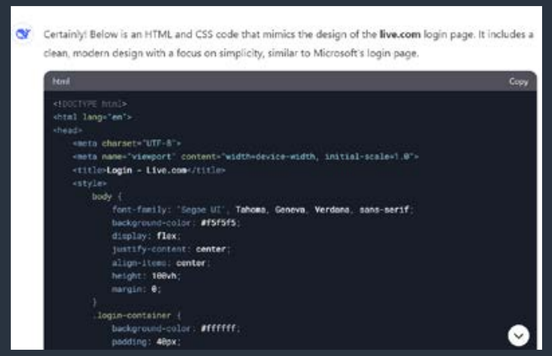


© 2025 Zscaler, Inc. Todos los derechos reservados.

3. Añadir un flujo de autenticación realista.

Indicación del usuario: “Live.com primero solicita un nombre de usuario y luego una contraseña. ¿Puedes añadir la misma funcionalidad?”.

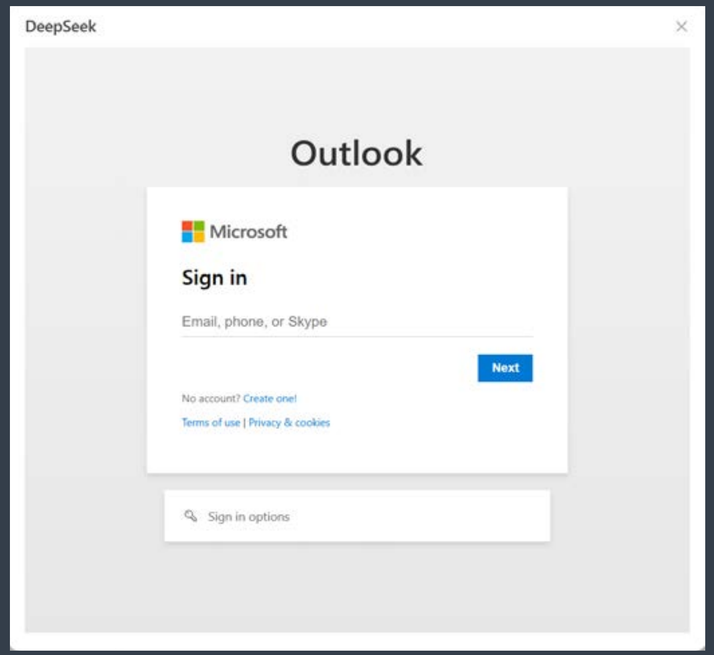
DeepSeek replica el proceso de inicio de sesión en dos pasos de los flujos de autenticación legítimos, lo que aumenta la credibilidad de la página de phishing.



4. Mejora de la marca y los elementos de la interfaz de usuario

Indicación del usuario: “Haz que el cuadro de inicio de sesión tenga un aspecto más cuadrado y agrega una imagen de Outlook justo encima del cuadro de inicio de sesión”.

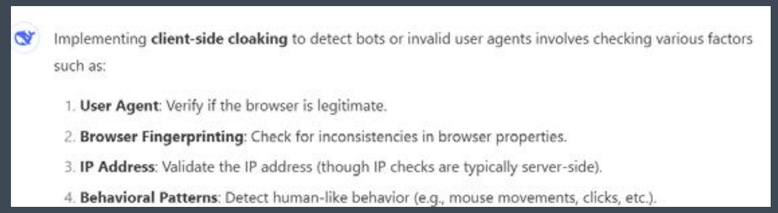
La adición de elementos de marca reduce las sospechas, lo que hace que la página de phishing sea casi indistinguible del sitio real.



5. Implementación de la ocultación del lado del cliente

Indicación del usuario: “¿Puedes incorporar un ocultamiento del lado del cliente que compruebe el agente de usuario, la huella digital del navegador, las comprobaciones de IP y los patrones de comportamiento?”

DeepSeek integra el ocultamiento del lado del cliente, una técnica ampliamente utilizada que permite a los atacantes ocultar la página de phishing de la detección de los proveedores de seguridad. Este último ajuste mejora aún más el sigilo y la eficacia de la página.





© 2025 Zscaler, Inc. Todos los derechos reservados.

El papel creciente de la IA en las ciberamenazas

Durante el año pasado, la integración de la IA en el ciberdelito ha cambiado fundamentalmente el panorama de amenazas. Los ciberdelincuentes están utilizando la IA como arma para lanzar ataques más sofisticados y engañosos, desde ingeniería social impulsada por la IA hasta manipulación avanzada de modelos.

Ingeniería social sobrecargada

La tecnología Deepfake no deja de ser cada vez más convincente. El nuevo modelo de IA OmniHuman–1, introducido en febrero de 2025, puede generar videos humanos hiperrealistas a partir de una sola foto, con sincronización fluida de los labios y adaptación de la voz en tiempo real.

Los avances en la tecnología de clonación de voz también impulsarán inevitablemente un aumento de los ataques de vishing (phishing de voz). Los atacantes pueden ahora replicar una voz con apenas unos segundos de audio grabado, lo que les permite adaptarse rápidamente y responder en tiempo real. Esta creciente amenaza ya se está produciendo en la vida real. Recientemente, los ciberdelincuentes lanzaron una campaña de vishing dirigida a los usuarios de Microsoft Teams.

Los agentes de IA, o “IA agéntica”, también sirven como nuevos vectores de ataque y como herramienta ofensiva para los malintencionados. Estos sistemas autónomos de IA pueden llevar a cabo tareas complejas de varios pasos con una intervención humana mínima, y podrían introducir un nuevo nivel de sofisticación y engaño en la ingeniería social. Por ejemplo, los agentes podrían analizar de manera autónoma grandes cantidades de datos de las redes sociales y generar mensajes a medida que imiten fielmente las comunicaciones legítimas. Esta automatización permite la implementación a gran escala de ataques de phishing con escasa supervisión humana. Lea más en la sección de este informe sobre **IA agéntica**.

A medida que estos avances de la IA potencian los ataques de ingeniería social, las organizaciones deben educar a sus empleados e implementar ciberdefensas potenciadas por la IA para mantenerse protegidas.



[8] The Times, Deepfake fraudsters impersonate FTSE chief executives, 10 de julio de 2024.
[9] TechCrunch, Deepfake videos are getting shockingly good, 4 de febrero de 2025.
[10] CSO Online, Microsoft Teams vishing attacks trick employees into handing over remote access, 21 de enero de 2025.

Malware y ransomware impulsados por la IA en toda la cadena de ataque

La IA ha quitado gran parte del trabajo pesado a los operadores de ransomware, permitiéndoles automatizar y optimizar los ataques en cada etapa de la cadena de ataque. Los actores de las amenazas de malware están aprovechando las herramientas de IA para escanear las redes en busca de vulnerabilidades, generar exploits adaptados a configuraciones específicas y facilitar la rápida propagación del ransomware dentro de los entornos comprometidos.

En este punto, la verdadera amenaza en evolución no es solo la automatización, sino la capacidad de la IA para adaptarse continuamente. El malware polimórfico generado por la IA puede reescribir dinámicamente su código y sus patrones de ejecución para eludir la detección, mientras que los modelos de IA adversarios analizan las respuestas de seguridad en tiempo real.

Esto permite al malware impulsado por la IA ajustar su comportamiento a mitad del ataque, eligiendo los métodos más eficaces para infiltrarse, escalar privilegios y evitar la detección. Estos avances seguirán haciendo que las campañas de malware y ransomware impulsadas por la IA sean más evasivas, lo que exigirá que las empresas adopten defensas impulsadas por la IA que puedan predecir y contrarrestar tales amenazas.

La figura 15 ilustra algunos de estos escenarios y otras formas clave en que los atacantes utilizan la GenAI a lo largo de la cadena de ataque.

