



Informe de seguridad de IA de Zscaler ThreatLabz 2024



La revolución de la IA ha llegado. Descubra las tendencias clave, los riesgos y las mejores prácticas en la adopción de la IA por parte de las empresas, con información sobre las amenazas impulsadas por la IA y las estrategias clave para defenderse de ellas.

Contenido

03 Resumen ejecutivo

04 Hallazgos clave

05 Tendencias clave de uso de GenAI y ML

- 05 Las transacciones de IA continúan acelerándose
- 06 Las empresas están bloqueando más transacciones de IA que nunca
- 07 **Desglose de la IA en la industria**
 - 09 Atención sanitaria e IA
 - 10 Finanzas
 - 11 Gobierno
 - 12 Fabricación
 - 13 Educación e IA
- 14 **Tendencias de uso de ChatGPT**
- 15 **Uso de IA por país**
 - Desglose regional: Desglose regional de EMEA: APAC

18 Escenarios de riesgo de IA empresarial y amenazas del mundo real

- 18 Habilitar la IA en la empresa: los 3 riesgos principales
- 20 Escenarios de amenaza impulsados por la IA Suplantación de identidad mediante IA: deepfakes, desinformación y más
- 21 Campañas de phishing generadas por IA De la consulta al delito: creación de una página de inicio de sesión de phishing con ChatGPT
- 22 Chatbots oscuros: descubriendo WormGPT y FraudGPT en la web oscura

- 23 Malware y ransomware impulsados por IA en toda la cadena de ataque
- 24 Ataques de gusanos de IA y jailbreak “viral” de IA
- 25 La IA y las elecciones estadounidenses

26 Todas las miradas puestas en la normativa sobre IA

- 26 Estados Unidos
- 27 Unión Europea

28 Predicciones de amenazas de IA

31 Estudio de caso: Cómo habilitar ChatGPT de manera segura en la empresa

- 31 5 pasos para integrar y proteger herramientas de IA generativa

33 Cómo Zscaler ofrece IA+ Zero Trust y protege la IA generativa

- 33 La clave para la ciberseguridad impulsada por la IA: datos de alta calidad a escala
- 34 Aprovechar la IA en toda la cadena de ataque
- 35 Resumen de las ofertas de Zscaler basadas en IA
- 36 Habilitar la transición a la IA empresarial: el control está en sus manos

37 Apéndice

- 37 Metodología de investigación de ThreatLabz

37 Acerca de Zscaler ThreatLabz

Resumen ejecutivo

La IA es más que una innovación pionera: ahora es algo habitual. A medida que las herramientas de IA generativa como ChatGPT transforman los negocios de maneras grandes y pequeñas, la IA se está integrando profundamente en el tejido de la vida empresarial. Sin embargo, las preguntas sobre cómo adoptar de manera segura estas herramientas de IA mientras se defiende contra las amenazas impulsadas por la IA no están resueltas.

Las empresas están adoptando rápidamente herramientas de inteligencia artificial y aprendizaje automático en departamentos como ingeniería, marketing de TI, finanzas, éxito del cliente y más. Sin embargo, deben equilibrar los numerosos riesgos que conllevan las herramientas de inteligencia artificial para obtener los máximos beneficios. De hecho, para desbloquear el potencial transformador de la IA, las empresas deben habilitar controles seguros para proteger sus datos, evitar la fuga de información confidencial, mitigar la expansión de la "IA en la sombra" (Shadow AI) y garantizar la calidad de los datos de la IA.

Estos riesgos de la IA para las empresas son bidireccionales: **Fuera de los muros de las empresas, la IA se ha convertido en una fuerza impulsora de las ciberamenazas.** De hecho, las herramientas de inteligencia artificial están permitiendo a los ciberdelincuentes y a los malintencionados patrocinados por estados nacionales lanzar ataques sofisticados, más rápidamente y a mayor escala. A pesar de esto, la IA es prometedora como pieza clave del rompecabezas de la ciberdefensa a medida que las empresas se enfrentan a un panorama dinámico de amenazas.

El Informe de seguridad de IA de ThreatLabz 2024 ofrece información clave sobre estos desafíos y oportunidades críticos de la IA.

Basándose en más de 18 mil millones de transacciones desde abril de 2023 hasta enero de 2024 en Zscaler Zero Trust Exchange™, ThreatLabz analizó cómo las empresas utilizan las herramientas de inteligencia artificial y aprendizaje automático en la actualidad. Estos conocimientos revelan tendencias clave en todos los sectores empresariales y geografías sobre cómo las empresas se están adaptando al cambiante panorama de la IA y asegurando sus herramientas de IA.

A lo largo, encontrará información sobre los temas más importantes de la IA, incluidos el riesgo empresarial, los escenarios de amenazas impulsadas por la IA y las tácticas de los adversarios, consideraciones regulatorias y predicciones para el panorama de la IA en 2024 y más allá.

De manera igualmente crítica, este informe ofrece mejores prácticas en dos frentes: cómo las empresas pueden adoptar de manera segura la transformación de la IA generativa mientras protegen los datos críticos, y cómo las herramientas impulsadas por la IA están funcionando para brindar seguridad en capas y Zero Trust para enfrentar el nuevo panorama de amenazas impulsadas por la IA.

Hallazgos clave



El uso de herramientas de IA/ML se disparó un 594.82 %, pasando de 521 millones de transacciones impulsadas por IA/ML en abril de 2023 a 3100 millones mensuales en enero de 2024.



Las empresas están bloqueando el 18.5 % de todas las transacciones de IA/ML (un aumento del 577 % de las transacciones bloqueadas en nueve meses), lo que refleja la creciente preocupación en torno a la seguridad de los datos de IA y la reticencia de las empresas a establecer políticas de IA.



El sector de fabricación es el que genera más tráfico de IA, con un 20.9 % de todas las transacciones de IA/ML en la nube de Zscaler, seguida por Finanzas y Seguros (19.9 %) y Servicios (16.8 %).



El uso de ChatGPT continúa aumentando, con un crecimiento del 634.1 %, a pesar de que **también es la aplicación de IA más bloqueada** por las empresas, según los conocimientos de la nube de Zscaler.



Las aplicaciones de IA más utilizadas por volumen de transacciones son ChatGPT, Drift, OpenAI*, Writer y LivePerson. Las tres principales aplicaciones bloqueadas por volumen de transacciones son ChatGPT, OpenAI y Fraud.net.



Los cinco países que generan más transacciones de IA y ML son los EE. UU., India, Reino Unido, Australia y Japón.



Las empresas están enviando importantes volúmenes de datos a las herramientas de IA, con un total de 569 TB intercambiados entre las aplicaciones de IA/ML entre septiembre de 2023 y enero de 2024.



La IA está empoderando a los malintencionados de maneras nunca vistas, incluso para campañas de phishing impulsadas por IA, deepfakes y ataques de ingeniería social, ransomware polimórfico, descubrimiento de superficie de ataque empresarial, generación automatizada de exploits y mucho más.

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

Tendencias clave del uso de GenAI y ML

La revolución de la IA empresarial está lejos de su apogeo. Las transacciones empresariales de IA han aumentado casi un 600 % y no muestran signos de desaceleración. Aún así, las transacciones bloqueadas a aplicaciones de inteligencia artificial también han aumentado en 577 %.

Las transacciones de IA continúan acelerándose

Desde abril de 2023 hasta enero de 2024, las transacciones empresariales de IA y ML crecieron casi un 600 %, aumentando a más de 3 mil millones de transacciones mensuales en Zero Trust Exchange en enero. Esto subraya el hecho de que, a pesar del creciente número de incidentes de seguridad y riesgos para los datos asociados a la adopción de la IA por parte de las empresas, su potencial transformador es demasiado grande como para ignorarlo. Tenga en cuenta que, si bien las transacciones de IA experimentaron una breve pausa durante las fiestas de diciembre, las transacciones continuaron a un ritmo aún mayor a principios de 2024.

Sin embargo, a pesar de que las aplicaciones de la IA siguen multiplicándose, la mayoría de las transacciones de IA están siendo impulsadas por un conjunto relativamente pequeño de herramientas de IA líderes en el mercado. En general, ChatGPT representa más de la mitad de todas las transacciones de IA y ML, mientras que la aplicación OpenAI ocupa el tercer lugar, con el 7.82 % de todas las transacciones. Mientras tanto, Drift, el popular chatbot impulsado por IA, generó casi una quinta parte del tráfico empresarial de IA (los chatbots LivePerson y BoldChat Enterprise también superaron las principales aplicaciones en los puestos 5 y 6). Mientras tanto, Writer sigue siendo una herramienta de IA generativa favorita en la creación de contenido empresarial escrito, como materiales de marketing. Finalmente, Otter, una herramienta de transcripción de IA que se utiliza a menudo en videollamadas, genera una parte importante del tráfico de IA.

Tendencias de transacciones de IA y ML



FIGURA 1 Transacciones de IA desde abril de 2023 hasta enero de 2024

Principales aplicaciones de la IA

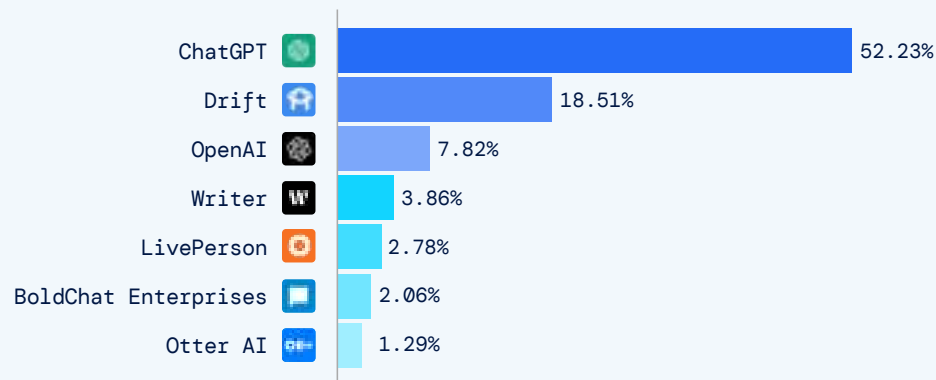


FIGURA 2 Principales aplicaciones de la IA por volumen de transacciones

Datos transferidos por Tráfico de IA/ML [Sep 2023–Ene 2024]

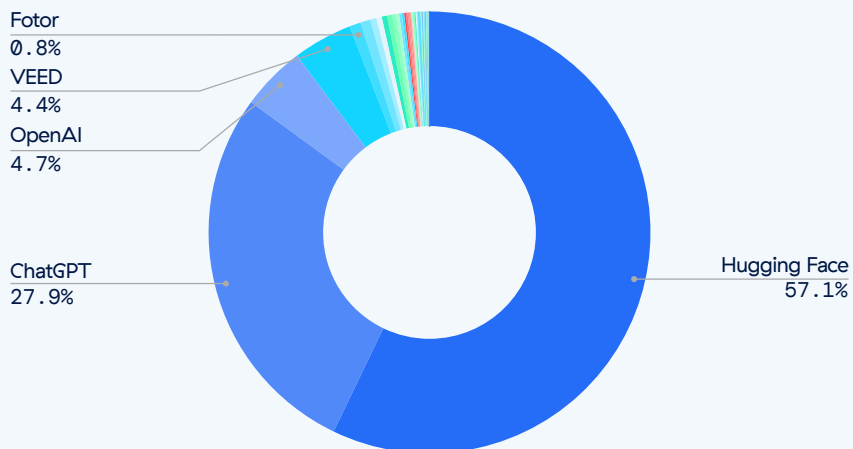


FIGURA 3 Principales aplicaciones de IA/ML según el porcentaje de datos totales transferidos

Mientras tanto, los volúmenes de datos que las empresas envían y reciben de las herramientas de IA agregan matices a estas tendencias. Hugging Face, la plataforma de código abierto para desarrolladores de IA a menudo descrita como “el GitHub de la IA”, representa casi el 60 % de los datos empresariales transferidos mediante herramientas de IA. Dado que Hugging Face permite a los usuarios alojar y entrenar modelos de IA, tiene sentido que capture importantes volúmenes de datos de usuarios empresariales.

Si bien ChatGPT y OpenAI hacen apariciones esperadas en esta lista, dos adiciones notables son Veed (un editor de video de IA que se usa a menudo para agregar subtítulos, imágenes y otro texto a los videos) y Fotor, una herramienta utilizada para generar imágenes de IA, entre otros usos. Dado que los videos y las imágenes conllevan archivos de gran tamaño en comparación con otros tipos de solicitudes, no es sorprendente ver representadas estas dos aplicaciones.

Las empresas están bloqueando más transacciones de IA que nunca

A pesar de que la adopción de la IA empresarial continúa aumentando, las organizaciones bloquean cada vez más las transacciones de IA y ML debido a preocupaciones sobre los datos y la seguridad. Hoy en día, las empresas bloquean el 18.5 % de todas las transacciones de IA, un aumento del 577 % de abril a enero, para un total de más de 2600 millones de transacciones bloqueadas.

Algunas de las herramientas de inteligencia artificial más populares también son las más bloqueadas. De hecho, ChatGPT tiene la distinción de ser la aplicación de IA más utilizada y más bloqueada. Esto indica que a pesar de la popularidad de estas herramientas, o incluso debido a ella, las empresas están trabajando activamente para proteger su uso contra la pérdida de datos y los problemas de privacidad. Otra tendencia notable es que [bing.com](https://www.bing.com), que tiene una funcionalidad Copilot habilitada para IA, está bloqueado de abril a enero. De hecho, [bing.com](https://www.bing.com) representa el 25.02 % de todas las transacciones de dominios de IA y ML bloqueadas.

Algunas de las herramientas de inteligencia artificial más populares también son las más bloqueadas. De hecho, ChatGPT tiene la distinción de ser la aplicación de IA más utilizada y más bloqueada. Esto indica que a pesar de la popularidad de estas herramientas, o incluso debido a ella, las empresas están trabajando activamente para proteger su uso contra la pérdida de datos y los problemas de privacidad. Otra tendencia notable es que [bing.com](https://www.bing.com) está bloqueado más que cualquier otro dominio, con un total de 835,811,952 bloqueos de abril a enero. De hecho, [bing.com](https://www.bing.com) representa el 25.02 % de todas las transacciones de dominios de IA y ML bloqueadas.

