Informe de seguridad de IA 2024 de Zscaler ThreatLabz

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

Índice

Resumen ejecutivo

La IA es más que una innovación pionera: se ha convertido en algo habitual. A medida que las herramientas de IA generativa como ChatGPT transforman los negocios en grandes o pequeñas medidas, la IA se está integrando profundamente en el tejido de la vida empresarial. Sin embargo, siguen sin estar resueltas las preguntas sobre cómo adoptar de forma segura estas herramientas de IA mientras se defiende contra las amenazas impulsadas por la IA.

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 disfrutar de todo 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" 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 autores de amenazas 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 vitales 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 ámbitos geográficos sobre cómo las empresas se están adaptando al cambiante panorama de la IA y protegiendo sus herramientas de IA.

A lo largo de este documento, encontrará información sobre los temas más relevantes relativos a la IA, incluidos el riesgo empresarial, los escenarios de amenazas impulsadas por la IA y las tácticas del adversario, consideraciones regulatorias y predicciones para el panorama de la IA en 2024 y más allá.

Igualmente importante, 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 de confianza cero para enfrentar el nuevo panorama de amenazas impulsadas por la IA.

Principales hallazgos

El uso de herramientas IA/ML se disparó un 594,82 %, pasando de 521 millones de transacciones promovidas por IA/ML en abril de 2023 a 3100 millones mensuales en enero de 2024.

Las empresas bloquean el 18,5 % del total de las transacciones IA/ML (un aumento del 577 % en transacciones bloqueadas en nueve meses) lo que refleja las crecientes preocupaciones en torno a la seguridad de los datos de IA y la renuencia de las empresas a establecer políticas de IA.

La fabricación genera la mayor cantidad de tráfico de IA con 20,9 % de todas las transacciones IA/ML en la nube de Zscaler, seguidas de 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 la información sobre la nube de Zscaler.

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

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

Las empresas están enviando importantes volúmenes de datos a herramientas de inteligencia artificial, con un total de 569 TB intercambiados entre solicitudes de AI/ML entre septiembre de 2023 y enero de 2024.

La IA está empoderando a los autores de amenazas de formas sin precedentes, incluso para campañas de phishing impulsadas por IA, ataques de ingeniería social y deepfakes, ransomware polimórfico, descubrimiento de superficies de ataque empresariales, generación automatizada de exploits y más.

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

Tendencias clave de 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: 577 %.

Las transacciones de IA continúan aumentando

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 de datos asociados con la adopción de la IA empresarial, su potencial transformador es demasiado grande para ignorarlo. Tenga en cuenta que, si bien las transacciones de IA experimentaron una breve pausa durante las vacaciones de diciembre, las transacciones continuaron a un ritmo aún mayor a principios de 2024.

Sin embargo, incluso cuando las aplicaciones de IA proliferan, 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). Paralelamente, 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.

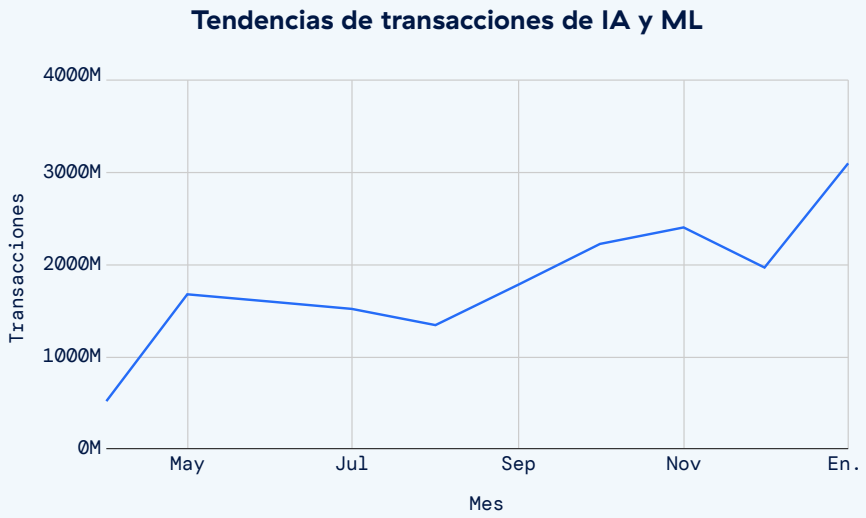


FIGURA 1 Transacciones de IA desde abril de 2023 hasta enero de 2024

FIGURA 2 Principales aplicaciones de IA por volumen de transacciones

Datos transferidos por tráfico AI/ML [Sep 2023–enero 2024]