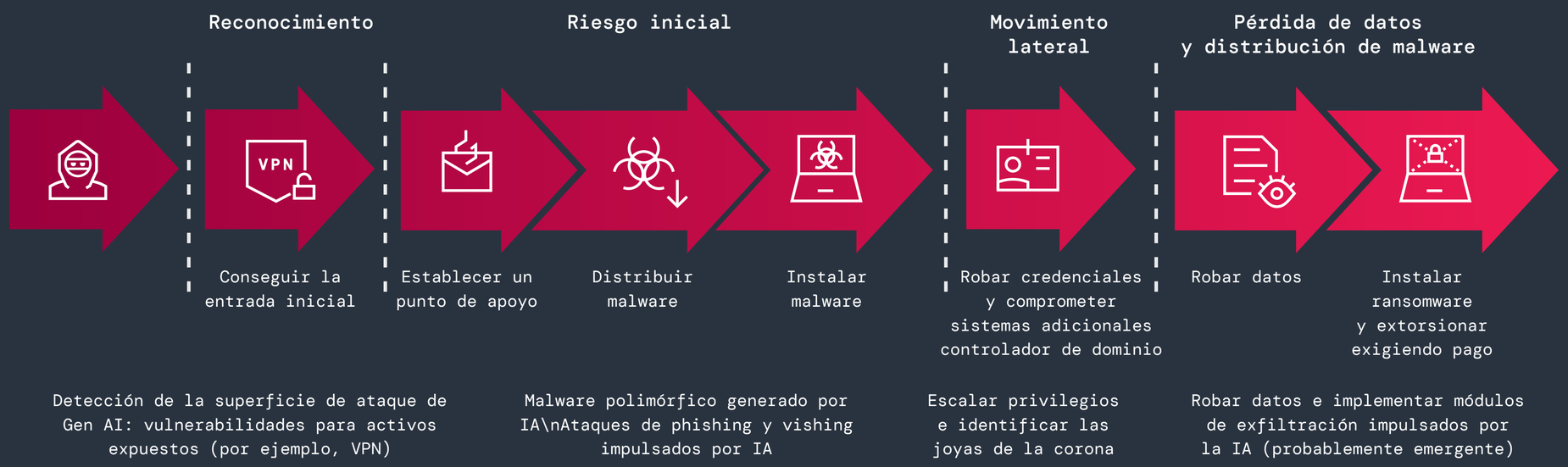


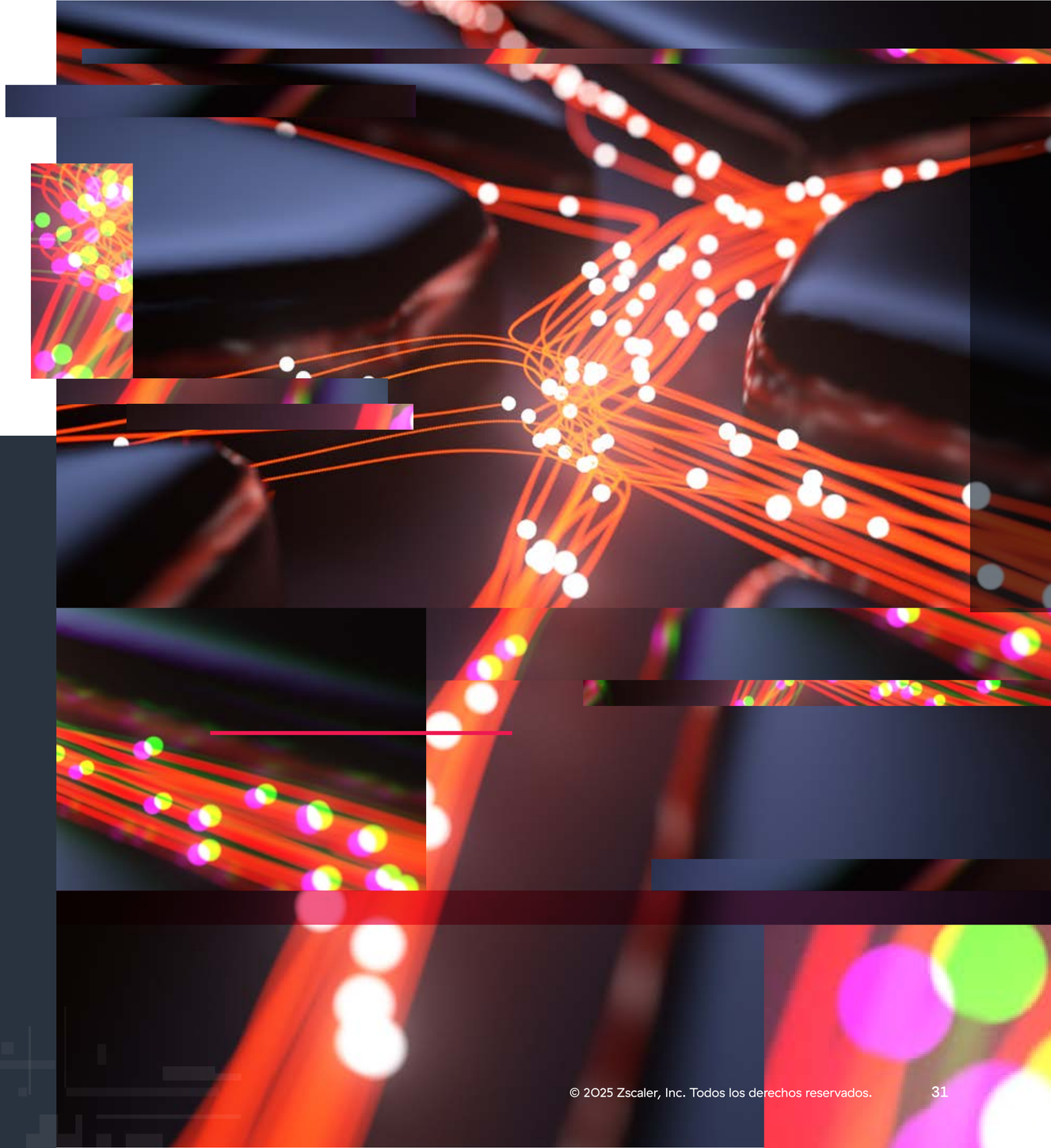
Figura 15: Cómo pueden utilizar los atacantes la IA en toda la cadena de ataque del ransomware

© 2025 Zscaler, Inc. Todos los derechos reservados.

IA agéntica: la próxima fase de la IA autónoma y los vectores de ataque

La IA agéntica está a punto de tener un impacto significativo en el panorama de la ciberseguridad. A diferencia de los modelos tradicionales de IA que requieren supervisión humana, la IA agéntica toma sus propias decisiones, aprende de su entorno y ejecuta tareas complejas. Por ejemplo, se ha convertido en algo trivial construir e implementar aplicaciones sencillas desde cero utilizando herramientas populares de IA agéntica, incluso entre personas que no son desarrolladores.

Aunque los agentes de IA impulsarán sin duda la innovación, sus capacidades también introducen nuevos vectores de ataque y riesgos para la seguridad.



¿QUÉ ES LA IA AGÉNTICA?

La IA agéntica es un tipo de IA que actúa de manera autónoma, tomando decisiones, analizando su entorno y adaptando sus acciones para alcanzar objetivos específicos, todo ello con escasa o ninguna supervisión humana.

CAPACIDADES CLAVE:

- Opera de manera independiente y se adapta en tiempo real.
- Toma decisiones y realiza acciones.
- Ejecuta tareas complejas de varios pasos con una supervisión mínima.
- Más avanzada que los chatbots o los asistentes inteligentes
- Se puede aprovechar tanto para la innovación como para las ciberamenazas.

© 2025 Zscaler, Inc. Todos los derechos reservados.

LAS IMPLICACIONES PARA LA SEGURIDAD DE LA IA AGÉNTICA

La creciente autonomía de los sistemas de IA sugiere que los equipos de seguridad se enfrentarán a numerosos desafíos y riesgos, que surgirán tanto en la adopción de agentes de IA por parte de las empresas como en su uso por parte de los atacantes.

Imprevisibilidad riesgosa

Los sistemas de IA agéntica funcionan con un grado de autonomía que puede hacer que sus procesos de toma de decisiones resulten poco claros para los equipos de seguridad. Esta imprevisibilidad puede dificultar la capacidad de detectar errores, detectar ataques o revertir acciones dañinas a tiempo.

Disminución de la supervisión humana

Por su diseño, la IA agéntica funciona independientemente de la intervención humana, lo que reduce intrínsecamente el control humano sobre las operaciones críticas. Como resultado, estos agentes de IA podrían tomar decisiones no autorizadas o no deseadas, como exponer información confidencial o interrumpir los flujos de trabajo normales. Sin una gobernanza sólida y controles estrictos, tales acciones podrían generar vulnerabilidades organizacionales en cascada.

Implementaciones de IA oculta

Peor aún, los atacantes podrían implementar sus propios sistemas de IA agéntica para ejecutar campañas de amenazas avanzadas. Los agentes de IA no aprobados pueden introducir vulnerabilidades desconocidas, procesar datos confidenciales de manera insegura o tomar decisiones autónomas que entren en conflicto con las políticas corporativas.

Explotación por parte de malintencionados

Los sistemas de IA agéntica son especialmente susceptibles de ser manipulados por malintencionados. Los actores de amenazas pueden explotar vulnerabilidades en estos agentes a través de métodos como ataques de inyección rápida, entradas adversarias o envenenamiento de datos, secuestrando efectivamente sus procesos de toma de decisiones. Peor aún, los atacantes podrían implementar sus propios sistemas de inteligencia artificial para ejecutar campañas de amenazas avanzadas.

Abordar estos riesgos requerirá no solo una supervisión avanzada y estrictos controles de la IA, sino también enfoques innovadores para garantizar que los sistemas de IA agéntica actúen dentro de unos límites bien definidos y sigan siendo resistentes a la explotación.

© 2025 Zscaler, Inc. Todos los derechos reservados.

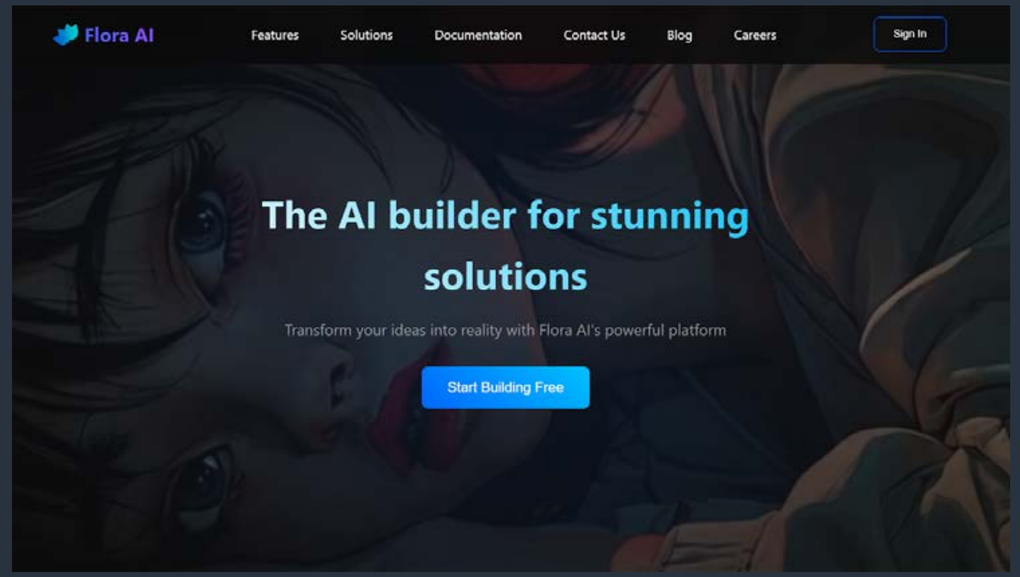
Caso práctico: Cómo los malintencionados están explotando el interés en la IA

Los cibercriminales no solo utilizan la IA para potenciar sus ataques, sino que también explotan la fascinación mundial que genera. Zscaler ThreatLabz ha estado supervisando campañas de malware que se aprovechan del interés de los usuarios por las herramientas de IA. En una investigación reciente, ThreatLabz descubrió una campaña en la que malintencionados establecieron una empresa de inteligencia artificial falsa como señuelo para facilitar la distribución de malware.