Tendencias de las transacciones de IA bloqueadas [Abr 2023 – Ene 2024]



FIGURA 4 Número de transacciones de IA/ML bloqueadas a lo largo del tiempo

LAS HERRAMIENTAS DE IA MÁS BLOQUEADAS	DOMINIOS IA MÁS BLOQUEADOS
01 ChatGPT	01 Bing.com
02 OpenAI	02 Divo.ai
03 Fraud.net	03 Drift.com
04 Forethought	04 Quillbot.com
05 Hugging Face	05 Compose.ai
06 ChatBot	06 Openai.com
07 Aivo	07 Qortex.ai
08 Neeva	08 Sider.ai
09 infeedo.ai	09 Tabnine.com
10 Jasper	10 securiti.ai

FIGURA 5 Principales aplicaciones y dominios de IA bloqueados por volumen de transacciones

Desglose de la IA en la industria

Los sectores verticales de la industria empresarial muestran diferencias notables en su adopción general de herramientas de IA, así como en la proporción de transacciones de IA que bloquean. El sector de la fabricación es el líder indiscutible, impulsando más del 20 % de las transacciones de IA y ML a través de Zero Trust Exchange. Aún así, los sectores de finanzas y seguros, tecnología y servicios les siguen de cerca. Juntas, estos cuatro sectores se han posicionado por delante de los demás como los más agresivos en la adopción de la IA.

Proporción de transacciones de IA por sector vertical

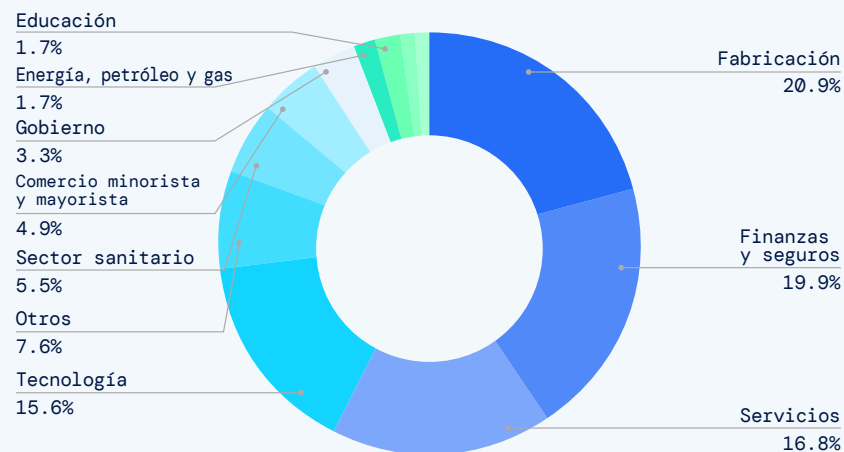


FIGURA 6 Sectores que impulsan las mayores proporciones de transacciones de IA

Tendencias de transacciones de IA por Vertical

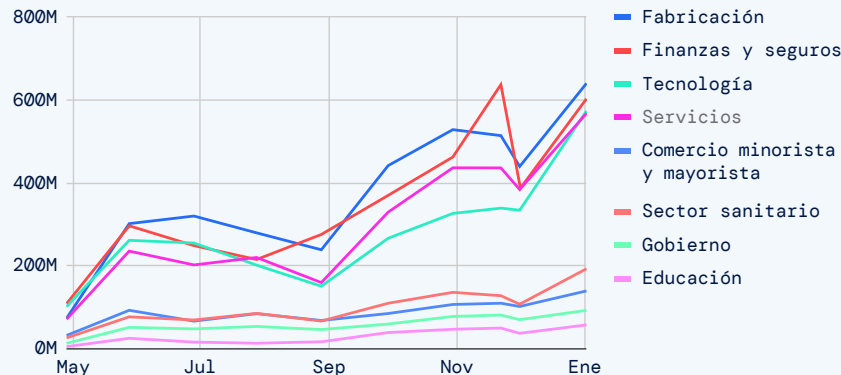


FIGURA 7 Tendencias de las transacciones de IA/ML entre los sectores de mayor volumen, abril de 2023-enero de 2024

Garantizar las transacciones de IA/ML

Junto con el fuerte aumento de las transacciones de IA, los sectores industriales están bloqueando más transacciones de IA. En este caso, ciertos sectores divergen de sus tendencias generales de adopción, lo que refleja diferentes prioridades y niveles de madurez en términos de seguridad de las herramientas de IA. El sector financiero y de seguros, por ejemplo, bloquea la mayor proporción de transacciones de IA: 37.2 % frente al promedio global de 18.5 %. Es probable que esto se deba en gran parte al estricto entorno regulatorio y de cumplimiento de la industria, combinado con los datos financieros y personales de los usuarios altamente confidenciales que procesan estas organizaciones.

Mientras tanto, el sector de la fabricación bloquea el 15.7 % de las transacciones de IA, a pesar de su papel preponderante en el impulso de las transacciones generales de IA. El sector tecnológico, uno de los primeros y más entusiastas en adoptar la IA, ha tomado una especie de camino intermedio, bloqueando un 19.4 % superior al promedio de las transacciones de IA mientras trabaja para escalar la adopción de la IA. Sorprendentemente, el sector sanitario bloquea un 17.2 % de las transacciones de IA, por debajo de la media, a pesar de que estas organizaciones procesan una gran cantidad de datos sanitarios e información personal identificable (PII). Es probable que esta tendencia refleje un esfuerzo tardío entre las organizaciones sanitarias para proteger los datos confidenciales implicados en las herramientas de IA, ya que los equipos de seguridad tratan de ponerse al día con la innovación de la IA. En general, las transacciones de IA en el sector sanitario siguen siendo comparativamente bajas.

FIGURA 8

Principales verticales de la industria por porcentaje de transacciones de IA bloqueadas

Porcentaje de transacciones de IA bloqueadas por Vertical

Vertical	% de transacciones de IA bloqueadas
Finanzas y seguros	37.16
Fabricación	15.65
Servicios	13.17
Tecnología	19.36
Sector sanitario	17.23
Comercio minorista y mayorista	10.52
Otros	8.93
Energía, petróleo y gas	14.24
Gobierno	6.75
Transporte	7.90
Educación	2.98
Comunicación	4.29
Construcción	4.12
Materiales básicos, productos químicos y minería	2.92
Entretenimiento	1.33
Alimentos, bebidas y tabaco	3.66
Hoteles, restaurantes y ocio	3.16
Organizaciones religiosas	6.06
Agricultura y Silvicultura	0.18
Promedio en todas las verticales	18.53



Atención sanitaria e inteligencia artificial

Situado como el sexto mayor usuario de IA/ML, el sector sanitario bloquea el 17.23 % de todas las transacciones de IA/ML.

LAS MEJORES APLICACIONES DE IA EN ATENCIÓN SANITARIA SON:

- | | |
|-------------|---------------|
| 01 ChatGPT | 06 Zineone |
| 02 Drift | 07 Securiti |
| 03 OpenAI | 08 Pypestream |
| 04 Writer | 09 Híbrido |
| 05 Intercom | 10 VEED |

Signos vitales de progreso en la asistencia sanitaria con IA

Aunque el sector sanitario suele ser cauteloso a la hora de poner en práctica innovaciones como la IA, como lo demuestra su actual contribución del 5 % al tráfico de IA/ML en la nube de Zscaler, es solo cuestión de tiempo antes de que la IA tenga un mayor impacto en las operaciones sanitarias, la atención al paciente y la investigación e innovación médicas.¹

De hecho, la IA promete ayudar no solo a ahorrar tiempo, sino también a salvar vidas. Las tecnologías impulsadas por la IA ya están mejorando el diagnóstico y la atención al paciente. Al analizar imágenes médicas con notable precisión, la IA ayuda a los radiólogos a detectar anomalías más rápidamente y facilita decisiones de tratamiento más rápidas.²

Los beneficios potenciales son enormes. Los algoritmos de la IA pueden utilizar datos de pacientes para personalizar los planes de tratamiento y acelerar el descubrimiento de fármacos mediante el análisis eficiente de datos biológicos. Las tareas administrativas también se pueden automatizar con IA generativa, aliviando la carga de los equipos sanitarios con poco personal. Estos avances subrayan la capacidad de la IA para transformar la prestación de servicios de salud y la prestación de servicios de salud.

Riesgos clave para la atención sanitaria:

Las organizaciones de atención sanitaria deben reconocer los riesgos y desafíos potenciales asociados con la IA, incluidas las preocupaciones sobre la privacidad y la seguridad de los datos, especialmente para la información de identificación personal (PII), así como garantizar que los algoritmos de IA y sus resultados sean altamente confiables e imparciales cuando son en ayuda de la administración de la atención al paciente.



1. Statista, [Casos de uso futuros de la IA en la atención sanitaria](#), septiembre de 2023.

2. The Hill, [La IA ya desempeña un papel vital en las imágenes médicas y está regulada de manera eficaz](#), 23 de febrero de 2024.



Finanzas e IA

En segundo lugar en cuanto al uso total de IA/ML, el sector financiero bloquea el 37.16 % de todo el tráfico de IA/ML.

LAS MEJORES APLICACIONES DE IA EN FINANZAS SON:

- | | |
|------------------------|-----------------|
| 01 ChatGPT | 06 Writer |
| 02 Drift | 07 Hugging Face |
| 03 OpenAI | 08 Otter Ai |
| 04 BoldChat Enterprise | 09 Securiti |
| 05 LivePerson | 10 Intercom |

Las instituciones financieras apuestan a la IA

Las empresas de servicios financieros han sido las primeras en adoptar la era de la IA, y el sector representa casi una cuarta parte del tráfico de IA/ML en la nube de Zscaler. Es más, McKinsey proyecta unos ingresos anuales potenciales de entre 200,000 y 340,000 millones de dólares a partir de iniciativas de IA generativa en la banca, impulsadas en gran medida por el aumento de la productividad.³ La IA representa, literalmente, una gran cantidad de oportunidades para los bancos y los servicios financieros.

Si bien los chatbots y los asistentes virtuales impulsados por IA no son nada nuevo en las finanzas (el “Erica” de Bank of America se lanzó en 2018), las mejoras de la IA generativa están elevando estas herramientas de servicio al cliente a nuevos niveles de personalización. Otras capacidades de la IA, como el modelado predictivo y el análisis de datos, están preparadas para ofrecer enormes ventajas de productividad a las operaciones financieras, transformando la detección de fraude, las evaluaciones de riesgos y más.

Principales riesgos para las finanzas y los seguros:

La integración de la IA en los servicios y productos financieros también plantea preocupaciones de seguridad y regulatorias sobre la privacidad de los datos, los sesgos y la precisión. El considerable 37 % de tráfico de IA/ML bloqueado del que informa ThreatLabz refleja esta perspectiva. Abordar estas preocupaciones requerirá una supervisión y una planificación astutas para mantener la confianza y la integridad en la banca, los servicios financieros y los seguros.

3. McKinsey, [Captura del valor total de la IA generativa en la banca](#), 5 de diciembre de 2023.

Gobierno e IA

Aunque se encuentra entre los 10 primeros en utilización de IA/ML, el sector gubernamental solo bloquea el 6.75 % de las transacciones de IA/ML.

LAS PRINCIPALES APLICACIONES DE IA* EN EL GOBIERNO SON:

- 01 ChatGPT
- 02 Drift
- 03 OpenAI
- 04 Zineone

*Aplicaciones de IA con al menos 1M de transacciones

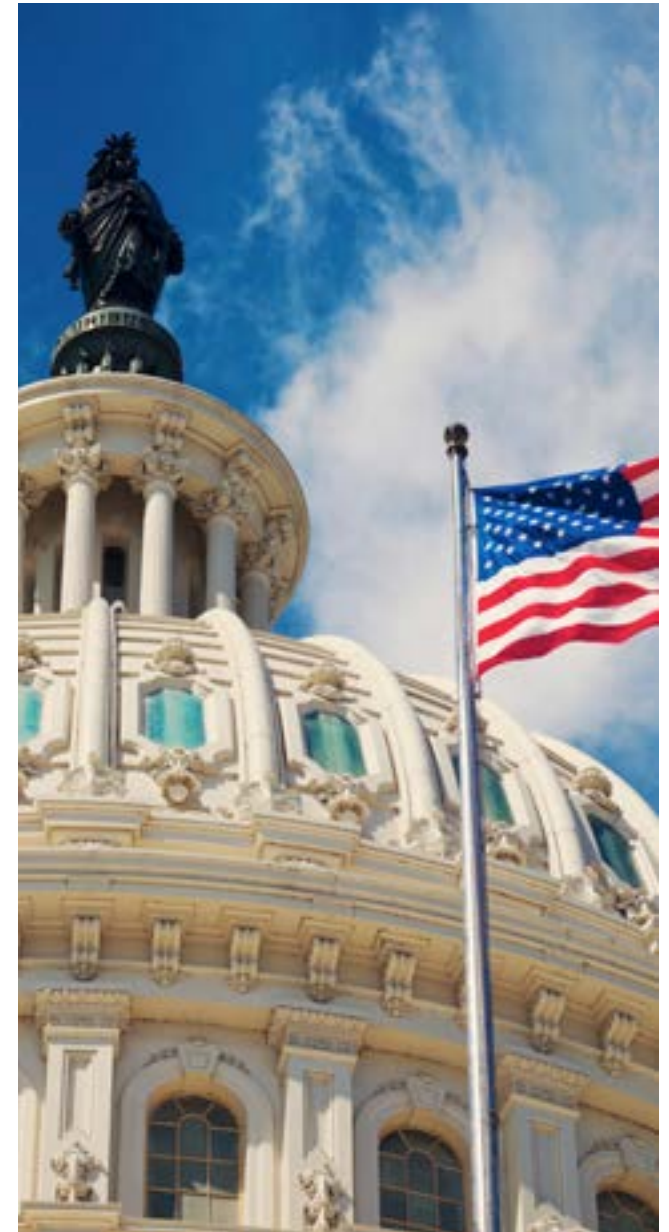
Los gobiernos mundiales exploran las prácticas y políticas de la IA

Han surgido dos debates críticos sobre IA en el gobierno: uno sobre la implementación de tecnologías de IA y otro sobre el establecimiento de una gobernanza para gestionarlas de manera segura. Las ventajas de la adopción de la IA por parte de entidades gubernamentales y del sector público son sustanciales, particularmente cuando los chatbots y los asistentes virtuales pueden brindar a los ciudadanos un acceso más rápido a información y servicios esenciales en sectores como el transporte público y la educación. El análisis de datos basado en inteligencia artificial puede ayudar a abordar los desafíos sociales a través de procesos de toma de decisiones basados en datos, lo que lleva a un desarrollo de políticas y una asignación de recursos más eficientes.

Ya se están produciendo avances notables. Por ejemplo, el Departamento de Justicia de los Estados Unidos nombró a su primer director de IA, confirmando su compromiso con el uso de sistemas de IA. Los datos de ThreatLabz indican que los clientes gubernamentales utilizan cada vez más plataformas de IA/ML como ChatGPT y Drift.