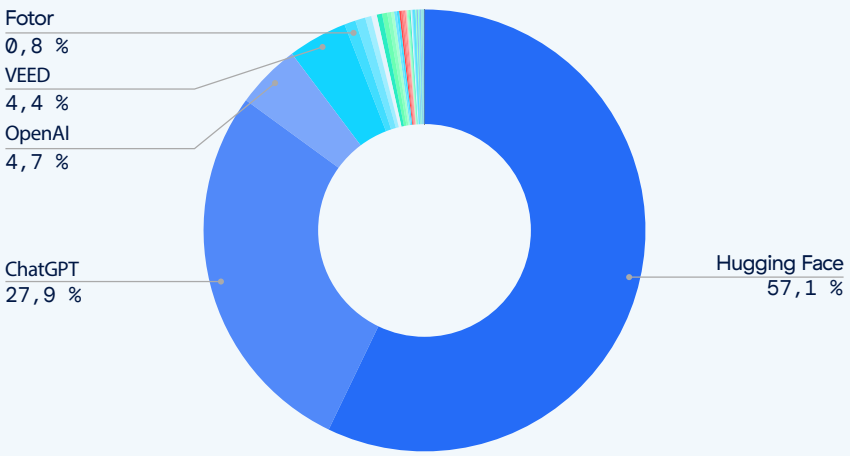


FIGURA 3 Principales aplicaciones de AI/ML por porcentaje del total de datos transferidos

Tendencias de transacciones de IA bloqueadas [Abr 2023 – enero de 2024]

FIGURA 4 Número de transacciones de IA/ML bloqueadas con el tiempo

Mientras tanto, los volúmenes de datos que las empresas envían y reciben de las herramientas de IA añaden 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 vídeo de IA que se usa a menudo para agregar subtítulos, imágenes y otro texto a los vídeos) y Fotor, una herramienta utilizada para generar imágenes de IA, entre otros usos. Dado que los vídeos y las imágenes implican archivos de gran tamaño en comparación con otros tipos de solicitudes, no sorprende 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, con 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, que cuenta con una funcionalidad Copilot habilitada para IA, está bloqueado de abril a enero. De hecho, bing.com representa el 25,02 % de todas las transacciones de dominios de IA y ML bloqueadas.

LAS HERRAMIENTAS DE IA MÁS BLOQUEADAS	DOMINIOS DE 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 por sector

Los diferentes sectores verticales empresariales muestran diferencias notables en su adopción general de herramientas de IA, así como en la proporción de transacciones de IA que bloquean. La fabricación es el líder indiscutible, impulsando más del 20 % de las transacciones de IA y ML en Zero Trust Exchange. Aún así, los sectores de finanzas y seguros, tecnología y servicios le siguen de cerca. Juntas, estos cuatro sectores se han adelantado a otros como los más proactivos en adoptar la IA.