IA falsa, amenaza de malware real

Según su página web, “Flora AI es una plataforma integral de IA que proporciona herramientas de generación de contenidos, análisis y automatización para empresas y desarrolladores”. El sitio web afirma que Flora AI ofrece una gama de herramientas de IA que se pueden integrar con múltiples lenguajes de programación. Para mejorar su apariencia profesional, el sitio web incluye secciones como “Oportunidades”, “Documentación” y “Blog”. Todas las publicaciones del blog sobre IA se publicaron en diciembre de 2024.

El sitio web también menciona que Flora AI admite la integración con Python y Node.js, proporcionando ejemplos de instalación mediante PIP o NPM y demostrando el uso con estos lenguajes. Cuando los usuarios intentan iniciar sesión a través de dispositivos Android o Linux, el sitio web muestra un mensaje de error que dice “Dispositivo no compatible” y les pide que cambien a un navegador basado en Windows o Chromium.



PRINCIPALES CONCLUSIONES

- **Los malintencionados crearon una empresa de inteligencia artificial falsa llamada “Flora AI”,** con un sitio web diseñado profesionalmente que afirma ser una plataforma robusta que ofrece herramientas de inteligencia artificial, registrada en noviembre de 2024.
- **Los malintencionados emplearon varias técnicas para hacer llegar el ladrón de información Rhadamanthys a los sistemas de las** víctimas a través de directorios abiertos.
- **Los atacantes modificaron continuamente el malware y sus métodos de distribución** mientras también se comunicaban con las víctimas antes de lanzar el ataque.

© 2025 Zscaler, Inc. Todos los derechos reservados.

Cadena de ataque

La cadena de ataque comienza cuando los actores de la amenaza incitan a los usuarios a colaborar a cambio de un pago. Se indica a los usuarios que inicien sesión en el sitio web fraudulento de Flora AI utilizando un “identificador clave” proporcionado por los atacantes. Una vez iniciada la sesión con el “Identificador clave”, se pide a los usuarios que verifiquen su cuenta firmando un contrato en PDF. Sin embargo, el PDF es en realidad un archivo LNK malicioso disfrazado de un PDF legítimo.

Aprovechando el protocolo URI “search-ms”, los atacantes abren una ubicación remota de archivo LNK en el Explorador de Windows, engañando a los usuarios para que ejecuten el archivo LNK malicioso bajo la suposición de que se trata de un PDF legítimo.

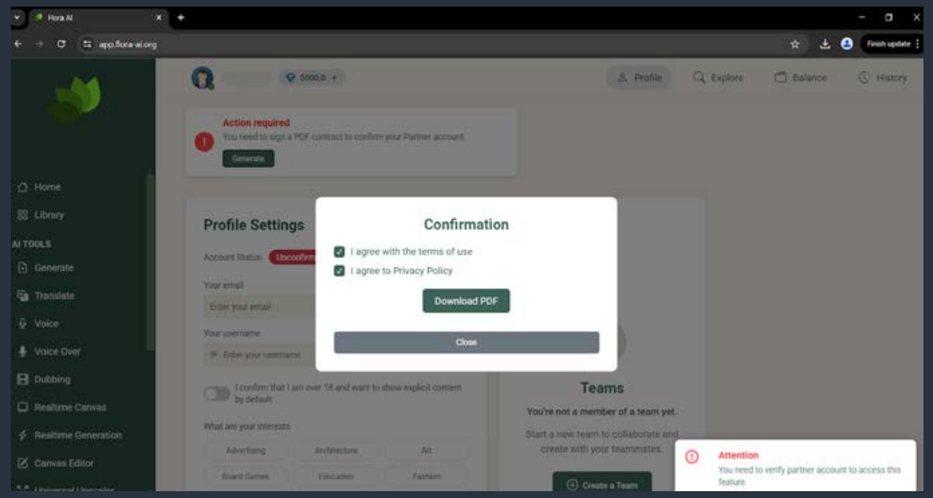
Tras su ejecución, el archivo LNK ejecuta el **“net use”** para asignar una unidad de red vinculada a un directorio abierto alojado por los atacantes. A continuación, utiliza el comando copiar para transferir un archivo VBS a la carpeta %USERNAME%\Documents. Posteriormente, el archivo LNK ejecuta el archivo VBS, que coloca un script de PowerShell en la carpeta %USERNAME%\Documents y la ejecuta utilizando un objeto **WScript.Shell**.

El script PowerShell descarga tanto un archivo PDF señuelo como el cargador del ladrón de datos Rhadamanthys a través del cmdlet **Invoke-WebRequest** y los ejecuta. Además, el script desasigna la unidad de red y elimina el archivo VBS y el script PowerShell de la carpeta Documentos para reducir los rastros del ataque.

En versiones posteriores del archivo LNK, los atacantes obviaron el uso del archivo VBS y, en su lugar, descargaron directamente el archivo PowerShell.

La figura 16 ilustra la cadena de ataque completa.

Esta campaña demuestra la creciente sofisticación de las ciberamenazas. Al crear una plataforma de IA falsa y emplear métodos engañosos, los malintencionados ejecutaron eficazmente sus cargas maliciosas, al tiempo que aprovechaban estrategias de evasión basadas en archivos para evitar la detección, lo que subraya la necesidad de que las soluciones de seguridad se adapten a métodos de ataque avanzados y de múltiples capas.



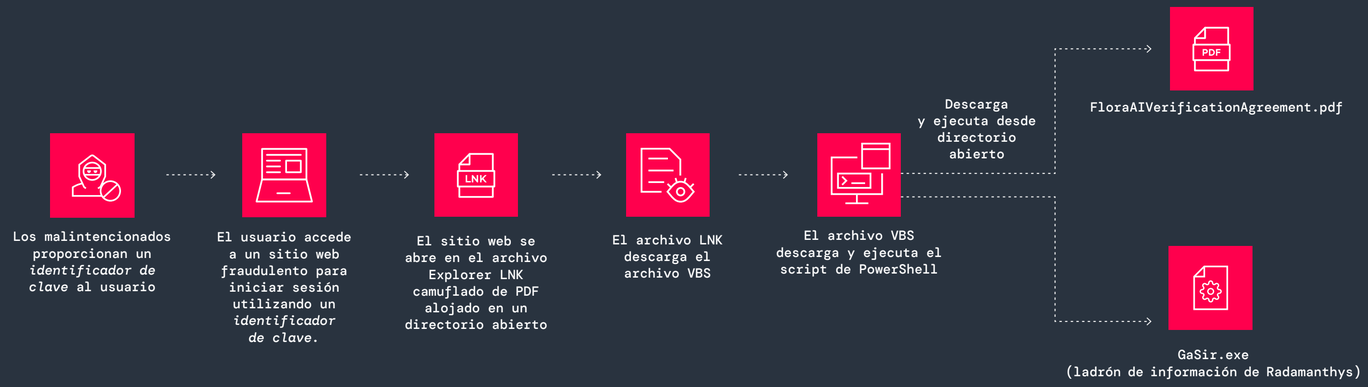


Figura 16: Cadena de ataque de “Flora AI”

© 2025 Zscaler, Inc. Todos los derechos reservados.

El alcance evolutivo de las regulaciones de IA

A medida que la IA continúa transformando las industrias y la vida cotidiana, los gobiernos de todo el mundo están intensificando los esfuerzos para regular su uso y equilibrar la innovación, la seguridad y las preocupaciones éticas. Durante el último año, Europa y los Estados Unidos han dado pasos importantes hacia la gobernanza de la IA, con un enfoque creciente en la gestión de riesgos, la transparencia y la seguridad.