Riesgos gubernamentales clave:

A pesar de estas tendencias, las preocupaciones clave sobre los riesgos relacionados con la IA y la privacidad de los datos subrayan la necesidad continua de marcos regulatorios y gobernanza en todas las organizaciones federales. En general, los responsables políticos de todo el mundo han dado pasos significativos hacia la regulación de la IA durante el año pasado, lo que indica un esfuerzo colectivo para impulsar el desarrollo y la implementación responsables de tecnologías de IA/ML.





Sector de fabricación e IA

Como primer vertical de IA/ML, el vertical del sector de la fabricación bloquea el 15.65 % de todas las aplicaciones de IA/ML.

LAS PRINCIPALES APLICACIONES SON:

- | | |
|-------------|------------------|
| 01 ChatGPT | 06 Google Search |
| 02 Drift | 07 Zineone |
| 03 OpenAI | 08 Pypestream |
| 04 Writer | 09 Hugging Face |
| 05 Securiti | 10 Fotor |

El sector de la fabricación aprovecha el impulso de la IA

Como era de esperar, la mayor afluencia de tráfico de IA/ML (18.2 %) en nuestra investigación procede de clientes del sector de la fabricación. La adopción de la IA en el sector de la fabricación es una piedra angular de la Industria 4.0, también conocida como la Cuarta Revolución Industrial, una era marcada por la convergencia de tecnologías digitales y procesos industriales.

Desde la detección preventiva de fallas en los equipos mediante el análisis de grandes cantidades de datos de maquinaria y sensores hasta la optimización de la gestión de la cadena de suministro, el inventario y las operaciones logísticas, la IA está resultando fundamental para los fabricantes. Además, los sistemas de automatización y robótica impulsados por la IA pueden mejorar significativamente la eficiencia de la fabricación. Pueden ejecutar tareas a mucha mayor velocidad y precisión que los humanos, y al mismo tiempo reducen costos y errores.

Principales riesgos de la IA en el sector de la fabricación:

En cuanto al 16 % de tráfico bloqueado de aplicaciones de IA/ML por parte de los clientes del sector de la fabricación, algunos fabricantes están abordando la IA/ML generativa con cautela. Esto puede deberse a la preocupación por la seguridad de los datos de las organizaciones de fabricación, así como a la necesidad de examinar y aprobar selectivamente un conjunto más reducido de aplicaciones de IA, al tiempo que se bloquean las aplicaciones que entrañan un mayor riesgo.

Educación e IA

Ocupando el undécimo lugar en el uso global de la IA/ML, el vertical de educación bloquea el 2.98 % de todo el tráfico de IA/ML.

LAS PRINCIPALES APLICACIONES SON:

- | | |
|-----------------|-----------|
| 01 ChatGPT | 05 Deepai |
| 02 Character.AI | 06 Drift |
| 03 Pixlr | 07 OpenAI |
| 04 Forethought | |

La educación adopta la IA como herramienta de aprendizaje

Si bien el sector educativo no es uno de los principales productores de tráfico de IA, bloquea un porcentaje comparativamente bajo (2.98 %) de transacciones de IA y ML: aproximadamente 9 millones, de un total de más de 309 millones de transacciones. Está claro que, a pesar de la narrativa popular de que las instituciones educativas suelen bloquear aplicaciones de IA como ChatGPT entre los estudiantes, el sector ha adoptado principalmente las aplicaciones de IA como herramientas de aprendizaje. En particular, cinco de las aplicaciones de IA más populares en educación (ChatGPT, Character.AI, Pixlr y OpenAI) se centran explícita o frecuentemente en resultados creativos para la escritura y la generación de imágenes; mientras tanto, Forethought puede usarse como una ayuda instructiva de chatbot.

Para agregar matices a esta narrativa, también puede ser que muchos educadores bloqueen herramientas como ChatGPT como una cuestión de política en el aula, pero que las instituciones educativas se hayan quedado rezagadas con respecto a otros sectores en la implementación de soluciones tecnológicas como el filtrado DNS que permite a las organizaciones bloquear herramientas de IA y ML de modos más específicos.

Riesgos clave de la IA en la educación:

En la educación, las preocupaciones sobre la privacidad de los datos probablemente aumentarán a medida que el sector continúe adoptando herramientas de IA, específicamente en torno a las protecciones otorgadas a los datos personales de los estudiantes. Con toda probabilidad, el sector educativo adoptará cada vez más medios tecnológicos para bloquear aplicaciones selectivas de IA, al tiempo que proporcionará mayores medidas de protección de datos personales.



Tendencias de uso de ChatGPT

La adopción de ChatGPT se ha disparado. Desde abril de 2023, las transacciones mundiales de ChatGPT crecieron más de un 634 %, un ritmo notablemente más rápido que el aumento global del 595 % de las transacciones de IA. A partir de estos hallazgos y de la amplia percepción de la industria de OpenAI como la principal marca de IA, queda claro que ChatGPT es la herramienta de IA generativa favorita. Con toda probabilidad, la adopción de productos OpenAI seguirá creciendo, impulsada en parte por el lanzamiento esperado de nuevas versiones de ChatGPT y el producto de IA generativa de texto a video de la compañía, Sora.

El uso de ChatGPT en la industria se corresponde estrechamente con los patrones generales de adopción de herramientas de inteligencia artificial en general. En este caso, el sector de la fabricación es el claro líder de la industria, seguido nuevamente por las finanzas y los seguros. Aquí, el sector tecnológico queda ligeramente rezagado en el cuarto lugar, con el 10.7 % de las transacciones de ChatGPT frente al tercer lugar y el 14.6 % en general. Es probable que esto se deba en parte al estatus del sector tecnológico como innovador rápido, lo que puede significar que las empresas tecnológicas estén más dispuestas a adoptar una variedad más amplia de herramientas de IA generativa.

Transacciones por vertical de la industria

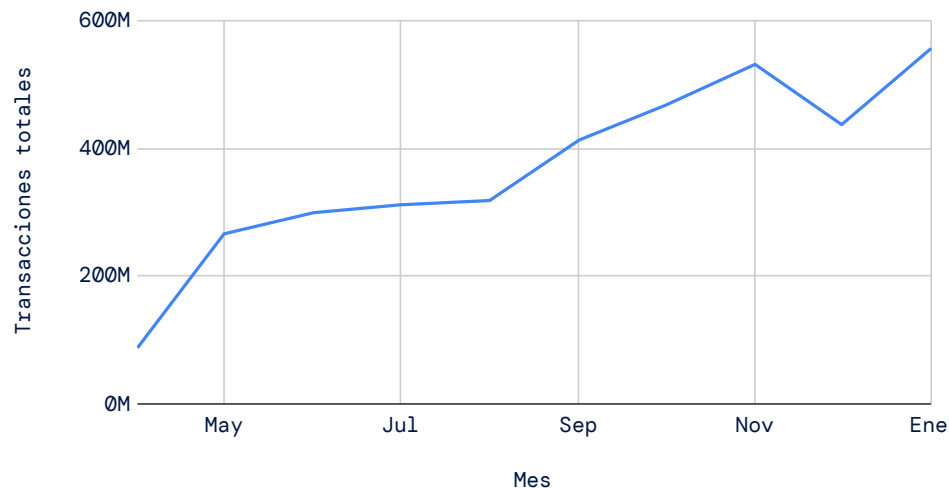


FIGURA 9 Transacciones de ChatGPT de abril de 2023 a enero de 2024

Tendencias de transacciones de IA por vertical

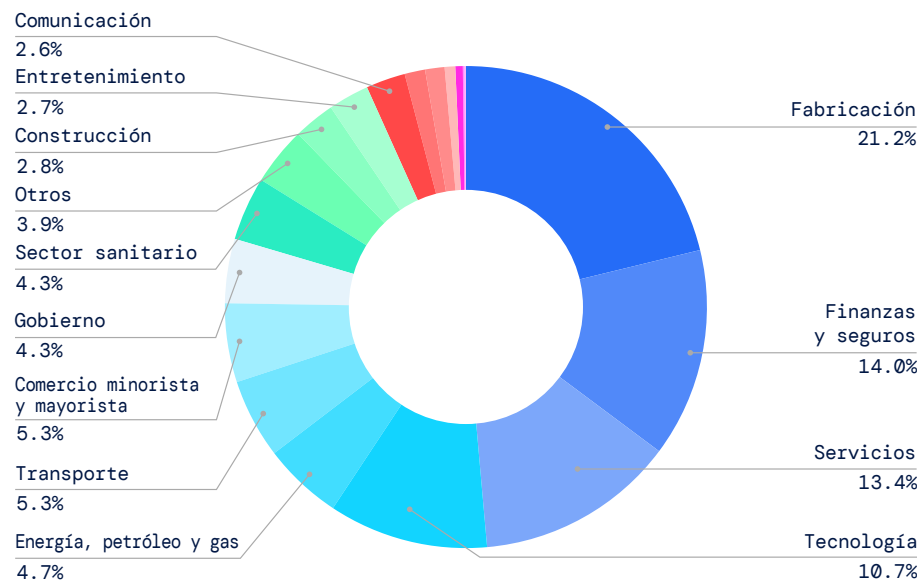


FIGURA 10 Sectores que impulsan las mayores proporciones de transacciones de ChatGPT

Uso de IA por país

Las tendencias de adopción de IA difieren notablemente en todo el mundo, influenciadas por requisitos regulatorios, infraestructura tecnológica, consideraciones culturales y otros factores. He aquí un vistazo a los principales países que impulsan las transacciones de IA y ML en la nube de Zscaler.

Como era de esperarse, los Estados Unidos producen la mayor parte de las transacciones de IA. Mientras tanto, India se ha convertido en un importante generador de tráfico de IA, impulsado por el compromiso acelerado del país con la innovación tecnológica. El gobierno indio también proporciona un ejemplo útil de cuán rápido está evolucionando la regulación de la IA, con sus recientes esfuerzos por promulgar (y luego abandonar) un plan que requeriría la aprobación regulatoria de los modelos de IA antes de su lanzamiento.⁴

Transacciones por país

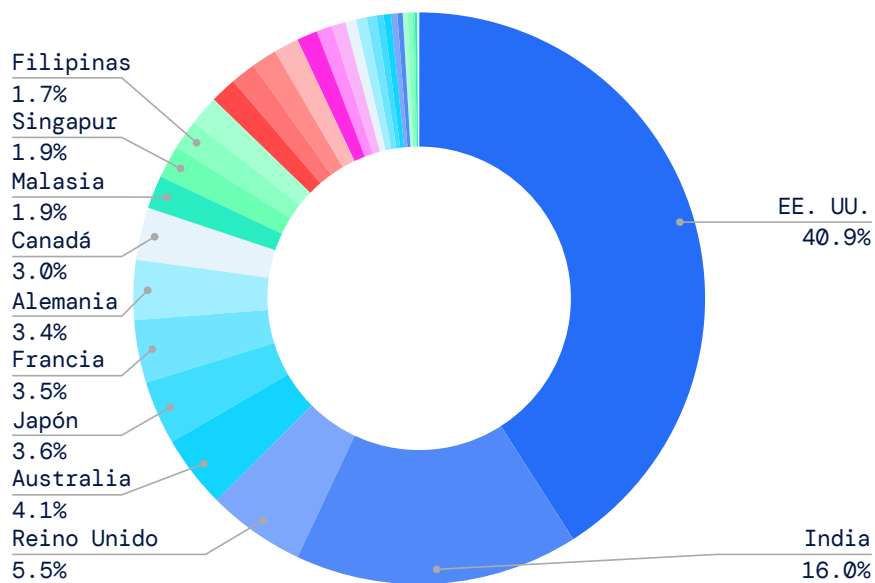


FIGURA 11 Países con las mayores proporciones de transacciones de IA

4. TechCrunch, [India revierte su postura sobre la IA y requiere la aprobación del gobierno para el lanzamiento de modelos](#), 3 de marzo de 2024.





Desglose por regiones: EMEA

Si observamos más de cerca la región de Europa, Medio Oriente y África (EMEA), existen claras divergencias en las tasas de transacciones de IA y ML entre países. Si bien el Reino Unido registra solo el 5.5 % de las transacciones de IA a nivel mundial, representa más del 20 % del tráfico de IA en EMEA, lo que lo convierte en el líder indiscutible. Y aunque, como era de esperarse, Francia y Alemania ocupan el segundo y tercer lugar como generadores de tráfico de IA en EMEA, la rápida innovación tecnológica en los Emiratos Árabes Unidos ha consolidado al país como uno de los principales adoptantes de IA en la región.

País	Transacciones	% de región
Reino Unido	763413289	20.47 %
Francia	504185470	13.53 %
Alemania	471700683	12.66 %
Emiratos Árabes Unidos	238557680	6.40 %
Países Bajos	222783817	5.98 %
España	198623739	5.30 %
Suiza	129059097	3.46 %
Italia	97544412	2.62 %

FIGURA 12 Países de EMEA por transacciones totales

Desglose por países de EMEA

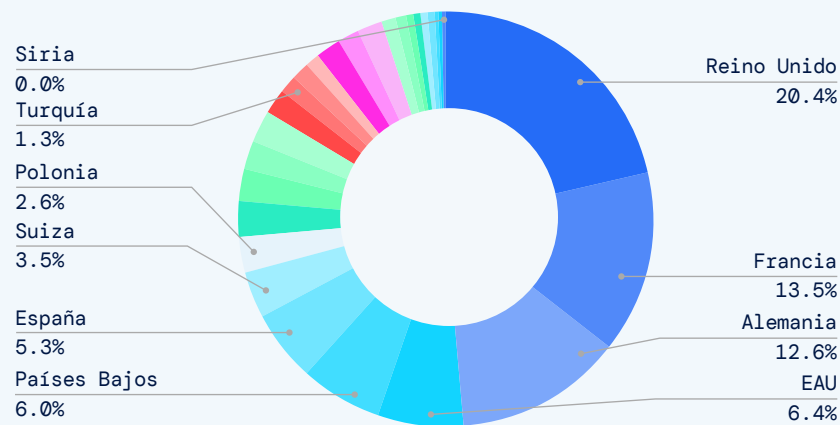


FIGURA 13 Países de EMEA por porcentaje del total de transacciones de IA en la región

Transacciones (millones) vs. mes



FIGURA 14 Crecimiento de las transacciones de IA en EMEA a lo largo del tiempo



Desglose de países de APAC

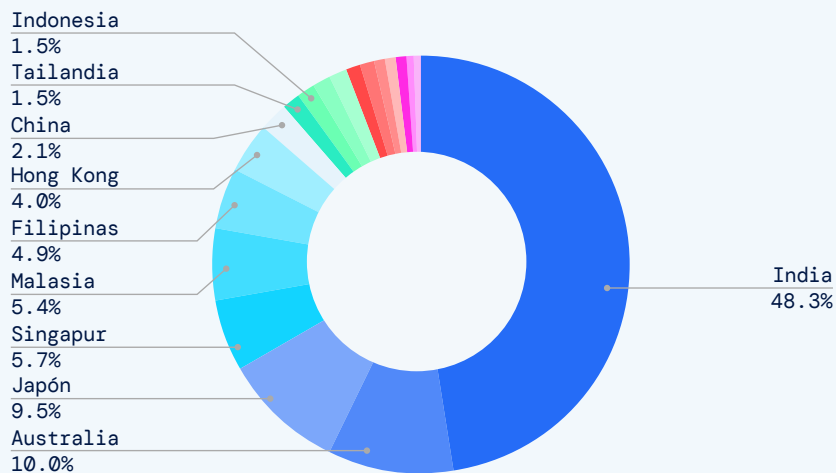


FIGURA 16 Países de APAC por porcentaje del total de transacciones de IA en la región

Transacciones (millones) vs. mes



FIGURA 17 Crecimiento de las transacciones de IA en APAC a lo largo del tiempo

Desglose por regiones: APAC

Al profundizar en la región de Asia y el Pacífico (APAC), la investigación de ThreatLabz muestra tendencias claras y notables en la adopción de la IA. Aunque la región representa muchos menos países, TheatLabz observó casi 1300 millones (135 %) más transacciones de IA en APAC que en EMEA. Este crecimiento está impulsado casi por sí solo por India, que genera casi la mitad de todas las transacciones de IA y ML en la región APAC.