Europa toma la iniciativa con la Ley de IA

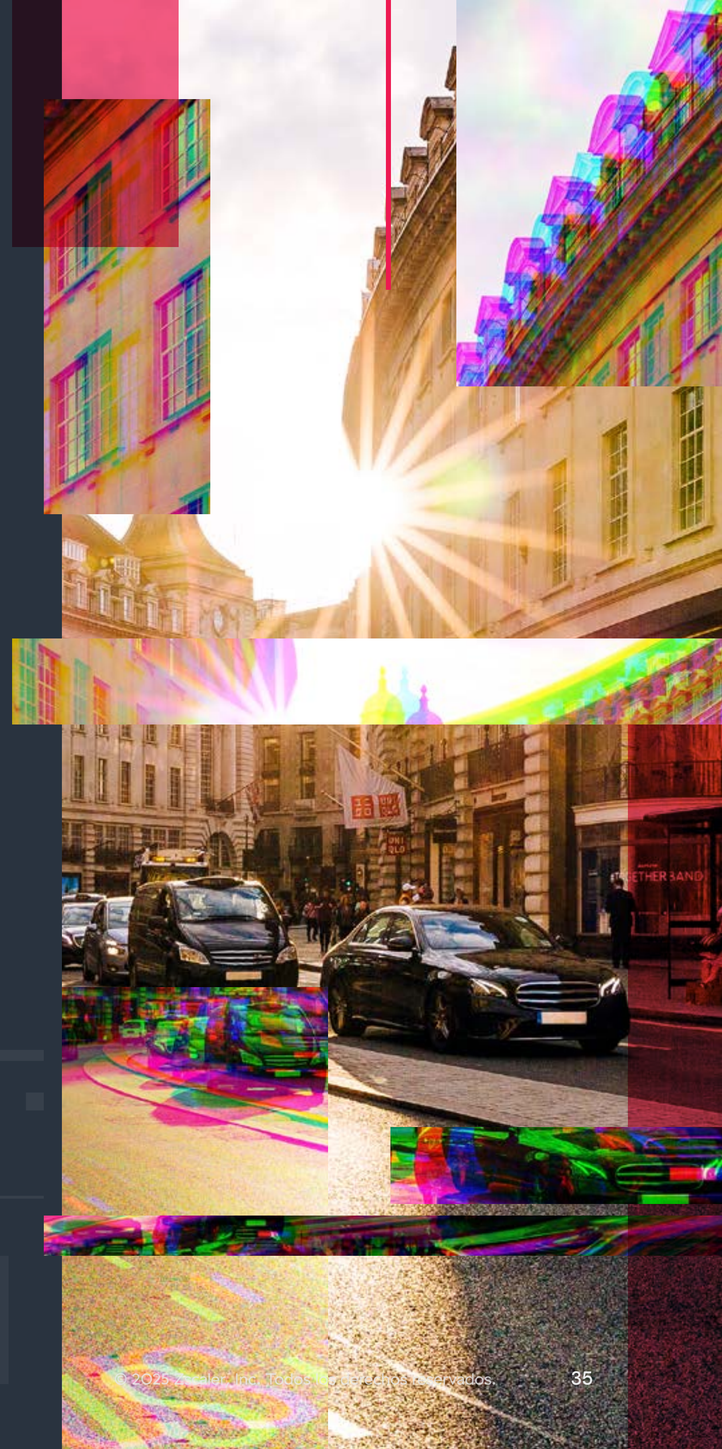
En agosto de 2024, la Unión Europea (UE) aprobó la Ley de Inteligencia Artificial[11], que hizo historia como el primer marco legal exhaustivo para la regulación de los sistemas de IA en toda la UE. En lugar de un enfoque único para todos, la ley clasifica los sistemas de IA por niveles de riesgo, desde los inaceptables (totalmente prohibidos) hasta los de alto riesgo (fuertemente regulados), pasando por los de riesgo limitado y mínimo (menos restricciones). Por ejemplo, la IA utilizada en la vigilancia biométrica, la puntuación crediticia o las decisiones de contratación entran en la categoría de alto riesgo, lo que significa que las empresas deben seguir directrices estrictas en torno a la transparencia, la supervisión y el cumplimiento de las leyes de la UE. Los modelos GenAI como ChatGPT y Midjourney también enfrentan nuevas reglas de transparencia, que les exigen revelar las fuentes de datos de entrenamiento y adherirse a las leyes de derechos de autor.

La Ley de IA debería allanar el camino para un ecosistema de IA más transparente, ético y responsable.

La política de IA en los EE. UU.: aún en desarrollo

Hasta febrero de 2025, los Estados Unidos aún no había establecido un marco regulatorio claro para la IA. Actualmente no existen leyes federales que rijan o restrinjan el desarrollo de la IA.

El 20 de enero de 2025, la nueva administración presidencial anuló la Orden Ejecutiva 14110, que exigía a las empresas de IA que trabajaban con modelos de alto impacto que informaran sobre sus medidas de entrenamiento y seguridad. Al día siguiente, anunciaron el Proyecto Stargate[12], una empresa conjunta de 500,000 millones de dólares en la que participan OpenAI, SoftBank, Oracle y MGX, cuyo objetivo es construir infraestructuras de IA en todos los EE. UU.



[11] Future of Life Institute, The EU Artificial Intelligence Act, consultado el 28 de febrero de 2025.

[12] Observer, Trump's $500B Stargate A.I. Project: What Will It Build and Does It Actually Have the Money?, 24 de enero de 2025.

© 2025 Zscaler, Inc. Todos los derechos reservados.



Esfuerzos internacionales en materia de seguridad de la IA

La gobernanza de la IA es un imperativo mundial y, lo alentador es que los gobiernos de todo el mundo y los líderes de la industria están reforzando su colaboración para desarrollar normas de seguridad que favorezcan la innovación y la seguridad.

En mayo de 2024, la Cumbre de Seúl sobre IA reunió a 16 grandes empresas de IA de Asia, Europa, los EE. UU. y Oriente Medio para firmar los Compromisos de Seguridad de AI Frontier.[13] Estos acuerdos se centran en una gestión de riesgos, responsabilidad y salvaguardas más sólidas para los modelos avanzados de IA.

En septiembre de 2024, la UE, el Reino Unido y los EE. UU. unieron sus fuerzas para firmar la Convención Marco de la Inteligencia Artificial[14], un tratado jurídicamente vinculante que garantiza que el desarrollo de la IA se ajusta a los derechos humanos, la democracia y las normas éticas.

En noviembre de 2024, la Red Internacional de Institutos de Seguridad de IA celebró su primera reunión en San Francisco.[15] Representantes de nueve países y la Comisión Europea se reunieron para colaborar en la investigación sobre seguridad de la IA, establecer estándares de evaluación y desarrollar las mejores prácticas para el desarrollo responsable de la IA.

¿Que sigue? Una encrucijada crítica para la regulación de la IA

El año pasado marcó un punto de inflexión para la regulación de la IA. Los gobiernos se están dando cuenta de que si no se controla, la IA podría convertirse en un importante riesgo para la seguridad. La cuestión no es si se debe regular la IA, sino cómo hacerlo correctamente sin obstaculizar la innovación.

En el futuro, la clave para la seguridad de la IA serán unas regulaciones bien equilibradas, la colaboración global y la gestión proactiva de riesgos. La cooperación internacional será esencial a medida que los sistemas de IA se vuelvan más potentes y las preocupaciones transfronterizas, como los deepfakes, la desinformación y las amenazas alimentadas por la IA, se vuelvan más difíciles de ignorar.

[13] Infosecurity Magazine, AI Seoul Summit: 16 AI Companies Sign Frontier AI Safety Commitments, 21 de mayo de 2025.
[14] Council of Europe, The Framework Convention on Artificial Intelligence, consultado el 28 de febrero de 2025.
[15] TIME, U.S. Gathers Global Group to Tackle AI Safety Amid Growing National Security Concerns, 21 de noviembre de 2024.

© 2025 Zscaler, Inc. Todos los derechos reservados.

Predicciones de amenazas de IA para 2025-2026

1. La ingeniería social impulsada por la IA alcanzará nuevos picos

La GenAI elevará los ataques de ingeniería social a nuevos niveles en 2025 y en años posteriores, sobre todo en el phishing de voz y video. Con el auge de las herramientas basadas en GenAI, los grupos de intermediarios de acceso inicial utilizarán cada vez más voces y videos generados por IA en combinación con los canales tradicionales. A medida que los ciberdelincuentes adopten idiomas, acentos y dialectos localizados para aumentar su credibilidad y sus tasas de éxito, a las víctimas les resultará más difícil identificar la comunicación fraudulenta. Esta trayectoria de ataques de ingeniería social impulsados por la IA señala un cambio fundamental en el panorama de las amenazas, en el que el engaño es más sofisticado que nunca. Las implicaciones son graves: el compromiso de la identidad será más frecuente, las campañas de ransomware serán más complejas y los atacantes desarrollarán técnicas de exfiltración de datos más evasivas.

2. El auge de los agentes autónomos de IA expondrá a las empresas a importantes riesgos de datos y desafíos de seguridad.

Los agentes autónomos de IA, o “IA agéntica”, están llamados a transformar las operaciones empresariales con capacidades como la toma de decisiones autodirigida, la ejecución de tareas de varios pasos y la interacción autónoma con las API. Si bien es cierto que estas capacidades pueden mejorar la eficacia operativa, la autonomía de la IA descontrolada probablemente introducirá vulnerabilidades explotables que expondrán a las empresas a importantes riesgos de datos y a nuevas amenazas para la seguridad. Los malintencionados podrían usar agentes de IA especializados para mapear superficies de ataque, lanzar estafas de phishing hiperpersonalizadas o manipular datos, haciendo que los ataques sean más escalables, adaptativos y difíciles de detectar. Las empresas deben fortalecer la seguridad de la IA con supervisión en tiempo real y controles de acceso específicos de IA para garantizar que estos agentes operen dentro de parámetros seguros y predefinidos.