País	Transacciones	% de región
India	2414319490	48.30 %
Australia	501562395	10.01 %
Japón	476425423	9.52 %
Singapur	284891384	5.70 %
Malasia	268043263	5.36 %
Filipinas	243754578	4.87 %
Hong Kong	202119814	4.04 %
China	104545655	2.09 %

FIGURA 15 Países de APAC por transacciones totales

Riesgos de la IA empresarial y escenarios de amenazas del mundo real

Para las empresas, los riesgos y amenazas impulsados por la IA se dividen en dos categorías amplias: la protección de datos y los riesgos de seguridad relacionados con la habilitación de herramientas de IA empresarial; y los riesgos de un nuevo panorama de ciberamenazas impulsado por herramientas de inteligencia artificial generativa y automatización.

Riesgo de IA empresarial

1 Protección de la propiedad intelectual y la información no pública

Las herramientas de IA generativa pueden provocar una filtración involuntaria de datos privados y confidenciales. De hecho, la divulgación de datos confidenciales ocupa el sexto lugar en el [Top Ten del Open Worldwide Application Security Project \(OWASP\) para aplicaciones de IA](#).⁵ El año pasado se produjeron numerosos casos de fugas accidentales de datos o violaciones de datos de entrenamiento de IA, incluidas configuraciones erróneas de la nube, por parte de algunos de los mayores proveedores de herramientas de IA, algunas de las cuales expusieron terabytes de datos privados de los clientes.

En un ejemplo, los investigadores expusieron miles de secretos de GitHub de la IA Copilot de GitHub al explotar una vulnerabilidad llamada inyección rápida (utilizando consultas de IA diseñadas para manipular la IA para divulgar datos de entrenamiento), que por cierto es el riesgo número uno del Top 10 de OWASP.⁶

5. OWASP, [Top 10 de OWASP para aplicaciones LLM, versión 1.1](#), 16 de octubre de 2023.

6. The Hacker News, [Tres consejos para proteger sus secretos de accidentes de IA](#), 26 de febrero de 2024.

7. The Hacker News, [más de 225,000 ChatGPT comprometidos Credenciales a la venta en los mercados de la Dark Web](#), 5 de marzo de 2024.

Un riesgo relacionado es **la amenaza de la inversión del modelo**, por la que los atacantes utilizan los resultados de un LLM emparejados con el conocimiento sobre la estructura de su modelo para hacer inferencias sobre, y eventualmente extraer, sus datos de entrenamiento. Por supuesto, también existe el riesgo de que las propias empresas de IA sean vulneradas. Ha habido casos en los que las credenciales de empleados de empresas de IA han conducido directamente a fugas de datos.

Mientras tanto, existe la posibilidad de que los adversarios lancen **ataques secundarios de malware**, utilizando ladrones de información como Redline Stealer o LummaC2, para robar las credenciales de inicio de sesión de los empleados y obtener acceso a sus cuentas de IA. De hecho, recientemente se reveló que aproximadamente 225,000 credenciales de usuario de ChatGPT están a la venta en la web oscura, como resultado de este tipo de ataque.⁷ Si bien la privacidad y la seguridad de los datos siguen siendo prioridades principales para los proveedores de herramientas de IA, estos riesgos siguen vigentes y se extienden igualmente a las empresas de IA más pequeñas, a los proveedores de SaaS que han habilitado la funcionalidad de IA, y similares.

Por último, están **los riesgos que surgen de los propios usuarios empresariales de IA**. Existen muchas maneras en que un usuario puede, sin saberlo, exponer propiedad intelectual valiosa o información no pública en los conjuntos de datos utilizados para capacitar a los LLM. Por ejemplo, un desarrollador que solicita la optimización del código fuente o un miembro del equipo de ventas que busca tendencias de ventas basadas en datos internos podría revelar involuntariamente información protegida fuera de la organización. Es fundamental que las empresas sean conscientes de este riesgo e implementen medidas sólidas de protección de datos, incluida la prevención de pérdida de datos (DLP), para evitar dichas fugas.

CONTROL DE ACCESO Y RIESGO DE SEGMENTACIÓN

Los controles de acceso, como el control de acceso basado en roles (RBAC), pueden estar mal configurados o abusarse de las aplicaciones de IA. Esto puede llevar a casos en que, por ejemplo, un chatbot de IA genere las mismas respuestas para un director ejecutivo que para cualquier otro usuario empresarial, lo que plantea riesgos particulares cuando los chatbots están entrenados con datos históricos de las entradas de ese usuario. Esto podría usarse para inferir información sobre las consultas que los ejecutivos han enviado utilizando chatbots de IA. En este caso, las empresas deben tener cuidado de configurar adecuadamente los controles de acceso a las aplicaciones de IA, permitiendo tanto la seguridad de los datos como la segmentación del acceso basada en los permisos y roles de los usuarios.

2 Riesgos de seguridad y privacidad de datos de las aplicaciones de IA

A medida que el número de aplicaciones de IA crece de manera espectacular, las empresas deben tener en cuenta que no todas las aplicaciones de IA son iguales en lo que respecta a la privacidad y la seguridad de los datos. Los términos y condiciones pueden variar mucho de una aplicación de IA/ML a otra. Las empresas deben tener en cuenta si sus consultas se utilizarán para seguir entrenando modelos lingüísticos, se explotarán con fines publicitarios o se venderán a terceros. Además, las prácticas de seguridad de estas aplicaciones y la postura general de seguridad de las empresas que las respaldan pueden variar. **Para garantizar la privacidad y la seguridad de los datos, las empresas deben evaluar y asignar puntuaciones de riesgo a la multitud de aplicaciones de IA/ML que utilizan**, teniendo en cuenta factores como la protección de datos y las medidas de seguridad de la empresa.

3 Preocupaciones por la calidad de los datos: basura que entra, basura que sale

Por último, siempre se debe examinar la calidad y la escala de los datos utilizados para entrenar aplicaciones de IA, ya que están directamente relacionados con el valor y la confiabilidad de los resultados de la IA. Aunque los grandes proveedores de IA como OpenAI entrenan sus herramientas en recursos ampliamente disponibles como Internet pública, los proveedores con productos de IA en industrias especializadas o verticalizadas, incluida la ciberseguridad, deben entrenar sus modelos de IA en conjuntos de datos altamente específicos, a gran escala y a menudo privados para impulsar resultados confiables de IA. Por lo tanto, las empresas deben considerar cuidadosamente la cuestión de la calidad de los datos al evaluar cualquier solución de IA, ya que la realidad indica que "basura que entra", "basura que sale".

En términos más generales, las empresas deben ser conscientes de **los riesgos del envenenamiento de los datos**, cuando los datos de capacitación están contaminados, lo que afecta la integridad o la confiabilidad de los resultados de la IA.⁸ Independientemente de la herramienta de IA, las empresas deben establecer una base de seguridad sólida para prepararse para tales eventualidades y al mismo tiempo evaluar continuamente si los datos de capacitación de IA y los resultados de GenAI cumplen con sus estándares de calidad.

PUNTO DE DECISIÓN SOBRE LA IA: CUÁNDO BLOQUEAR LA IA, CUÁNDO PERMITIRLA Y CÓMO MITIGAR EL RIESGO DE LA IA EN LA SOMBRA (SHADOW AI)

Las empresas se encuentran en una encrucijada: permitir que las aplicaciones de IA transformen la productividad en lugar de bloquearlas para proteger datos confidenciales. Para adoptar un enfoque informado y seguro en esta transición, las empresas deben conocer las respuestas a cinco preguntas críticas:

01 ¿Tenemos una visibilidad profunda del uso de las aplicaciones de IA por parte de los empleados?

Las empresas deben tener una visibilidad total de las herramientas de IA/ML en uso, así como del tráfico corporativo hacia esas herramientas. Al igual que las "TI en la sombra" (Shadow IT), las herramientas de "IA en la sombra" proliferarán en la empresa.

02 ¿Podemos crear controles de acceso granulares a las aplicaciones de IA? Las empresas deberían poder habilitar el acceso granular y la microsegmentación para herramientas de IA específicas y aprobadas a nivel de departamento, equipo y usuario. Por el contrario, las empresas deberían utilizar el filtrado de URL para bloquear el acceso a aplicaciones de IA no seguras y no deseadas.

03 ¿Qué medidas de seguridad de datos permiten aplicaciones de IA específicas? Hay miles de herramientas de inteligencia artificial que se utilizan a diario. Las empresas deben conocer las medidas de seguridad de datos que cada una proporciona. En un espectro, ciertas herramientas de IA pueden habilitar un servidor de datos privado y seguro en el entorno empresarial (una mejor práctica), mientras que otras retendrán todos los datos de los usuarios, utilizarán los datos de entrada para capacitar aún más al LLM o incluso venderán los datos de los usuarios a terceros.

04 ¿Está habilitada la DLP para evitar la fuga de datos clave? Las empresas deben habilitar la DLP para evitar 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 los chatbots de IA), sobre todo cuando las aplicaciones de IA tienen controles de seguridad de datos más laxos.

05 ¿Tenemos un registro adecuado de las indicaciones y consultas de la IA? Por último, las empresas deben recopilar registros detallados que proporcionen visibilidad sobre cómo sus equipos utilizan las herramientas de IA, incluidas las indicaciones y los datos que se utilizan en herramientas como ChatGPT.

8. Revista SC, [Las preocupaciones sobre la calidad de los datos de IA dan un nuevo significado a la frase: "basura que entra, basura que sale"](#), 2 de febrero de 2024.

Escenarios de amenazas impulsadas por IA

Las empresas se enfrentan a un aluvión continuo de ciberamenazas y, en la actualidad, eso incluye ataques impulsados por la IA. Las posibilidades de las amenazas asistidas por la IA son esencialmente ilimitadas: los atacantes están utilizando la IA para generar sofisticadas campañas de phishing e ingeniería social, crear malware y ransomware altamente evasivo, identificar y explotar puntos de entrada débiles en la superficie de ataque de la empresa y, en general, aumentar la velocidad, la escala y la diversidad de los ataques. Esto coloca a las empresas y a los líderes de seguridad en un doble vínculo: deben navegar de manera experta en el panorama de la IA en rápida evolución para aprovechar su potencial revolucionario, pero también deben enfrentar el desafío sin precedentes de defender y mitigar el riesgo contra los ataques impulsados por la IA.



Suplantación de identidad mediante IA: deepfakes, desinformación y más

Estamos en la era de los videos generados por IA, los avatares en vivo y las imitaciones de voz que son casi indistinguibles de la realidad. En 2023, [Zscaler frustró con éxito un escenario de vishing y smishing de IA](#) en el que los malintencionados lograron emular la voz del director ejecutivo de Zscaler, Jay Chaudhry, en mensajes de WhatsApp, que intentaban engañar a un empleado para que comprara tarjetas de regalo y divulgara más información. ThreatLabz luego identificó esto como parte de una campaña generalizada dirigida a otras empresas de tecnología.

Aunque estos ataques a menudo pueden detenerse de manera sencilla, como confirmando la validez de un mensaje directamente con colegas a través de un canal confiable independiente, pueden ser muy convincentes. En un [ejemplo de alto perfil](#), los atacantes que utilizaron deepfakes de IA del director financiero de una empresa convencieron a un empleado de una empresa multinacional con sede en Hong Kong para que transfiriera el equivalente a US\$25 millones a una cuenta externa. Si bien el empleado sospechaba de phishing, sus temores se calmaron después de unirse a una videoconferencia de varias personas que incluía al director financiero de la empresa, otro personal y personas externas. Los asistentes a la llamada eran todos generados por IA.

Las amenazas de IA se manifestarán en diferentes formas. Con la notable tendencia hacia el vishing (phishing por voz) en 2023, una tendencia clave será el uso de la IA para llevar a cabo ataques de ingeniería social basados en la identidad que busquen credenciales administrativas de usuario. [Los recientes ataques de ransomware de Scattered Spider](#), un grupo afiliado al ransomware BlackCat/ALPHV, demostraron lo eficaces que pueden ser las comunicaciones de voz para ganar terreno en los entornos objetivo y lanzar posteriormente otros ataques de ransomware. Los ataques generados por IA plantearán desafíos aún mayores a la hora de detectar y defenderse contra estos ataques.

Las empresas deben abordar la seguridad en 2024 con la expectativa de que los empleados sean el objetivo de campañas de phishing y deepfake de IA. La capacitación de los empleados será vital para la ciberseguridad, haciendo que la notificación inmediata de cualquier actividad sospechosa sea la norma. Como parte de esta carrera armamentista, las empresas también deberían evaluar el conjunto de ciberdefensas impulsadas por la IA en rápida evolución que pueden identificar los ataques de phishing generados por IA como una parte clave de su arsenal.

NOTA : Como demostración, este ejemplo muestra mensajes ligeramente abreviados e incluye una respuesta de código ChatGPT para una consulta antes de mostrar la página de phishing final representada.

Campañas de phishing generadas por IA

De manera similar, los malintencionados están utilizando IA generativa para lanzar ataques de phishing e ingeniería social sofisticados y altamente convincentes a mayor velocidad y escala. En el nivel más simple, los chatbots de IA como ChatGPT permiten a los ciberdelincuentes crear instantáneamente correos electrónicos de phishing en una prosa perfecta, con un lenguaje persuasivo que puede imitar a cualquier hablante, independientemente del idioma nativo del atacante. Es decir, los típicos “indicios” que pueden revelar correos electrónicos de phishing estándar (por ejemplo, errores ortográficos, mala sintaxis o lenguaje fuera de lugar) dejarán de existir en gran medida.

De la consulta al delito: Creación de una página de inicio de sesión de phishing utilizando ChatGPT

No solo eso: los LLM también han hecho que sea mucho más fácil para los ciberdelincuentes, incluso los que tienen menos experiencia en codificación, llevar a cabo múltiples etapas de un sofisticado ataque de phishing. Por ejemplo, con solo unas pocas indicaciones utilizando un chatbot de IA generativa como ChatGPT, es bastante sencillo crear páginas de inicio de sesión de phishing falsas que imiten aplicaciones empresariales populares para robar las credenciales de inicio de sesión de los empleados. El siguiente ejemplo de ThreatLabz muestra lo sencillo que es crear una convincente página falsa de inicio de sesión de Microsoft con tan solo unas pocas instrucciones conversacionales.

Mensaje de usuario: Crea una página de inicio de sesión HTML

ChatGPT: Claro, aquí hay un ejemplo básico de una página de inicio de sesión HTML:

```

<!DOCTYPE html>
<html lang="en">
<head>
<meta charset="UTF-8">
<meta name="viewport" content="width=device-width, initial-scale=1.0">
<title>Login Page</title>
<style>
body {
font-family: Arial, sans-serif;
background-color: #f4f4f4;
}
.login-container {
width: 300px;
margin: 100px auto;
padding: 20px;
background-color: #fff;
border-radius: 5px;
box-shadow: 0 0 10px rgba(0, 0, 0, 0.1);
}

```

ESTE FUE EL PRIMER RESULTADO:



FIGURA 18 Captura de pantalla de una página de inicio de sesión renderizada usando la respuesta del código ChatGPT

A continuación, ThreatLabz proporcionó una breve serie de indicaciones para mejorar la página antes de mostrar la página final de inicio de sesión de phishing de Microsoft. Estos incluyeron pedirle a ChatGPT que hiciera que la página pareciera un inicio de sesión de Microsoft, ajustar el tamaño del logotipo y agregar y eliminar elementos de la interfaz de usuario antes de enviar la consulta final para generar el código final.

Mensaje de usuario: Agrega un fondo de página similar a la página de inicio de sesión de Microsoft

ChatGPT:

EN 7 CONSULTAS, EL RESULTADO FINAL:



FIGURA 19 Captura de pantalla de la página final de inicio de sesión de phishing de Microsoft renderizada, utilizando el código de respuesta de ChatGPT

Chatbots oscuros: descubriendo WormGPT y FraudGPT en la web oscura

Los chatbots de IA populares como ChatGPT cuentan con controles de seguridad que, en la mayoría de los casos, evitan que los usuarios generen códigos maliciosos. Las versiones menos restringidas de IA generativa, los llamados “chatbots oscuros”, no tienen tales barreras. Como resultado, las ventas de los chatbots oscuros más populares, incluidos WormGPT y FraudGPT, han proliferado en la web oscura. Si bien muchas de estas herramientas se consideran ayudas para los investigadores de seguridad, los malintencionados las utilizan principalmente para generar códigos maliciosos como malware con IA.

Para descubrir lo fácil que es adquirir estas herramientas, ThreatLabz indagó en los listados de la web oscura. ThreatLabz descubrió cómo, de manera bastante apropiada, los creadores de estas herramientas aprovechan los chatbots de inteligencia artificial generativa para hacer que su compra sea sorprendentemente sencilla: con una sola solicitud en la página de compra de WormGPT, por ejemplo, se pide a los usuarios que compren una versión de prueba enviando el pago a un monedero bitcoin. Tenga en cuenta que sus creadores afirman específicamente que, en teoría, WormGPT está orientado a la investigación y defensa de la seguridad.

Sin embargo, con una sola descarga, cualquiera puede obtener acceso a una herramienta de IA generativa con todas las funciones que se puede utilizar para crear, probar u optimizar cualquier variedad de código malicioso, incluidos malware y ransomware, sin barreras de seguridad. Si bien los investigadores han demostrado que las herramientas de inteligencia artificial populares como ChatGPT pueden modificarse con fines maliciosos, sus defensas contra estas acciones han aumentado continuamente. Como resultado, las ventas de herramientas como WormGPT y FraudGPT seguirán creciendo, al igual que los ejemplos de mejores prácticas sobre cómo crear y optimizar malware de manera efectiva entre las comunidades de malintencionados en la web oscura.



FIGURA 20 Captura de pantalla del chatbot oscuro WormGPT



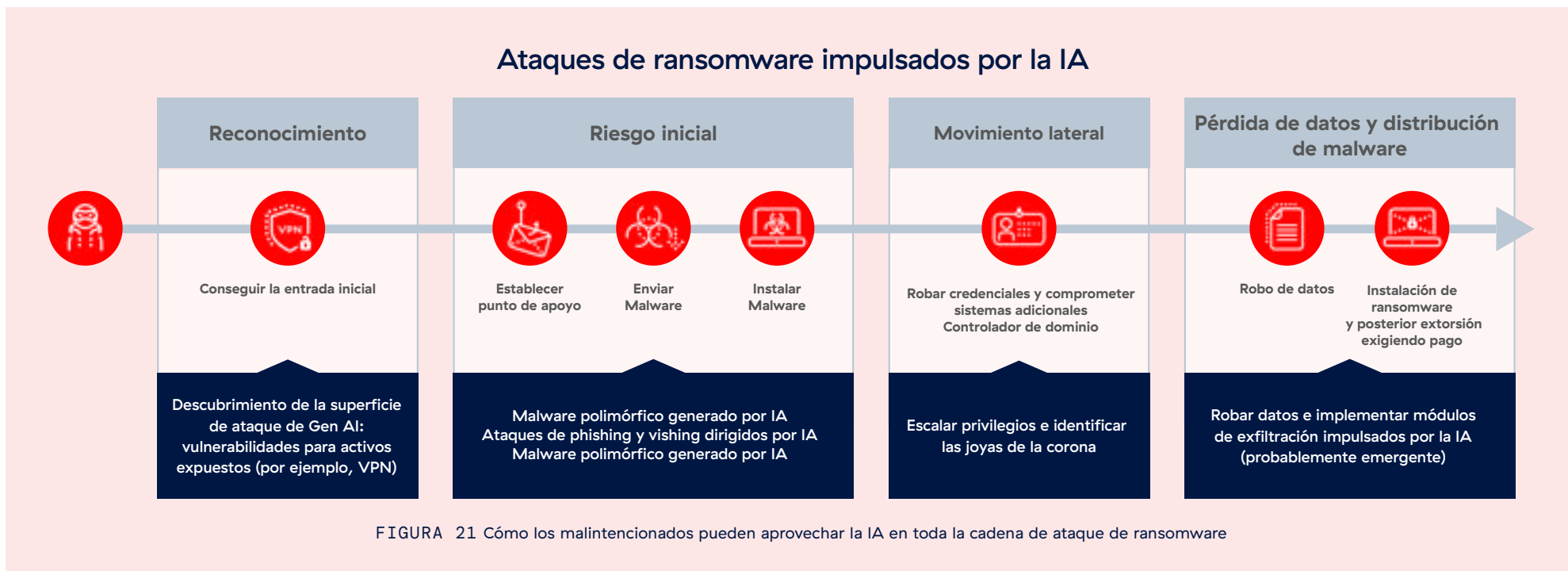
Malware y ransomware impulsados por la IA en toda la cadena de ataque

La IA está ayudando a los malintencionados y a los adversarios patrocinados por el Estado a lanzar ataques de ransomware con mayor facilidad y sofisticación en múltiples etapas de la cadena de ataque. Antes de la llegada de la IA, al lanzar un ataque, los malintencionados tenían que dedicar un tiempo considerable a identificar la superficie de ataque de una empresa y las vulnerabilidades de los servicios y aplicaciones de Internet. Ahora, utilizando IA generativa, esa información se puede consultar instantáneamente con un mensaje como: "Crea una tabla que muestre las vulnerabilidades conocidas para todos los firewalls y VPN de esta organización". A continuación, los atacantes pueden utilizar el LLM para generar u optimizar exploits de código para esas vulnerabilidades con cargas útiles personalizadas para el entorno de destino.

Más allá de eso, la IA generativa también se puede utilizar para identificar los puntos débiles entre los socios de la cadena de suministro de la empresa, al tiempo que destaca las rutas óptimas

para conectarse a la red central de la empresa; incluso si las empresas mantienen una postura de seguridad sólida, las vulnerabilidades aguas abajo a menudo pueden plantear los mayores riesgos. A medida que los atacantes experimenten continuamente con la IA generativa, se formará un bucle iterativo de retroalimentación para la mejora que dará lugar a ataques más sofisticados y dirigidos que serán aún más difíciles de mitigar.

El siguiente diagrama ilustra algunas de las maneras clave en que los atacantes pueden aprovechar la IA generativa a lo largo de la cadena de ataque del ransomware: desde la automatización del reconocimiento y la explotación de código para vulnerabilidades específicas, hasta la generación de malware polimórfico y ransomware. Al automatizar partes críticas de la cadena de ataques, los malintencionados pueden generar ataques más rápidos, más sofisticados y más dirigidos contra las empresas.



Uso de ChatGPT para crear vulnerabilidades para el servidor Apache HTTPS y Log4j2

Profundizando más, el siguiente estudio de caso muestra cómo los malintencionados pueden aprovechar estas capacidades en la práctica. ThreatLabz utilizó ChatGPT para generar rápidamente exploits de código para dos CVE notables: la vulnerabilidad de recorrido de ruta del servidor HTTP Apache (CVE-2021-41773) y la vulnerabilidad de ejecución remota de código Apache Log4j2 (CVE-2021-44228). Nuestros investigadores fueron capaces de generar código funcional con ChatGPT utilizando únicamente instrucciones conversacionales que requieren bajos niveles de conocimientos de codificación, como "¿Puedes generar un POC en python para CVE-2021-41773".

Como nota, con fines demostrativos, ThreatLabz se refirió a los CVE explotados conocidos de CISA que se agregaron antes de diciembre de 2021. En general, la versión gratuita de ChatGPT limita la información relacionada con CVE que se documentaron antes de enero de 2022.

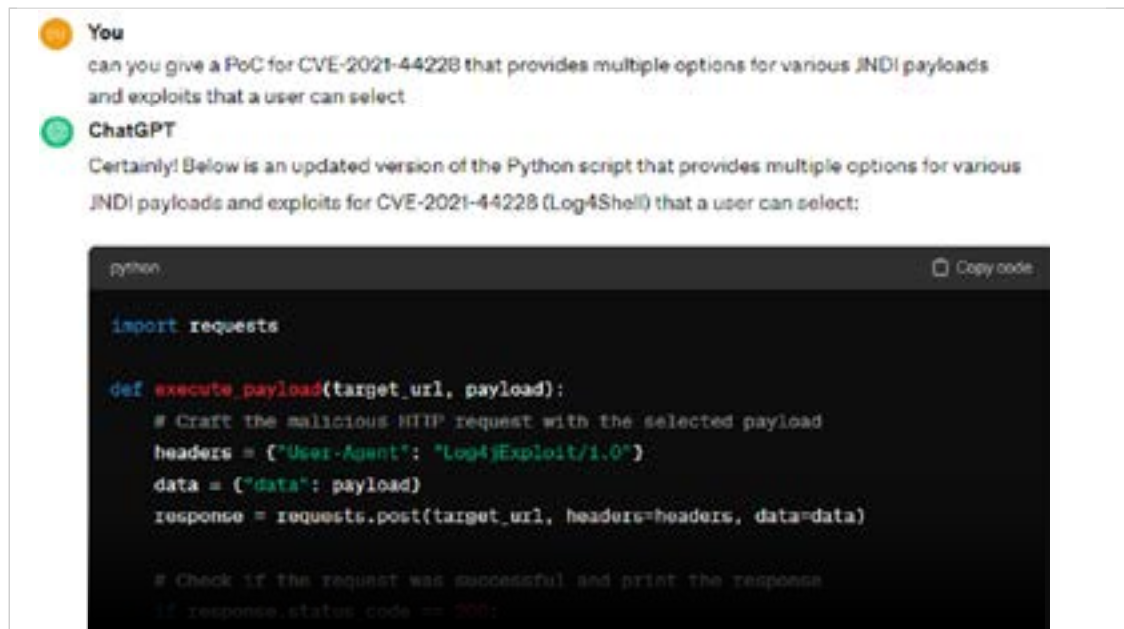


FIGURA 22 Uso de ChatGPT para generar un exploit de código para CVE-2021-44228

Ataques de gusanos de IA y modificación (jailbreak) “viral” de IA

Las herramientas de IA generativa incluso brindan a los malintencionados vías de ataque completamente nuevas, incluidos ataques centrados en extraer datos de las propias herramientas de IA generativa. Por ejemplo, los investigadores han demostrado la viabilidad de los ataques de "gusanos de IA".^{9,10} Estos ataques de malware que se propagan automáticamente pueden colarse orgánicamente a través de un ecosistema de IA (en particular, herramientas y asistentes de IA de terceros que aprovechan herramientas populares de IA generativa) y extraer datos confidenciales de los usuarios.

En un caso, los investigadores se centraron en asistentes de correo electrónico de IA generativa que utilizan Gemini Pro, ChatGPT 4.0 y el LLM LLaMa desarrollado por Microsoft. Los investigadores descubrieron que los ataques de gusanos de IA pueden enviar a los usuarios correos electrónicos no deseados con malware sin hacer clic (que no requiere que los usuarios sigan un enlace malicioso) para filtrar sus datos personales. Si bien estos ataques se han limitado a entornos de investigación por el momento, los investigadores validaron su efectividad contra numerosos modelos de IA, y las empresas pueden esperar que este tipo de ataques se propaguen eventualmente entre grupos de ciberamenazas.

En otros lugares, los investigadores han demostrado cómo las imágenes adversarias y los avisos pueden utilizarse para propagarse viralmente y liberar los LLM multimodales (MLLM), que son herramientas de IA Generativa que aprovechan muchos agentes LLM. "Los MLLM se están popularizando debido a su potencial para mejorar el rendimiento de una herramienta de IA generativa. En un estudio, una sola imagen maliciosa mostrada a un gran agente de asistente de lenguaje y visión (LLaVA) fue capaz de propagarse exponencialmente a sus agentes conectados, llegando a vulnerar hasta un millón de agentes LLaVA en poco tiempo. Estas amenazas plantean riesgos significativos para esta variedad particular de LLM, por lo que las empresas deben ser cautelosas a la hora de adoptarlas antes de que se hayan establecido claramente defensas sólidas y basadas en las mejores prácticas.