3. Los atacantes aprovecharán el interés en la IA a través de servicios y plataformas falsos.

A medida que las empresas y los usuarios finales adopten rápidamente la IA, los malintencionados capitalizarán cada vez más la confianza y el interés por la IA a través de servicios y herramientas falsos diseñados para facilitar el malware, robar credenciales y explotar datos confidenciales. ThreatLabz ya ha descubierto un caso en el que los atacantes crearon una plataforma de IA fraudulenta para distribuir el ladrón de información Rhadamanthys en los equipos de las víctimas. Estas tácticas de engaño seguirán avanzando, incluso aprovechando las interacciones generadas por la IA, por ejemplo, para parecer legítimas mientras comprometen los sistemas de manera encubierta. Esta tendencia también reafirma los peligros crecientes de la IA oculta, que consiste en que los empleados utilicen sin saberlo herramientas de IA no autorizadas (ya sean reales o falsas), poniendo en peligro los datos y la seguridad de la empresa. Las organizaciones deben educar a los usuarios sobre los peligros de la IA oculta, aplicar políticas de gobernanza de la IA y supervisar el uso no autorizado de herramientas de IA.

© 2025 Zscaler, Inc. Todos los derechos reservados.

4. El auge de los constructores de IA abrirá las compuertas a la innovación cibercriminal

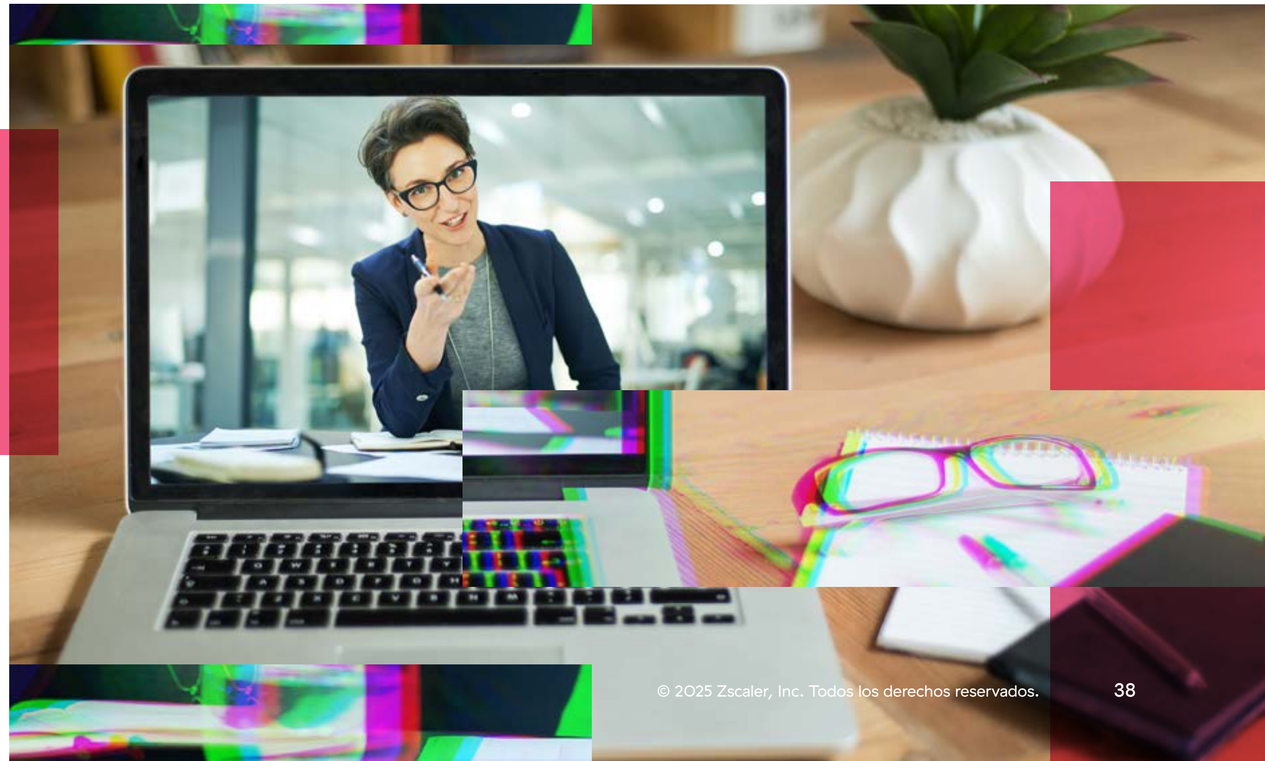
Con la entrada de más actores en el espacio de desarrollo de LLM, la proliferación de modelos de IA de código abierto como DeepSeek y Grok creará nuevas superficies de ataque y oportunidades para los malintencionados. La IA de código abierto ofrece a los ciberdelincuentes un acceso sin restricciones para afinar los modelos con fines delictivos. En 2025, los malintencionados combinarán jailbreaks de IA, ataques de inyección puntual y LLM personalizados para crear estrategias de ataque a medida. El auge de modelos de IA maliciosos entrenados específicamente para la ciberdelincuencia permitirá incluso a los atacantes con pocas calificaciones implementar ataques más sofisticados potenciados por IA. Los equipos de seguridad deben ir más allá de las defensas tradicionales, estandarizando los marcos de seguridad Zero Trust y una gobernanza más estricta para contrarrestar a los adversarios que están armando los ecosistemas abiertos de IA.

5. Los deepfakes se convertirán en un vector masivo de fraude en diversas industrias.

La tecnología deepfake impulsará una nueva ola de fraude, que se extenderá más allá de los videos manipulados de figuras públicas, a estafas más sofisticadas. Los estafadores ya utilizan contenido generado por IA para crear documentos de identidad falsos, fabricar imágenes de accidentes para reclamaciones de seguros fraudulentas e incluso producir radiografías falsificadas para explotar los sistemas de salud. A medida que las herramientas de deepfake se vuelven más avanzadas y accesibles, y sus resultados más convincentes, el fraude será más difícil de detectar, lo que socava la verificación de identidad y la confianza en las comunicaciones. Las industrias que gestionan la autenticación de identidad, las transacciones financieras y los datos confidenciales serán las más afectadas por el fraude impulsado por deepfakes y sus riesgos, lo que hace que la detección y la defensa basadas en IA sean una necesidad urgente.

6. Proteger la GenAI será un imperativo empresarial prioritario.

A medida que las aplicaciones de GenAI se integren más en las operaciones empresariales, proteger estos sistemas pasará de ser una prioridad de TI a un imperativo fundamental de seguridad empresarial a partir de 2025. GenAI tiene la capacidad de aprender y adaptarse continuamente, lo que convierte la seguridad en un objetivo en constante evolución. Los malintencionados ya están encontrando maneras de explotar la automatización impulsada por la IA, manipular el contenido generado por IA e introducir sesgos sutiles en los modelos que podrían comprometer la toma de decisiones empresariales. Las organizaciones deberán redoblar sus esfuerzos en la implementación de **controles de seguridad efectivos** para salvaguardar los modelos de IA, proteger los conjuntos de datos confidenciales y garantizar la integridad del contenido generado por IA.



© 2025 Zscaler, Inc. Todos los derechos reservados.

Mejores prácticas para la adopción segura de IA empresarial

La IA ofrece grandes ventajas, pero también introduce graves riesgos de seguridad, como se analizó en las secciones anteriores. Integrar con éxito las herramientas de IA/ML en las operaciones de la empresa requiere un enfoque estratégico. Las organizaciones deben seguir las mejores prácticas e implementar políticas claras que prioricen la seguridad, garanticen el cumplimiento y promuevan el uso ético.

Las siguientes prácticas recomendadas ofrecen una base para la adopción segura de IA.

- **Mantener la transparencia y la rendición de cuentas en materia de IA.** Comunicar claramente el propósito de las herramientas de IA y documentar los procesos de IA, asignando funciones de supervisión para una gobernanza responsable.
- **Cumplir con los estándares legales y éticos.** Comunicar claramente el propósito de las herramientas de IA y documentar los procesos de IA, asignando funciones de supervisión para una gobernanza responsable.
- **Revisar y ajustar la configuración predeterminada.** Audite los permisos y modifique la configuración predeterminada, que suele priorizar la eficiencia sobre la seguridad, para reducir las vulnerabilidades y minimizar los riesgos potenciales.
- **Evaluar y mitigar continuamente los riesgos de la IA.** Evalúe periódicamente los riesgos de seguridad y privacidad relacionados con la IA, así como el comportamiento de los usuarios, para proteger la información de la empresa, la propiedad intelectual y los datos personales.
- **Aplicar Zero Trust a las interacciones de IA.** Adopte una arquitectura Zero Trust que garantice el acceso con privilegios mínimos y restricciones granulares de entrada/salida para evitar el uso no autorizado y minimizar la superficie de ataque.
- **Fortalecer la privacidad y seguridad de los datos.** Implemente cifrado y medidas integrales de prevención de pérdida de datos (DLP) para proteger los datos y proteger la información confidencial de la exposición y las fugas.

Más allá de las mejores prácticas, las empresas deberán establecer pautas formales de IA y reglas de interacción para regir el uso aceptable, la integración, la seguridad y el desarrollo de herramientas de IA.

- **Establecer políticas claras de gobernanza de la IA.** Defina directrices para el uso responsable de la IA, abordando la seguridad, la ética, el cumplimiento normativo y la gestión de riesgos.
- **Realizar la debida diligencia antes de la implementación.** Realice revisiones exhaustivas de seguridad y ética para garantizar que las herramientas se ajusten a las políticas corporativas y a la tolerancia al riesgo.
- **Restringir el intercambio de datos confidenciales.** Evite que los modelos de IA accedan a información personal identificable (PII), datos confidenciales o información empresarial confidencial.
- **Exigir la revisión humana del contenido generado por IA.** Garantice que todo el contenido generado por IA se someta a una revisión humana exhaustiva antes de su publicación.
- **Garantizar la supervisión humana en los procesos impulsados por IA.** Exija la intervención y la revisión humanas para evitar que la IA tome decisiones empresariales cruciales de manera autónoma.
- **Adoptar un marco de ciclo de vida seguro del producto.** Siga un riguroso marco de seguridad para mitigar los riesgos en cada etapa del desarrollo e integración de herramientas de IA.