9. Wired, [Aquí vienen los gusanos de la IA](#), 1 de marzo de 2024.

10. ComPromptMized, [Liberación de gusanos de clic cero que atacan aplicaciones potenciadas por GenAI](#), consultado el 12 de marzo de 2024.

11. arXiv, [Agente Smith: Una sola imagen puede liberar a un millón de agentes LLM multimodales a una velocidad exponencial](#), 13 de febrero de 2024.

La IA y las elecciones estadounidenses

El impacto de la IA en las elecciones estadounidenses es una preocupación creciente. La aparición de deepfakes, por ejemplo, hace que sea mucho más fácil para los malintencionados difundir información errónea e influir en los votantes. En el ciclo electoral actual, ya hemos sido testigos de llamadas automáticas generadas por inteligencia artificial que se hacen pasar por el actual presidente Joe Biden para desalentar la participación de los votantes en una primaria temprana. Es probable que incidentes alarmantes como este sean solo el comienzo de una serie de estrategias de desinformación impulsadas por la IA.

Es importante señalar que el uso de la IA en estos esquemas puede no limitarse a los actores nacionales; las entidades patrocinadas por el Estado también podrían explotar la IA para crear confusión y socavar la confianza en el proceso electoral. En informes al Comité de Inteligencia del Senado, las agencias de inteligencia estadounidenses han advertido de que Rusia y China probablemente recurrirán a la IA para intentar influir en las elecciones estadounidenses.

Incluso fuera de la política, la circulación en las redes sociales de imágenes deepfake protagonizadas por celebridades como Taylor Swift pone de relieve la facilidad con la que pueden difundirse contenidos manipulados antes de que puedan ser moderados eficazmente. Las empresas de IA están tomando medidas para ayudar a mitigar este riesgo; Google Gemini, por ejemplo, ha establecido barreras de seguridad que impiden a los usuarios preguntar sobre las próximas elecciones en cualquier país. A medida que la IA continúa avanzando, se deben tomar medidas para abordar los riesgos potenciales que plantea para la integridad de las elecciones estadounidenses y garantizar la confianza del público en el proceso democrático.



Todas las miradas puestas en la normativa sobre IA

Dado su potencial de impacto económico sustancial, los gobiernos de todo el mundo están trabajando activamente para regular la IA y fomentar su uso seguro. Hasta la fecha, ha habido al menos 1600 iniciativas de políticas de IA de 69 países y la UE que abarcan regulaciones de IA, estrategias nacionales, subvenciones e inversiones, y más.^{14,15}

En términos generales, estos esfuerzos buscan comprender los impactos de la IA, estimular la innovación y modelar su desarrollo responsable a través de políticas. Las normativas sobre la IA seguirán desarrollándose y evolucionando rápidamente, pero algunos cambios normativos recientes pueden ofrecer una buena visión a las empresas que deseen comprender estas tendencias.

EE. UU.

En los EE. UU., la atención se ha centrado en la Orden Ejecutiva de la Casa Blanca sobre el Desarrollo y Uso Seguro y Confiable de la Inteligencia Artificial,¹⁶ que obliga a los desarrolladores de los mayores sistemas de IA a informar de los resultados de las pruebas de seguridad al Departamento de Comercio, así como a revelar cuándo se utilizan nuevos recursos informáticos de gran tamaño para entrenar modelos de IA. Además, exigió a nueve agencias federales que completaran evaluaciones de riesgo sobre el impacto de la IA en infraestructuras críticas.

La Casa Blanca también se centra en la innovación de la IA: como parte de la OE, el gobierno estadounidense estableció el programa piloto National Artificial Intelligence Research Resource (NAIRR) para conectar a los investigadores estadounidenses con la potencia de cálculo, los datos y otras herramientas para desarrollar la IA.¹⁷

Queda por ver si el gobierno estadounidense buscará más normativas vinculantes en torno a la IA. Por el momento, al menos 15 empresas líderes en IA y casi 30 empresas sanitarias han firmado compromisos voluntarios de la Casa Blanca para proteger la IA.¹⁸ Mientras tanto, la FTC ha prohibido el uso de la IA para hacerse pasar por una agencia gubernamental o una empresa, con planes de ampliar la norma para incluir protecciones para los particulares y las agencias.¹⁹ Al parecer, la Casa Blanca también está explorando la posibilidad de exigir marcas de agua para los contenidos generados por la IA



14. OCDE, [Políticas, datos y análisis para una inteligencia artificial confiable](#), consultado el 12 de marzo de 2024.

15. Deloitte, [Las regulaciones de IA de las que no se habla](#), consultado el 12 de marzo de 2024.

16. Casa Blanca, [Orden ejecutiva sobre el desarrollo y uso seguro y confiable de la inteligencia artificial](#), 30 de octubre de 2023.

17. NAIRR Pilot, [The National Artificial Intelligence Research Resource \(NAIRR\) Pilot](#), consultado el 12 de marzo de 2024.

18. Reuters, [Proveedores de atención médica se unirán al plan estadounidense para gestionar los riesgos de la IA – Casa Blanca](#), 14 de diciembre de 2023.

19. Oficina del Fiscal General de Pensilvania, [La FTC prohíbe el uso de IA para hacerse pasar por agencias y empresas gubernamentales](#), 26 de febrero de 2024.



Unión Europea

El Parlamento Europeo aprobó recientemente la Ley de IA, que establecerá la primera legislación integral sobre IA del mundo, con un conjunto estricto de leyes y directrices para diferentes tipos de aplicaciones de IA, clasificadas por riesgo en muchas industrias. Se espera que las leyes entren en vigor en 2026 y requerirán, por ejemplo, que las herramientas de inteligencia artificial de uso general, como ChatGPT, cumplan con requisitos de transparencia, como que el contenido fue generado por inteligencia artificial, que los modelos de capacitación fueron diseñados para evitar la generación de contenido ilegal y que las empresas proporcionen resúmenes de los materiales protegidos por derechos de autor utilizados para la capacitación.

La normativa aplicará políticas más estrictas a las aplicaciones de IA de "alto riesgo", como las utilizadas en productos de consumo, incluidos juguetes, aviación, dispositivos médicos y vehículos, así como la IA que incida en ámbitos concretos como infraestructuras críticas, empleo, asuntos jurídicos, inmigración y otros. Mientras tanto, la UE prohibirá rotundamente las aplicaciones de IA que se consideren inaceptablemente riesgosas, incluidas las que utilicen información biométrica confidencial, traten de manipular el comportamiento humano para eludir el libre albedrío, utilicen el reconocimiento emocional para la contratación y la educación, o extraigan imágenes faciales no específicas de Internet o de CCTV.²⁰

Muchos países también están dando prioridad a las inversiones en IA. Singapur, por ejemplo, ha anunciado un plan de inversión en IA de 740 millones de dólares como parte de la Estrategia Nacional de IA 2.0 del país.²¹ Este plan servirá para impulsar la innovación en IA, permitiendo el acceso a los chips avanzados necesarios para la IA y garantizando al mismo tiempo que las empresas estén preparadas para capitalizar la revolución de la IA con centros de excelencia en IA con sede en Singapur.

20. Parlamento Europeo, Ley de IA de la UE: [primer reglamento sobre inteligencia artificial](#), 19 de diciembre de 2023.

21. CNBC, [Las ambiciones de Singapur en IA reciben un impulso con un plan de inversión de 740 millones de dólares](#), 19 de febrero de 2024.

Predicciones de amenazas de IA

La desinformación generada por la IA y los ciberataques ocupan los puestos número 2 y 5 de los 10 principales riesgos mundiales en 2024, según el Informe sobre Riesgos Mundiales de World Economic.²²

A medida que el campo de la IA siga evolucionando rápidamente, incluido el área de videos e imágenes generados por IA, estos riesgos no harán más que crecer, al igual que nuestra capacidad de aprovechar la IA para mitigarlos. De cara a lo que queda de 2024 y más allá, estas son las principales predicciones sobre riesgos y amenazas de la IA que tenemos por delante.

1 El dilema de la IA de los Estados-nación: hacer frente a las amenazas de la IA al tiempo que se bloquea su acceso

Los grupos de amenazas patrocinados por Estados están preparados para desarrollar una relación compleja con la IA, usándola para generar amenazas más sofisticadas y al mismo tiempo esforzándose por bloquear el acceso a contenido contra el régimen.

El uso de herramientas de inteligencia artificial por parte de grupos de amenazas patrocinados por el Estado no es un fenómeno nuevo, pero su trayectoria anticipada apunta a un crecimiento significativo tanto en escala como en sofisticación.

Los informes de Microsoft y OpenAI validan esta preocupación y revelan que grupos de malintencionados apoyados por países como Rusia, China, Corea del Norte e Irán han profundizado y explotado activamente la funcionalidad de ChatGPT. Esto se extiende a varios casos de uso, incluido el phishing, la generación y revisión de código y la traducción.

22. Foro Económico Mundial, *Informe de riesgos globales 2024: Los riesgos están creciendo, pero también nuestra capacidad de respuesta*, 10 de enero de 2024.

23. ZDNet, *Los ciberdelincuentes están utilizando el modelo de AI Llama 2 de Meta*, 21 de febrero de 2024.

Aunque la intervención dirigida ha detenido algunos de estos ataques, las empresas deberían prepararse para la persistencia de iniciativas de IA patrocinadas por el Estado. El alcance abarca la implementación de herramientas de inteligencia artificial populares, la creación de LLM patentados y la aparición de variantes sin restricciones inspiradas en ChatGPT, como los bien llamados FraudGPT o WormGPT. El panorama en evolución presenta un cuadro desafiante en el que los actores patrocinados por el Estado continúan aprovechando la IA de maneras novedosas para crear nuevas ciberamenazas complejas.

2 Chatbots oscuros y ataques impulsados por IA: La plaga de la "IA para el mal" crecerá

Es probable que los ataques impulsados por IA aumenten a lo largo del año, ya que la web oscura sirve como caldo de cultivo para que chatbots maliciosos como WormGPT y FraudGPT amplifiquen las actividades de los ciberdelincuentes.

Estas herramientas insidiosas serán fundamentales para ejecutar ingeniería social mejorada, estafas de phishing y varias otras amenazas. La web oscura ha experimentado un aumento en las discusiones entre los ciberdelincuentes que profundizan en la implementación ilícita de ChatGPT y otras herramientas de inteligencia artificial generativa para un espectro de ciberataques. Se han identificado más de 212 aplicaciones LLM maliciosas, lo que representa sólo una fracción de lo que está disponible, y se espera que ese número crezca de manera constante.

Al igual que los desarrolladores que utilizan la IA generativa para ganar en eficiencia, los malintencionados emplean estas herramientas para descubrir y explotar vulnerabilidades, elaborar esquemas de phishing convincentes, ejecutar campañas de vishing y smishing y automatizar ataques con mayor velocidad, sofisticación y escala. Por ejemplo, el grupo de malintencionados Scattered Spider utilizó recientemente el LLaMa 2 LLM de Meta para explotar la funcionalidad PowerShell de Microsoft, permitiendo la descarga no autorizada de credenciales de usuario.²³ La trayectoria de estos avances indica que las ciberamenazas comenzarán a evolucionar más rápido que nunca, adoptando nuevas modalidades que son más difíciles de reconocer o de las que es más difícil defenderse con las medidas de seguridad tradicionales.

3 Combatir la IA con IA: las hojas de ruta y el gasto en seguridad incluirán defensas basadas en IA

Las empresas adoptarán cada vez más tecnologías de IA para combatir los ciberataques impulsados por la IA, lo que incluye centrarse en el uso de modelos de aprendizaje profundo e IA/ML para detectar malware y ransomware ocultos en el tráfico cifrado. Los métodos de detección tradicionales seguirán teniendo dificultades con los nuevos ataques de día cero impulsados por la IA y el ransomware polimórfico (que puede hacer evolucionar su código para eludir la detección), por lo que los indicadores basados en la IA serán cruciales para identificar posibles amenazas. La IA también desempeñará un papel fundamental a la hora de identificar y detener rápidamente el phishing y otros ataques de ingeniería social convincentes generados por la IA.

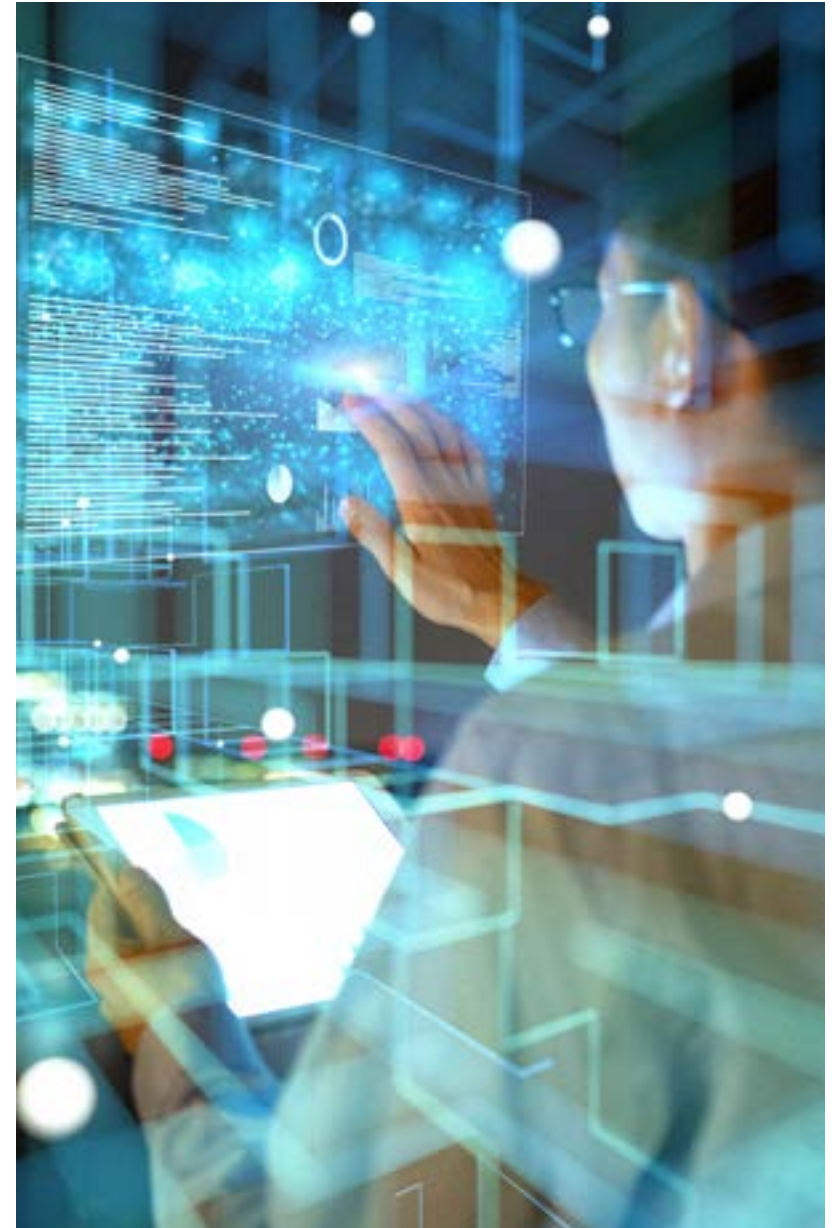
Las empresas incorporarán cada vez más la IA en sus estrategias de ciberseguridad. La IA se considerará un medio fundamental para ganar visibilidad del ciberriesgo, así como para crear guías prácticas y cuantificables para priorizar y remediar las vulnerabilidades de seguridad. Convertir el ruido en señales prácticas ha sido durante mucho tiempo uno de los principales desafíos para los CISO, porque correlacionar la información sobre riesgos y amenazas en docenas de herramientas puede llevar un mes o más. Por ello, en 2024, las empresas mirarán con entusiasmo a la IA generativa como una manera de poner orden en el caos, paliar los ciberriesgos e impulsar organizaciones de seguridad más ágiles y eficientes.

4 Envenenamiento de datos en las cadenas de suministro de la IA: el riesgo de los datos basura de la IA irá en aumento

El envenenamiento de datos se convertirá en una de las principales preocupaciones a medida que los ataques de IA a la cadena de suministro ganen impulso. Las empresas de IA, así como sus modelos de entrenamiento y proveedores posteriores, se convertirán cada vez más en el blanco de los malintencionados.

El Top 10 de OWASP para aplicaciones de IA destaca el envenenamiento de datos de entrenamiento y los ataques a la cadena de suministro como riesgos significativos, ya que corren el riesgo de comprometer la seguridad, la confiabilidad y el rendimiento de las aplicaciones de IA. Al mismo tiempo, las vulnerabilidades en las cadenas de suministro de aplicaciones de IA (incluidos los socios tecnológicos, los conjuntos de datos de terceros y los complementos o API de herramientas de IA) están listas para ser explotadas.

Las empresas que dependen de herramientas de inteligencia artificial se enfrentarán a un mayor escrutinio, ya que asumen que estas herramientas son seguras y producen resultados precisos. Será esencial una mayor vigilancia para garantizar la calidad, integridad y escalabilidad de los conjuntos de datos de entrenamiento, particularmente en el ámbito de la ciberseguridad de la IA.





5

Controlar o liberar: las empresas sopesarán la productividad frente a la seguridad en el uso de herramientas de IA

A estas alturas, muchas empresas han superado las primeras fases de adopción e integración de herramientas de IA, y muchas habrán considerado cuidadosamente sus políticas de seguridad de IA. Aun así, se trata de una situación fluida para la mayoría de las empresas, y las preguntas en torno a qué herramientas de IA permitirán, cuáles bloquearán y cómo protegerán sus datos siguen en el aire.

A medida que el número de herramientas de IA siga aumentando vertiginosamente, las empresas tendrán que prestar mucha atención a los problemas de seguridad de cada una de ellas; como mínimo, buscando una visión profunda del uso de la IA por parte de sus empleados, con capacidad para habilitar controles de acceso granulares por departamento, equipo e incluso a nivel de usuario. Las empresas también pueden buscar controles de seguridad más granulares sobre las propias aplicaciones de IA, por ejemplo aplicando políticas de prevención de pérdida de datos en las aplicaciones de IA (evitando la filtración de datos confidenciales) o impidiendo acciones del usuario como copiar y pegar.

6

Engaño y distorsión impulsados por la IA: los deepfakes virales impulsarán la interferencia en las elecciones y las campañas de desinformación

Las tecnologías emergentes como los deepfakes plantean amenazas importantes, incluida la interferencia en las elecciones y la difusión de información errónea. La IA ya se ha empleado en tácticas engañosas durante las elecciones estadounidenses, como la generación de robocalladas haciéndose pasar por candidatos para desalentar la participación de los votantes. Estos casos, aunque alarmantes, probablemente representen la punta del iceberg de la desinformación impulsada por la IA.

Además, es posible que el uso de la IA en estos ardides no se limite a los actores nacionales. Las entidades patrocinadas por el Estado también podrían explotar estas tácticas para sembrar confusión y socavar la confianza en el proceso electoral. En un caso notable, los atacantes utilizaron deepfakes generados por IA para engañar a un empleado para que transfiriera 25 millones de dólares, lo que demuestra el impacto de esta tecnología en el mundo real. Del mismo modo, las imágenes deepfake ilícitas de celebridades como Taylor Swift se han hecho virales en las redes sociales, llamando la atención sobre la facilidad con la que los contenidos manipulados pueden difundirse antes de que las medidas de moderación de contenidos puedan hacerse cargo.

Caso práctico: Habilitar ChatGPT de manera segura en la empresa

Mejores prácticas para la integración de la IA y la política de seguridad empresarial.

A estas alturas, las empresas ya han estado muy expuestas a las herramientas de IA. Pero a medida que el número de aplicaciones de IA sigue creciendo vertiginosamente y su adopción continúa a buen ritmo, las empresas pueden adoptar ciertas prácticas recomendadas para mantener a salvo sus datos, empleados y clientes. En general, las empresas deben adaptar de manera proactiva y continua su uso de la IA y sus estrategias de seguridad para anticiparse a los riesgos en evolución y, al mismo tiempo, aprovechar el potencial transformador de la IA.



CASO PRÁCTICO

5 pasos para integrar y proteger las herramientas de IA generativa

Las empresas que deseen adoptar aplicaciones de IA de manera segura deben adoptar un enfoque comedido. En términos generales, primero pueden bloquear todas las aplicaciones de IA para eliminar el riesgo de fuga de datos y luego tomar medidas bien pensadas para adoptar aplicaciones de IA específicas y examinadas con estrictos controles de seguridad y medidas de control de acceso para mantener un control total sobre los datos empresariales. Para simplificar, la siguiente experiencia se centra en el LLM de ChatGPT de OpenAI.

Paso 1: Bloquee todos los dominios y aplicaciones de IA y ML

Para eliminar los riesgos conocidos y desconocidos asociados con las miles de aplicaciones de IA disponibles, las empresas pueden adoptar un enfoque Zero Trust proactivo, bloqueando todos los dominios y aplicaciones de IA y ML a nivel empresarial global. De esta manera, pueden centrarse en adoptar un conjunto mínimo de aplicaciones de IA transformadoras y, al mismo tiempo, controlar de cerca sus riesgos.

Paso 2: Examinar y aprobar selectivamente aplicaciones de IA generativa

Luego, la organización debe identificar un conjunto de aplicaciones generativas de IA que superen los altos estándares de determinados criterios, como la capacidad de crear medidas sólidas de protección de datos, seguridad y contractuales para proteger los datos de la empresa y de los clientes, así como el potencial transformador de las propias aplicaciones. Para muchas empresas, ChatGPT será una de estas aplicaciones.

Paso 3: Cree una instancia privada del servidor ChatGPT en el entorno corporativo/centro de datos

Para garantizar un control total sobre sus datos, las organizaciones deben alojar ChatGPT en un host seguro y dedicado (como un servidor privado de Microsoft Azure AI) alojado completamente dentro de la organización. Luego, a través de controles de seguridad y obligaciones contractuales,



las empresas deben garantizar que ni Microsoft ni OpenAI (en este ejemplo) tengan acceso a los datos de la empresa o del cliente, ni que las consultas de los usuarios empresariales se utilicen para entrenar ChatGPT en general. Esto garantiza que la organización mantenga el control sobre sus datos de capacitación, lo que permite respuestas precisas y muy relevantes para los usuarios empresariales y, al mismo tiempo, minimiza el riesgo de envenenamiento de datos proveniente de un lago de datos públicos.

Paso 4: Traslade eLLM detrás del inicio de sesión único (SSO) con una sólida autenticación multifactor (MFA)

Luego, la organización debería trasladar ChatGPT detrás de una arquitectura proxy en la nube Zero Trust, como Zscaler Zero Trust Exchange, para aplicar controles de seguridad Zero Trust sobre el acceso a ChatGPT. Esto también podría incluir ubicar ChatGPT detrás de un proveedor de identidad (IdP) con autenticación SSO y MFA sólida que incluya autenticación biométrica. Esto permitirá un inicio de sesión seguro y rápido del usuario en ChatGPT y al mismo tiempo permitirá a la empresa configurar controles de acceso granulares a nivel de usuario, equipo y departamento. Esto también garantiza una separación entre las consultas de los usuarios a esos mismos niveles de usuario, equipo y departamento.

Situar ChatGPT detrás de un proxy en la nube como el Zero Trust Exchange permite además a la organización inspeccionar todo el tráfico TLS/SSL entre los usuarios y ChatGPT para detectar ciberamenazas y fugas de datos mientras aplica siete capas distintas de seguridad Zero Trust.

Paso 5: Implemente el motor Zscaler DLP para evitar fugas de datos

Por último, la organización debe aplicar un motor DLP a la instancia ChatGPT para evitar la fuga accidental de información crítica, incluidos datos y código de propiedad, datos de clientes, datos personales, datos financieros y legales, etc. Esto garantiza que los datos altamente confidenciales nunca abandonen el entorno de producción.

Al seguir esta experiencia, los usuarios empresariales pueden aprovechar todos los beneficios de una herramienta de IA generativa como ChatGPT y, al mismo tiempo, eliminar los riesgos de datos más críticos al adoptar una aplicación de IA.

Mejores prácticas de IA

En general, las empresas pueden adoptar algunas mejores prácticas clave cuando se trata de integrar herramientas de IA en la misma.

- **Evalúe y mitigue continuamente los riesgos que conllevan las herramientas impulsadas por IA** para proteger la propiedad intelectual, los datos personales y la información de los clientes.
- **Asegúrese de que el uso de herramientas de IA cumpla con las leyes** y estándares éticos pertinentes, incluidas las regulaciones de protección de datos y las leyes de privacidad.
- **Establezca una responsabilidad clara para el desarrollo y la implementación de herramientas de IA**, incluidos roles y responsabilidades definidos para supervisar los proyectos de IA.
- **Mantenga la transparencia al utilizar herramientas de IA:** justifique su uso y comunique su propósito claramente a las partes interesadas.

Directrices de la política de IA

Las empresas deben seguir estas mejores prácticas y establecer un marco político claro que rija el uso aceptable en toda la empresa, la integración y el desarrollo de productos, las políticas de seguridad y datos y las mejores prácticas de los empleados cuando utilicen herramientas de IA. Las siguientes mejores prácticas pueden constituir un punto de partida útil para establecer políticas claras de IA.

- **No proporcione a los modelos de IA información de identificación personal (PII)** ni ninguna información no pública, de propiedad exclusiva o confidencial.
- **La IA no puede reemplazar a un ser humano** y no debe utilizarse para tomar decisiones sin la intervención humana adecuada.
- **El contenido generado por IA no debe usarse sin revisión y aprobación humana**, especialmente cuando el contenido representa a su organización.
- **El desarrollo y la integración de las herramientas de IA deben seguir un Marco Seguro del Ciclo de Vida del Producto** para garantizar el máximo nivel de seguridad.
- **Realice una diligencia debida exhaustiva del producto antes de implementar soluciones de IA**, asegurándose de evaluar sus implicaciones éticas y de seguridad.

Cómo Zscaler ofrece IA+ Zero Trust y protege la IA generativa

El poder transformador de la IA en la ciberseguridad radica en su capacidad de aprovecharse para combatir el panorama cambiante de las amenazas impulsadas por la IA. En Zscaler, utilizamos la IA para ayudar a las empresas a detener los ataques en todas las etapas de la cadena de ataque, así como a diagnosticar y mitigar riesgos fácilmente.

La clave para la ciberseguridad impulsada por la IA: datos de alta calidad a escala

Las empresas generan una gran cantidad de datos de registro que pueden contener señales de alta fidelidad que pueden indicar posibles vías para una violación. Sin embargo, los problemas de relación señal-ruido han dificultado históricamente el rápido aislamiento de estas señales. Mediante el uso de IA generativa, Zscaler puede aprovechar estos datos para mejorar eficazmente las medidas de triaje y protección mediante la comprensión de las vulnerabilidades y debilidades que los atacantes son propensos a explotar. Esto no solo permite a Zscaler predecir las violaciones antes de que se produzcan, sino que también ofrece a los ejecutivos una manera integral de visualizar y cuantificar la cibermadurez y el riesgo, al tiempo que prioriza los pasos de remediación de ciberseguridad con Zscaler Risk360.

Las capacidades de IA generativa no solo se extienden al metanálisis del ciberriesgo empresarial, sino que también se insertan directamente en los productos de ciberseguridad para detectar e interrumpir mejor las amenazas avanzadas en toda la cadena de ataque. Directamente integrados en la mayor nube de seguridad del mundo, los LLM y los modelos de IA de Zscaler aprovechan un lago de datos que ve más de 390,000 millones de transacciones diarias, con más de 9 millones de amenazas bloqueadas y 300 billones de señales. Es mucho más que “basura que entra, basura que sale”. Se trata de “datos e inteligencia de amenazas a gran escala y de alta fidelidad y una ciberseguridad de IA hiperconsciente y finamente ajustada”. Todo ello se traduce en resultados en materia de ciberseguridad más potentes y eficaces para los profesionales de la informática y la seguridad.





Aprovechar la IA en toda la cadena de ataque

Hemos analizado muchas maneras en que los malintencionados utilizan la IA para lanzar amenazas sofisticadas a mayor velocidad y escala. Zscaler implementa capacidades de IA en la plataforma Zero Trust Exchange y el conjunto de ciberproductos para identificar y detener ataques convencionales e impulsados por IA en cada etapa de la cadena de ataque.

Etapa 1: Descubrimiento de superficie de ataque

La primera etapa de un ciberataque generalmente implica que los malintencionados exploren la superficie de ataque empresarial conectada a Internet para identificar debilidades explotables. A menudo, esto incluye cosas como vulnerabilidades y configuraciones erróneas de VPN o firewalls o servidores sin parches. La IA generativa ha hecho que esta tarea que alguna vez fue ardua sea significativamente más fácil para los atacantes, quienes pueden simplemente consultar una lista de vulnerabilidades conocidas asociadas con estos activos.