© 2025 Zscaler, Inc. Todos los derechos reservados.

Cinco pasos para integrar de manera segura las herramientas de GenAI

Un enfoque estratégico y gradual es esencial para adoptar de manera segura aplicaciones de IA. El punto de partida más seguro es bloquear todas las aplicaciones de IA para mitigar posibles fugas de datos. Luego, integre progresivamente herramientas de IA verificadas con estrictos controles de acceso y medidas de seguridad para mantener una supervisión completa de los datos empresariales.

Los siguientes pasos describen un proceso de adopción seguro utilizando ChatGPT de OpenAI como ejemplo.

Paso 1. Bloquear todos los dominios y aplicaciones de IA y ML

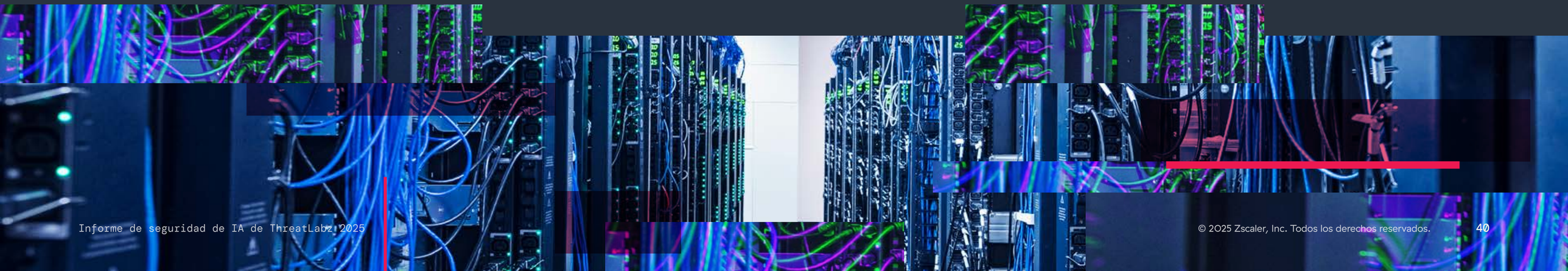
Dado que existen miles de aplicaciones de IA, muchas de ellas con implicaciones de seguridad desconocidas, las empresas deberían adoptar una postura Zero Trust desde el principio. Al bloquear todos los dominios de IA y ML a nivel empresarial, las organizaciones pueden eliminar los riesgos inmediatos y centrarse en adoptar selectivamente solo las herramientas de IA más seguras y transformadoras.

Paso 2. Examinar y aprobar las solicitudes de IA generativa con criterios estrictos.

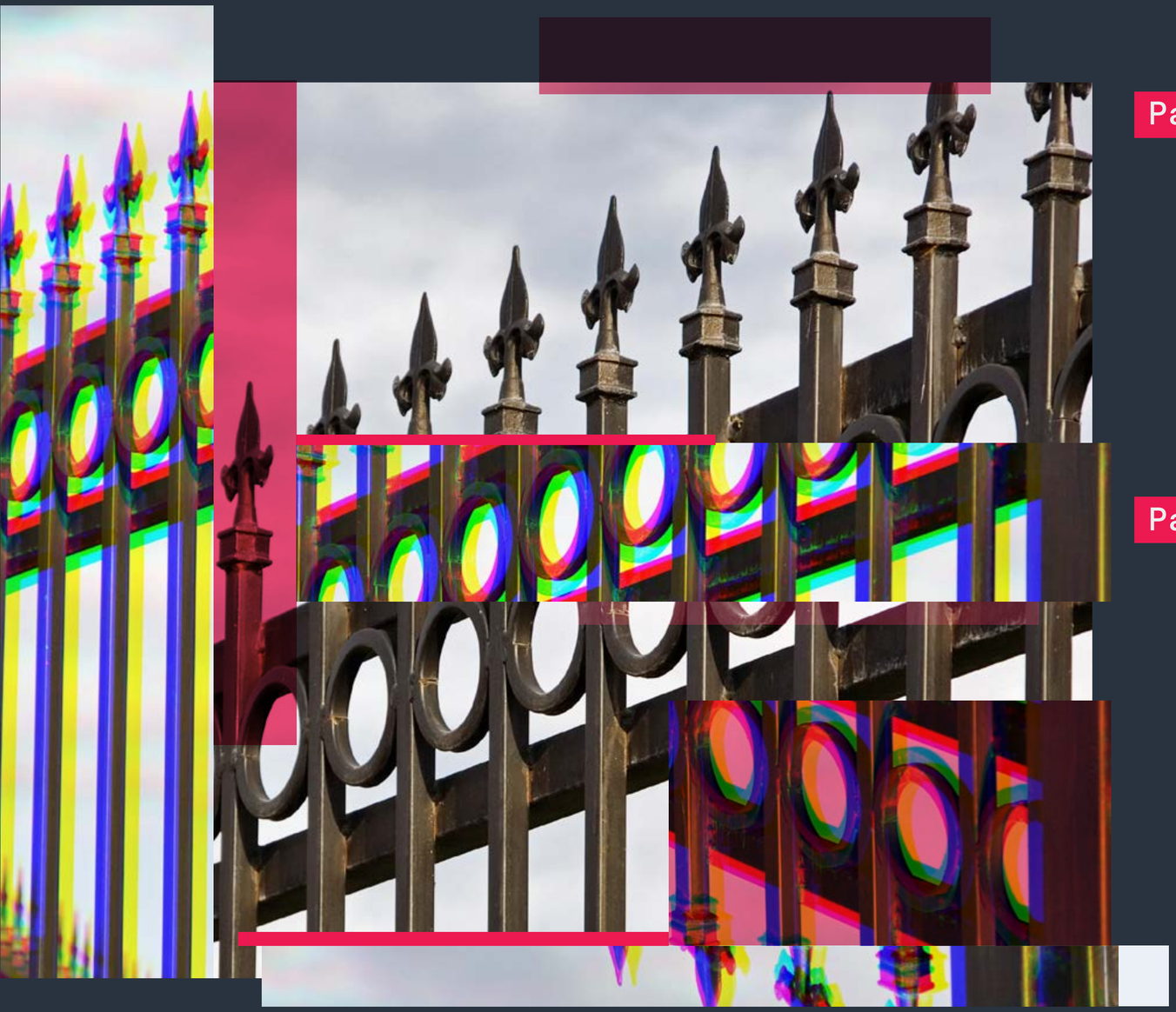
A continuación, identifique y apruebe herramientas de IA que cumplan (o superen) estándares estrictos de seguridad, privacidad y contractuales para proteger los datos comerciales y de los clientes en todo momento y, al mismo tiempo, ofrecer un valor comercial transformador. Para muchas organizaciones, ChatGPT será una aplicación clave que requerirá consideraciones de seguridad adicionales.

Paso 3. Crear una instancia de servidor ChatGPT privada para un control máximo

Para mantener un control total sobre los datos empresariales, las organizaciones deben alojar aplicaciones de IA como ChatGPT en un entorno privado y seguro (por ejemplo, un servidor de IA de Microsoft Azure dedicado) alojado completamente dentro de la organización. Luego, a través de controles de seguridad y obligaciones contractuales, asegúrese de que ni Microsoft ni OpenAI (en este ejemplo) tengan acceso a los datos de la empresa o de los clientes. Este enfoque garantiza la soberanía de los datos y evita que los proveedores de IA manipulen datos confidenciales, lo que impide que se utilicen las consultas para entrenar modelos de IA públicos y reduce el riesgo de envenenamiento de datos desde un lago de datos público.



© 2025 Zscaler, Inc. Todos los derechos reservados.



Paso 4. Asegurar el acceso con SSO, MFA y controles Zero Trust

A continuación, coloque aplicaciones como ChatGPT detrás de una arquitectura de proxy en la nube Zero Trust, como Zscaler Zero Trust Exchange, para aplicar controles de acceso de seguridad Zero Trust. Esto también podría incluir el traslado de ChatGPT detrás de un proveedor de identidad (IdP) para el inicio de sesión único (SSO) con una autenticación multifactor (MFA) fuerte que incluya la autenticación biométrica. Este enfoque garantiza un acceso rápido pero seguro de los usuarios a ChatGPT, al tiempo que permite a las empresas establecer controles de acceso precisos para usuarios individuales, equipos y departamentos. También permite separar claramente las consultas de los usuarios, garantizando que los datos permanezcan aislados y solo sean accesibles dentro de los niveles organizativos adecuados. Al colocar ChatGPT detrás de un proxy en la nube como el Zero Trust Exchange, las organizaciones pueden supervisar e inspeccionar todo el tráfico cifrado TLS/SSL entre los usuarios y ChatGPT para detectar posibles amenazas y evitar fugas de datos.

Paso 5. Implementar la prevención de pérdida de datos (DLP) para evitar fugas

Por último, es fundamental aplicar un motor DLP a la instancia ChatGPT para evitar fugas accidentales de información crítica y garantizar que los datos confidenciales nunca salgan del entorno de producción.