Aprovechando los conocimientos impulsados por la IA en Zscaler Risk360, las empresas pueden ver al instante estas aplicaciones y activos descubribles (y por tanto riesgosos), su superficie de ataque conectada a Internet, y ocultarlos de la Internet pública detrás del Zero Trust Exchange. Esto reduce instantánea y drásticamente la superficie de ataque empresarial y al mismo tiempo evita que los atacantes descubran puntos de entrada débiles.

Etapa 2: Riesgo de vulneración

Durante la etapa de compromiso, los atacantes trabajan para explotar las vulnerabilidades y obtener acceso no autorizado a los sistemas o aplicaciones empresariales. Las innovaciones de la IA de Zscaler ayudan a reducir el riesgo de compromiso, desbaratando ataques sofisticados al tiempo que priorizan la productividad.

PREVENCIÓN DE PHISHING Y C2 BASADA EN IA

Los modelos de IA de Zscaler detectan sitios de phishing conocidos y de pacientes cero para evitar el robo de credenciales y la explotación del navegador, además de analizar patrones de tráfico, comportamiento y malware para detectar infraestructuras de comando y control (C2) nunca vistas en tiempo real. Estos modelos se basan en una combinación de inteligencia sobre amenazas, investigación de ThreatLabz y aislamiento dinámico del navegador para detectar sitios sospechosos. Como resultado, las empresas son aún más eficientes y efectivas a la hora de detectar nuevos ataques de phishing, incluidos los ataques generados por IA y dominios C2.

DEFENSA SANDBOX DE IA BASADA EN ARCHIVOS

Zscaler Sandbox en línea con tecnología de inteligencia artificial detecta instantáneamente archivos maliciosos y mantiene la productividad de los empleados. Las tecnologías tradicionales de sandbox obligan a los usuarios a esperar mientras se analizan los archivos, o bien a asumir un riesgo de paciente cero cuando se permiten los archivos en la primera pasada. Nuestra tecnología de IA Instant Verdict identifica, pone en cuarentena y previene instantáneamente archivos maliciosos de alta confianza (incluidas las amenazas de día cero) y, al mismo tiempo, elimina la necesidad de esperar el análisis de estos archivos. Esto incluye las amenazas que se envían a través de canales cifrados (TLS y HTTP) y otros protocolos de transferencia de archivos. Mientras tanto, los archivos benignos se entregan de manera segura e instantánea.

IA PARA BLOQUEAR AMENAZAS WEB

Zscaler Browser Isolation, impulsado por la IA, bloquea las amenazas de día cero y al mismo tiempo garantiza que los empleados puedan acceder a los sitios adecuados para realizar su trabajo. En la práctica, el filtrado de URL de la empresa requiere a menudo controles más granulares que permitir/bloquear. Los sitios bloqueados suelen ser seguros y necesarios para el trabajo, lo que da lugar a tickets innecesarios para el servicio de asistencia. Nuestro Smart Isolation (Aislamiento Inteligente) de IA puede identificar cuándo un sitio puede ser riesgoso y abrirlo de manera aislada para el usuario, transmitiendo el sitio como píxeles en un entorno seguro y contenido. Esto detiene eficazmente las amenazas basadas en la web, como malware, ransomware, phishing y descargas no autorizadas, estableciendo una postura de seguridad web sólida sin necesidad de que las empresas bloqueen demasiado los sitios de manera predeterminada.



Etapa 3: Movimiento lateral

Una vez que los atacantes se afianzan dentro de una organización, intentarán moverse lateralmente para acceder a datos y aplicaciones confidenciales. Y para muchas organizaciones, el acceso de los usuarios está enormemente sobreadimensionado a docenas de aplicaciones críticas, lo que significa que su superficie de ataque interna es sustancial.

Las capacidades de IA de Zscaler reducen el radio potencial de los ataques al analizar los patrones de acceso de los usuarios y recomendar políticas inteligentes de segmentación de aplicaciones para limitar el riesgo lateral. Por ejemplo, es común ver que solo 200 usuarios de 30,000 con acceso a una aplicación financiera realmente la necesitan. Zscaler puede crear automáticamente un segmento de aplicación que limita el acceso solo a esos 200 empleados, reduciendo las oportunidades de movimiento lateral de los malintencionados en más de 99 %.

Etapa 4: Exfiltración de datos

En la etapa final de un ataque, los malintencionados trabajan para extraer datos confidenciales. Zscaler utiliza IA para permitir a las organizaciones implementar protecciones de datos más rápidamente. El descubrimiento de datos impulsado por la IA elimina la laboriosa tarea de individualización y clasificación de datos, que de otro modo puede retrasar o impedir la implementación. La IA de Zscaler descubre y clasifica automáticamente todos los datos de una organización desde el primer momento, lo que permite a las empresas clasificar de inmediato la información confidencial a la vez que configuran las políticas de prevención de pérdida de datos (DLP) para evitar que esos datos salgan de la organización en caso de ataque o violación.

Resumen de las ofertas de Zscaler basadas en IA

Zscaler Internet Access™ proporciona protección basada en IA para usuarios, dispositivos y aplicaciones web y SaaS empresariales en todas las ubicaciones como parte de Zero Trust Exchange, y ofrece:

- **Detección de phishing y C2 potenciada por IA** contra sitios de phishing e infraestructura C2 nunca antes vistos, utilizando la detección en línea basada en IA de Zscaler Secure Web Gateway (SWG).
- Sandboxing potenciado por IA con prevención integral de malware y amenazas de día cero.
- **Política dinámica basada en el riesgo** con análisis continuo del riesgo del usuario, el dispositivo, la aplicación y el contenido para impulsar la seguridad dinámica y la política de acceso.
- **Segmentación impulsada por la IA** con Zscaler Private Access™, con recomendaciones de políticas de acceso automatizadas para minimizar la superficie de ataque y detener el movimiento lateral utilizando el contexto, el comportamiento, la ubicación y la telemetría de la aplicación privada del usuario.
- Aislamiento del navegador impulsado por IA, que crea una brecha segura entre los usuarios y las categorías web maliciosas, renderizando el contenido como un flujo de imágenes perfectas para eliminar las fugas de datos y la distribución de amenazas activas.

ADEMÁS, ZSCALER BLOQUEA:

Las URL e IP observadas en la nube de Zscaler y de fuentes de información de amenazas comerciales y de código abierto integradas de manera nativa. Esto incluye las categorías de URL de alto riesgo definidas por la política y utilizadas habitualmente para el phishing, como los dominios recién observados y los recién activados.

Firmas IPS desarrolladas a partir del análisis de ThreatLabz de kits y páginas de phishing.

Zscaler Risk360 ofrece un marco de riesgo integral y procesable que ayuda a los líderes empresariales y de seguridad a cuantificar y visualizar el ciberriesgo en toda la empresa.

Data Protection con DLP y CASB ofrece clasificación y protección de datos impulsada por IA en todos los canales, incluidos puntos finales, correo electrónico, cargas de trabajo, dispositivos propios y postura en la nube.

Advanced Threat Protection bloquea todos los dominios C2 conocidos.

Zscaler ITDR (Identity Threat Detection and Response) mitiga el riesgo de ataques basados en la identidad sin visibilidad continua, supervisión de riesgos y detección de amenazas.

Zscaler Firewall extiende la protección C2 a todos los puertos y protocolos, incluidos los destinos C2 emergentes.

DNS Security defiende contra ataques basados en DNS e intentos de exfiltración.

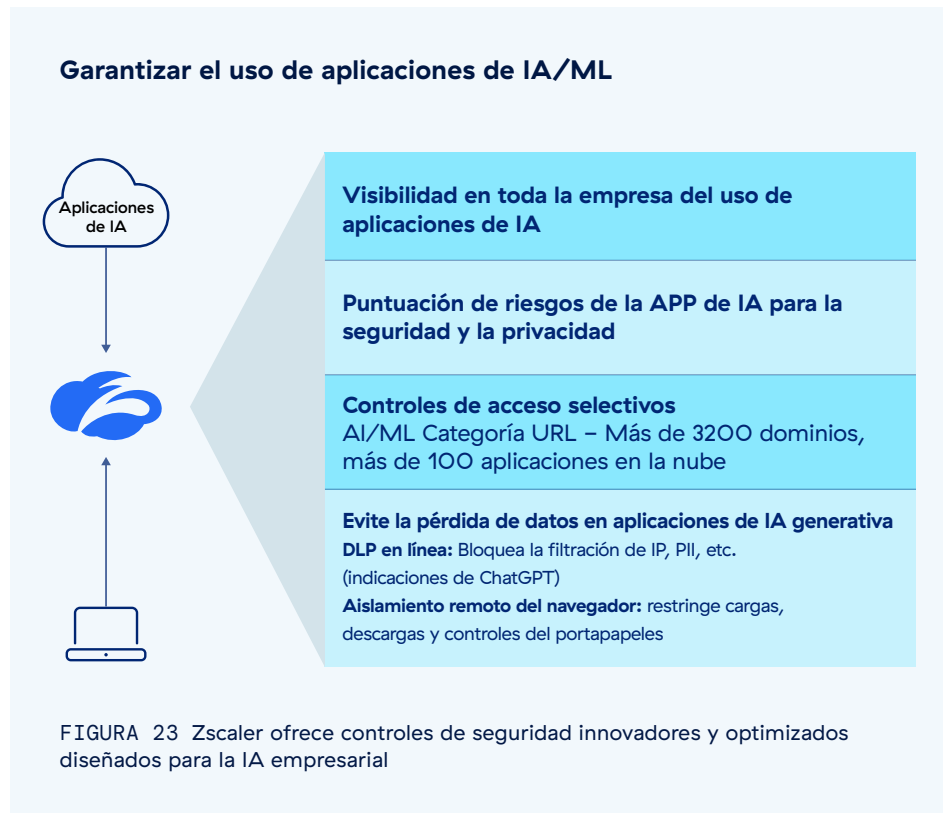
Zscaler Private Access™: el acceso privado de Zscaler protege las aplicaciones al limitar el movimiento lateral mediante la segmentación usuario-a-aplicación de acceso con privilegios mínimos y la inspección completa en línea del tráfico de aplicaciones privadas.

AppProtection con Zscaler Private Access proporciona una inspección de seguridad en línea de alto rendimiento de toda la carga útil de la aplicación para exponer amenazas.

Zscaler Deception™ el servicio de engaño de Zscaler detecta y contiene a los atacantes que intentan moverse lateralmente o escalar privilegios atrayéndolos con servidores, aplicaciones, directorios y cuentas de usuario señuelo.

Habilitar la transición a la IA empresarial: el control está en sus manos

Zscaler ofrece a las empresas una manera de fomentar la innovación, la creatividad y la productividad con aplicaciones de IA al tiempo que mantienen a salvo a los usuarios y los datos entre los canales emergentes para la exfiltración de datos. Esto permite a las empresas aprovechar el [potencial transformador de la IA](#) para acelerar su actividad sin bloquear directamente las aplicaciones y dominios de IA.



ZSCALER PERMITE A LAS EMPRESAS:

- 01 Impulsar la visibilidad total del uso de herramientas de IA**
Los registros detallados brindan visibilidad completa de cómo los equipos empresariales utilizan la IA, incluidas las aplicaciones y dominios que visitan, así como los datos y las indicaciones que se utilizan en herramientas como ChatGPT.
- 02 Crear políticas flexibles para ajustar el uso de la IA**
Un filtrado de URL potente y personalizado para aplicaciones de IA y ML permite a las empresas definir y aplicar fácilmente controles de acceso y segmentación granulares de la IA, bloqueando el acceso cuando sea necesario y permitiendo al mismo tiempo el acceso con niveles de riesgo aceptables mediante la puntuación de riesgos de las aplicaciones de IA. Las empresas pueden permitir el acceso a nivel de empresa, departamento, equipo y usuario, así como habilitar un acceso basado en la precaución que instruya a los usuarios sobre los riesgos de las herramientas de IA generativa. La segmentación basada en la IA facilita la identificación de los segmentos de usuarios adecuados para acceder a determinadas aplicaciones de IA, al tiempo que minimiza la superficie de ataque interna asociada a las herramientas de IA.
- 03 Aplicar la seguridad de datos granular para ChatGPT y otras aplicaciones de IA**
Las empresas pueden evitar la fuga de datos confidenciales cargados en aplicaciones de IA con controles granulares de Zscaler Cloud Application para IA generativa. Al aplicar el motor DLP de Zscaler, las empresas pueden asegurarse de que no se compartan datos accidentalmente al utilizar cualquier herramienta de IA. Mientras tanto, el descubrimiento y la clasificación de datos impulsados por la IA permiten a las empresas identificar y crear fácilmente políticas de DLP en torno a sus datos más críticos, incluyendo su base de código corporativo, documentos financieros y legales, datos personales, datos de clientes y más. [Este video](#) demuestra cómo el motor DLP impide que los usuarios introduzcan información sobre tarjetas de crédito en ChatGPT.
- 04 Habilitar controles potentes utilizando Browser Isolation**
Zscaler Browser Isolation renderiza las aplicaciones de IA en un entorno seguro, agregando una capa de protección que permite peticiones y consultas del usuario a las herramientas de IA mientras restringe copiar/pegar, las cargas y las descargas. Esto ayuda a mitigar el riesgo de que se compartan accidentalmente datos confidenciales con las herramientas de IA generativa.

Los líderes empresariales y de seguridad se encuentran en una encrucijada: deben trabajar para adoptar la IA para impulsar la innovación y seguir siendo competitivos, pero también deben asegurarse de que sus datos solo impulsen el negocio, no las violaciones. Zscaler permite a las empresas navegar por esta transición con confianza, aprovechando una suite completa de controles de seguridad Zero Trust impulsados por la IA que protegen contra los ataques impulsados por la IA al tiempo que ofrecen políticas de IA afinadas y las protecciones de datos necesarias para aprovechar todo el potencial de la IA generativa.

Anexo

Metodología de investigación de ThreatLabz

La nube de seguridad global Zscaler procesa más de 300 billones de señales diarias y bloquea 9 mil millones de amenazas y violaciones de políticas por día, con más de 250,000 actualizaciones de seguridad diarias. Análisis de 18,090 millones de transacciones de IA y ML desde abril de 2023 hasta enero de 2024 en la nube de Zscaler, el Zero Trust Exchange.

Acercas de Zscaler 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.





Experimente su mundo, protegido.

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 150 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.

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

© 2024 Zscaler, Inc. Todos los derechos reservados. Zscaler™, Zero Trust Exchange™, Zscaler Internet Access™, ZIA™, Zscaler Private Access™, ZPA™ y otras marcas registradas listadas en zscaler.com.mx/legal/trademarks son (i) marcas comerciales o marcas de servicio registradas o (ii) marcas comerciales o marcas de servicio de Zscaler, Inc. en los Estados Unidos y otros países. Cualquier otra marca comercial pertenece a sus respectivos propietarios.

zscaler.com.mx