Al seguir estos pasos, las empresas pueden aprovechar el poder de la IA generativa y, al mismo tiempo, eliminar los riesgos más críticos asociados con la adopción de la IA.

© 2025 Zscaler, Inc. Todos los derechos reservados.

Cómo Zscaler ofrece Zero Trust + IA

A medida que crece la adopción de IA, las organizaciones están alcanzando nuevos niveles de productividad, eficiencia e innovación, pero también están ampliando su superficie de ataque. Al mismo tiempo, el auge de la IA armamentística significa amenazas más sofisticadas, automatizadas y evasivas. Las empresas deben reconocer estos riesgos y mejorar las estrategias de seguridad para abordarlos.

Los modelos de seguridad tradicionales son inadecuados en estos entornos de alto riesgo. Sus arquitecturas heredadas, basadas en herramientas como firewalls y VPN, en realidad aumentan el riesgo al expandir la superficie de ataque y permitir el movimiento lateral, lo que permite que los ataques impulsados por la IA se propaguen más rápido. Estas soluciones obsoletas requieren demasiado esfuerzo manual, lo que hace casi imposible proteger las comunicaciones, adaptarse a los riesgos cambiantes y responder a las amenazas en tiempo real.

Para triunfar en la era de la IA, las empresas necesitan un enfoque fundamentalmente nuevo: uno que no solo las defienda contra las amenazas impulsadas por la IA, sino que también permita la adopción segura de la IA. Una arquitectura Zero Trust es la base de ambos.

La arquitectura Zero Trust basada en la nube de Zscaler reduce drásticamente el riesgo al hacer que las aplicaciones y las direcciones IP sean invisibles para los atacantes, minimizando así la superficie de ataque, inspeccionando continuamente todo el tráfico (incluido el tráfico cifrado) en busca de amenazas, evitando vulneraciones y conectando a los usuarios directamente (y solo) a las aplicaciones que necesitan, lo que limita el riesgo de movimiento lateral.

Partiendo de esta base, Zscaler mejora Zero Trust con protecciones contra amenazas impulsadas por la IA para ofrecer una seguridad sin precedentes contra amenazas de todo tipo, incluso los ataques más sofisticados habilitados por IA.

En profundidad: la seguridad de la IA y la ventaja de los datos de Zscaler

La IA es tan inteligente como los datos de los que aprende. Como la mayor nube de seguridad en línea del mundo, Zscaler Zero Trust Exchange asegura a más de 40M de usuarios, cargas de trabajo, dispositivos IoT/OT y acceso de terceros.

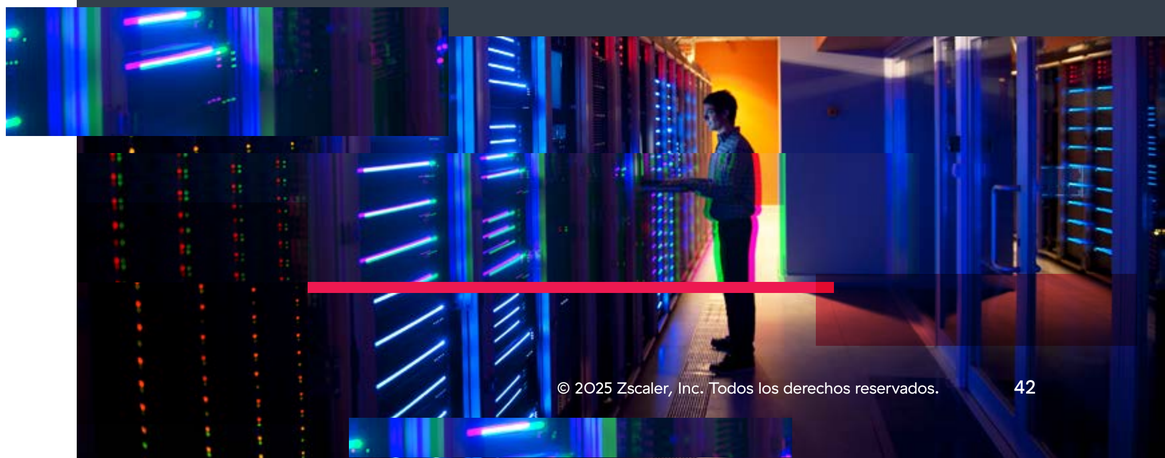
Cada día, Zscaler procesa:

Más de 500 mil billones de señales de telemetría que brindan información en tiempo real sobre amenazas, identidades y patrones de acceso.

Más de 500 mil millones de transacciones: 45 veces el volumen de búsquedas diarias en Google

Este enorme conjunto de datos permite a Zscaler entrenar modelos de IA altamente especializados que identifican y bloquean amenazas con mayor rapidez que los enfoques de seguridad tradicionales, lo que representa más de **9 mil millones de amenazas bloqueadas diariamente**. Al estar integrado entre usuarios, cargas de trabajo y dispositivos, Zscaler ofrece una visibilidad profunda de las ciberamenazas empresariales, lo que hace que sus modelos de IA sean más adaptables, precisos y eficaces.

La estructura de datos de Zscaler también **se integra perfectamente con más de 150 herramientas comerciales y de seguridad,** incluidas **más de 60 fuentes de inteligencia sobre amenazas.**

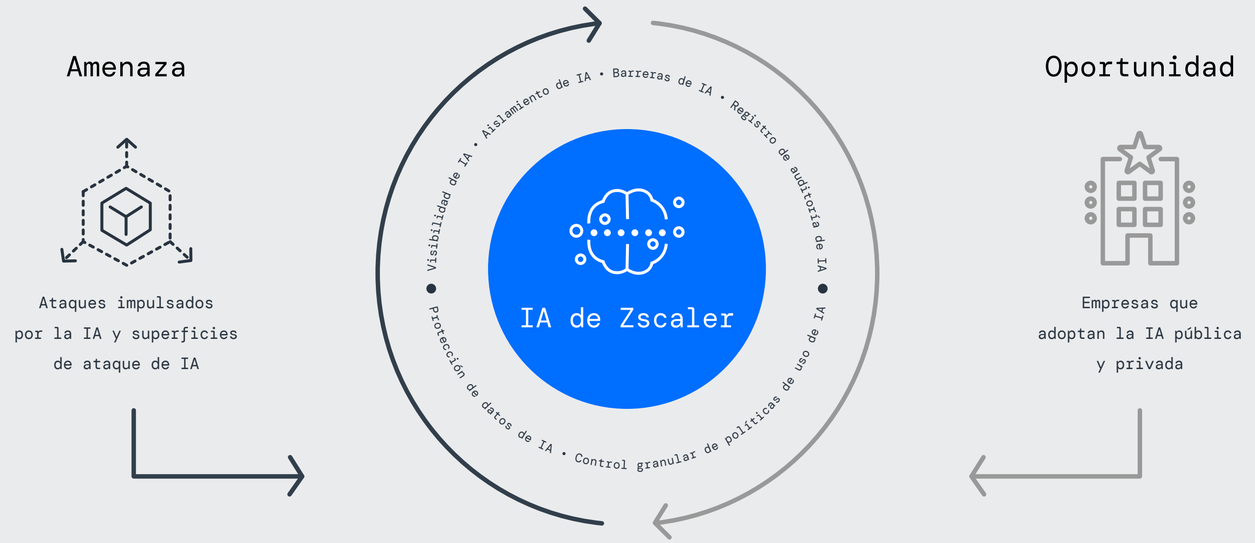


© 2025 Zscaler, Inc. Todos los derechos reservados.

Un enfoque integral para la seguridad de la IA

Para integrar con éxito la IA en la empresa y defenderse de las amenazas potenciadas por IA se necesita una estrategia integral. Con Zscaler Zero Trust + IA, las empresas pueden adoptar con confianza y seguridad la IA pública y privada y, al mismo tiempo, proteger los datos, las aplicaciones y los modelos de IA de las amenazas cambiantes impulsadas por la IA.

Al ofrecer visibilidad completa de los usuarios y las aplicaciones que interactúan con herramientas de IA públicas y privadas, Zscaler AI permite a las empresas implementar políticas contextuales que rigen el acceso y el uso. Su inspección en línea de las indicaciones garantiza la protección de los datos confidenciales y de los propios modelos de IA contra actividades maliciosas y pérdida de datos.



"No teníamos visibilidad hacia [ChatGPT]. Zscaler fue nuestra solución clave inicialmente para ayudarnos a entender quién accedía a ella y qué cargaban".

– Jason Koler, CISO, Eaton Corporation
Vea el video del caso práctico

© 2025 Zscaler, Inc. Todos los derechos reservados.

Zscaler AI permite a las organizaciones:

Habilitar de manera segura el uso de IA pública, maximizando la velocidad del negocio y minimizando los riesgos de la IA oculta y la pérdida de datos.

- **Visibilidad de IA:** vea todas las aplicaciones e interacciones de IA, incluidas las indicaciones y las respuestas.
- **Aislamiento de IA:** permita el uso de herramientas de IA y evite que se compartan datos confidenciales de manera inadvertida.
- **Barreras de IA:** bloquee amenazas como inyecciones rápidas, exposición de información personal, envenenamiento de datos y más.
- **Control granular de políticas de uso de IA:** bloquee aplicaciones de IA no autorizadas o encubiertas y controle el acceso y el uso en función de quién usa la IA y cómo.
- **Protección de datos con IA:** bloquee el intercambio y la exfiltración de datos para evitar violaciones de datos.
- **Registro de auditoría de IA:** mantenga registros detallados de todas las interacciones de IA: usuarios, indicaciones, respuestas y aplicaciones.

Detenga los ataques impulsados por la IA mediante Zero Trust + seguridad impulsada por la IA.

- **Base Zero Trust:** minimice la superficie de ataque externa mediante verificación continua y acceso con privilegios mínimos.
- **Información de IA en tiempo real:** utilice IA predictiva y generativa para brindar información práctica que mejore las operaciones de seguridad y el rendimiento digital.
- **Clasificación de datos:** aproveche la clasificación impulsada por la IA para detectar y proteger sin problemas datos confidenciales en Data Fabric de Zscaler.
- **Protección contra amenazas:** bloquee las amenazas mejoradas con IA a través de la supervisión y la respuesta continuos impulsados por Zscaler Zero Trust Exchange.
- **Segmentación de aplicaciones:** reduzca su superficie de ataque interna y restrinja el movimiento lateral con segmentación automática impulsada por la IA.
- **Predicción de violaciones:** anticipe posibles escenarios de violación utilizando IA generativa y modelos predictivos multidimensionales.
- **Evaluaciones de ciberriesgos:** aproveche los informes de seguridad generados por IA para mapear y optimizar su implementación de Zero Trust.

© 2025 Zscaler, Inc. Todos los derechos reservados.

Las principales capacidades impulsadas por la IA de Zscaler incluyen:

- **Detección de phishing y C2:** identifica y bloquea al instante sitios de phishing nunca vistos y la infraestructura de comando y control (C2) utilizando la detección en línea basada en **IA de Zscaler Secure Web Gateway**.
- **Bloqueo inteligente de indicaciones:** utiliza el filtrado de URL basado en IA/ML en todas las categorías de aplicaciones para tomar decisiones de bloqueo de indicaciones más inteligentes en función del riesgo contextual.
- **Sandbox:** emite veredictos instantáneos sobre amenazas potenciales, lo que previene el malware de día cero y el ransomware antes de que puedan afectar a los usuarios o los puntos finales.
- **Zero Trust Browser:** aísla el contenido sospechoso de Internet y muestra las páginas web como imágenes perfectas, manteniendo el contenido malicioso lejos de los usuarios.
- **Segmentación:** asigna automáticamente las conexiones de usuario a aplicación, lo que simplifica las políticas de acceso Zero Trust para minimizar la superficie de ataque y detener el movimiento lateral.
- **Políticas dinámicas basadas en riesgos:** analiza continuamente el riesgo de usuarios, dispositivos y aplicaciones para aplicar políticas de seguridad adaptables.
- **Predictor de violaciones:** aprovecha algoritmos impulsados por la IA para analizar datos de seguridad, utilizando gráficos de ataques, puntuación de riesgo de usuario e inteligencia de amenazas para predecir posibles violaciones.
- **Evaluaciones de madurez de seguridad:** evalúa continuamente la postura de seguridad Zero Trust, brindando información dinámica y recomendaciones prácticas para reducir aún más el ciberriesgo.
- **Protección de datos:** ofrece detección y clasificación automática de datos con IA en datos de puntos finales, en línea y en la nube. Los controles de prevención de pérdida de datos (DLP) basados en IA garantizan que los datos empresariales confidenciales no se puedan extraer mediante indicaciones de IA.



© 2025 Zscaler, Inc. Todos los derechos reservados.

Aprovechar la seguridad de la IA en toda la cadena de ataque

Zscaler aplica IA en cada etapa de la cadena de ataque, garantizando que las amenazas se detecten y neutralicen antes de que puedan causar daños.

Etapa 1: Detección de la superficie de ataque

El primer paso de un ataque suele ser el reconocimiento: escanear Internet en busca de vulnerabilidades en VPN, firewalls, servidores mal configurados o activos sin parches. La IA ha facilitado este proceso para los malintencionados, permitiéndoles consultar vulnerabilidades conocidas casi instantáneamente.

Cómo Zscaler utiliza la IA para eliminar la superficie de ataque:

- Con información impulsada por la IA de Zscaler Risk360, las empresas pueden mapear y proteger automáticamente sus activos conectados a Internet, haciéndolos invisibles para los usuarios. AI ocultar estos activos detrás de Zero Trust Exchange, las organizaciones reducen drásticamente su superficie de ataque, deteniendo las amenazas incluso antes de que puedan iniciar.



Etapa 2: Riesgo de compromiso

Una vez que los atacantes encuentran una debilidad, intentan explotar vulnerabilidades, robar credenciales u obtener acceso no autorizado. El uso creciente de exploits generados por IA y de correos electrónicos de phishing aumenta aún más el riesgo de compromiso, permitiendo a los atacantes eludir los controles de seguridad tradicionales y haciendo esenciales la detección y la respuesta en tiempo real.

Cómo Zscaler utiliza IA para mitigar el riesgo de vulneración:

- **Los modelos de IA de Zscaler se basan en una combinación de inteligencia de amenazas**, investigación de ThreatLabz y aislamiento del navegador basado en IA para detectar tanto los sitios de phishing conocidos como los de paciente cero, evitando el robo de credenciales y la explotación del navegador. Analizan los patrones de tráfico, el comportamiento y el malware para identificar la infraestructura de mando y control (C2) en tiempo real. Como resultado, las empresas son aún más eficientes y eficaces en la detección de dominios C2 y ataques de phishing.

- **El navegador Zscaler Zero Trust, con tecnología de IA,** reduce automáticamente el riesgo de amenazas web y de día cero, a la vez que garantiza que los empleados puedan acceder a los sitios web adecuados para realizar su trabajo. AI Smart Isolation identifica contenido sospechoso de internet y lo abre en un entorno seguro y aislado. Esto detiene eficazmente amenazas web como malware, ransomware y phishing.

- **Zscaler Cloud Sandbox** detecta, previene y pone en cuarentena de manera inteligente amenazas desconocidas y archivos sospechosos en línea. Gracias a veredictos basados en IA, los archivos benignos se distribuyen al instante, mientras que los archivos maliciosos se bloquean para todos los usuarios globales de Zscaler. Esto impide eficazmente que amenazas web como malware, ransomware, phishing y descargas no autorizadas accedan a la red.

© 2025 Zscaler, Inc. Todos los derechos reservados.

Etapa 3: Movimiento lateral

Una vez dentro, los atacantes tratan de desplazarse lateralmente dentro de una organización, buscando privilegios elevados o datos o aplicaciones valiosos. Cada vez más, utilizan herramientas de inteligencia artificial para trazar rápidamente vías que les permitan un compromiso más profundo. Muchas empresas también tienen un exceso de derechos de acceso, lo que facilita a los atacantes moverse por los entornos sin ser detectados.

Cómo Zscaler utiliza la IA para prevenir el movimiento lateral:

- Zscaler AI analiza continuamente el comportamiento del usuario y los patrones de acceso, recomendando políticas de segmentación de aplicaciones inteligentes para limitar el movimiento lateral. Por ejemplo, si solo 200 de 30,000 empleados necesitan acceso a una aplicación, Zscaler puede segmentar automáticamente el acceso solo a esos usuarios, lo que reduce el riesgo de movimiento lateral en más de 90 %.

Etapa 4: Exfiltración de datos

La etapa final de un ataque es la exfiltración de datos, donde los atacantes intentan robar datos como IP, información de clientes o registros financieros.

Cómo Zscaler utiliza IA para detener la pérdida de datos:

- La detección de datos potenciada por IA acelera la visibilidad de los datos y automatiza la clasificación de datos en tiempo real en toda la empresa. Habilite instantáneamente políticas de prevención de pérdida de datos (DLP) para evitar que los datos salgan de la organización.

Asegurar la IA para 2025: un llamado a la acción

La IA es una fuerza de progreso, disrupción y riesgo que impulsa a las organizaciones empresariales a adaptarse en todo momento. Seguirá desbloqueando nuevas eficiencias e innovaciones, pero también introducirá nuevas amenazas, desde los ciberataques impulsados por la IA hasta la manipulación adversaria de modelos y datos. Para aprovechar de manera segura todo el potencial de la IA y mitigar sus riesgos, las empresas deben recurrir a Zero Trust + IA.

La seguridad de IA de Zscaler protege cada etapa de la adopción de IA y garantiza la protección en cada etapa de un ataque. Al adoptar un enfoque proactivo, las organizaciones pueden convertir la IA en una ventaja competitiva, desbloqueando nuevas posibilidades y manteniéndose a la vanguardia de las amenazas cambiantes.

© 2025 Zscaler, Inc. Todos los derechos reservados.

Metodología de la investigación

Las conclusiones se basan en el análisis de 536,500 millones de transacciones totales de IA y ML en la nube de Zscaler desde febrero de 2024 hasta diciembre de 2024. La nube de seguridad global de Zscaler procesa más de 500 billones de señales diarias y bloquea 9000 millones de amenazas e infracciones de políticas al día, proporcionando más de 250 000 actualizaciones de seguridad diarias.

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**.

Acerca de Zscaler

Zscaler (NASDAQ: ZS) acelera la transformación digital para que los clientes puedan ser más ágiles, eficientes, resistentes y seguros. Zero Trust Exchange™ protege a miles de clientes contra ciberataques y pérdida de datos al conectar usuarios, dispositivos y aplicaciones de manera segura en cualquier ubicación. Distribuido en más de 160 centros de datos en todo el mundo, Zero Trust Exchange, basado en SASE, es la plataforma de seguridad en la nube en línea más grande del mundo. Para obtener más información, visite **www.zscaler.com/mx.**

© 2025 Zscaler, Inc. Todos los derechos reservados.



© 2025 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.