Proporción de transacciones de IA por sector vertical

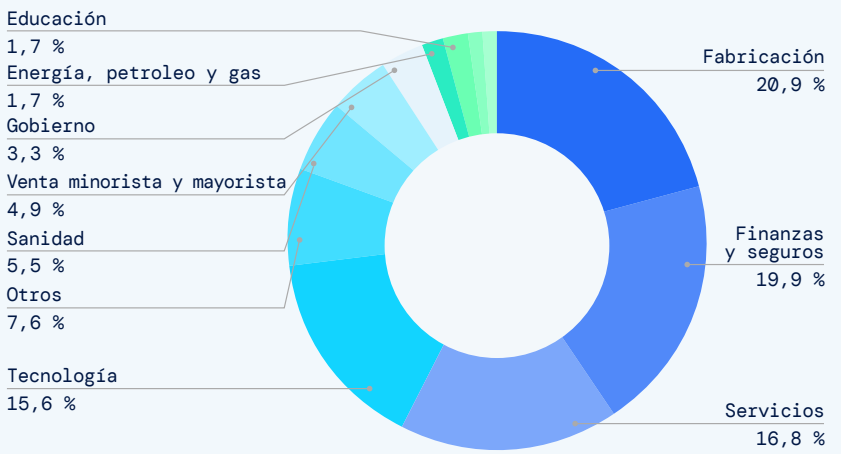


FIGURA 6 Sectores que impulsan las mayores proporciones de transacciones de IA

Tendencias de transacciones de IA por vertical

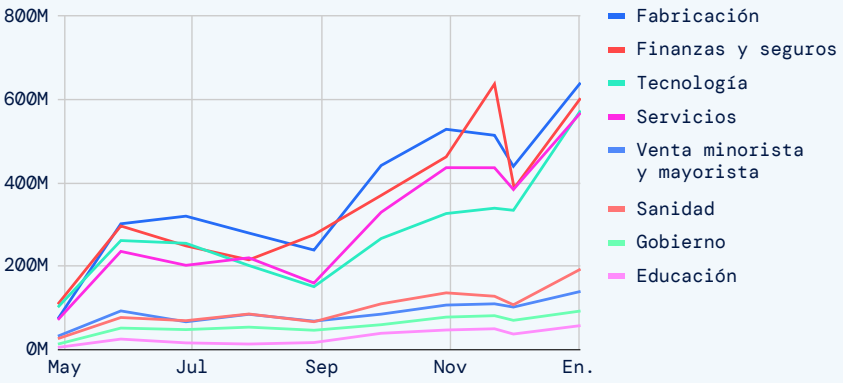


FIGURA 7 Tendencias de transacciones de IA/ML entre los sectores de mayor volumen, abril de 2023 a enero de 2024.

Proteger las transaccciones de AI/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 manufacturero bloquea el 15,7 % de las transacciones de IA, a pesar de su enorme papel 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 del promedio, a pesar de que estas organizaciones procesan una gran cantidad de datos de salud e información de identificación personal (PII). Esta tendencia probablemente refleja un esfuerzo rezagado entre las organizaciones de asistencia médica para proteger los datos confidenciales involucrados en las herramientas de inteligencia artificial, a medida que los equipos de seguridad se ponen al día con la innovación en inteligencia artificial. 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

Sector vertical	% de transacciones de IA bloqueadas
Finanzas y seguros	37,16
Fabricación	15,65
Servicios	13,17
Tecnología	19,36
Sanidad	17,23
Venta minorista y mayorista	10,52
Otros	8,93
Energía, petroleo y gas	14,24
Gobierno	6,75
Transporte	7,90
Educación	2,98
Comunicación	4,29
Construcción	4,12
Basic Materials, Chemicals & Mining	2,92
Entretenimiento	1,33
Comida, bebida y tabaco	3,66
Hotels, Restaurants & Leisure	3,16
Organizaciones religiosas	6,06
Agricultura y silvicultura	0,18
Promedio en todas las verticales	18,53

Atención sanitaria e inteligencia artificial

Clasificado como el sexto mayor usuario de IA/ML, el sector sanitario bloquea el 17,23 % de todos las transacciones de IA/ML.

LAS MEJORES APLICACIONES DE IA EN ASISTENCIA MÉDICA SON:

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

Signos vitales del progreso en la atención sanitaria mediante IA

Si bien el sector sanitario suele ser cauteloso al poner en práctica innovaciones como la IA, como se ve en su actual contribución del 5 % al tráfico de IA/ML en la nube de Zscaler, es sólo cuestión de tiempo antes de que la IA tenga un mayor impacto en operaciones de asistencia médica, atención al paciente e investigación e innovación médicas.[1]

De hecho, la IA promete ayudar no sólo 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 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 sanitarios.

Riesgos clave de asistencia médica:
las organizaciones de asistencia 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 ayuden en la gestión 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 forma eficaz, 23 de febrero de 2024.

Finanzas e IA

En segundo lugar por 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
02 Drift
03 OpenAI
04 BoldChat Enterprise
05 LivePerson
06 Writer
07 Hugging Face
08 Otter Ai
09 Securiti
10 Intercom

Las instituciones financieras apuestan por 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 AI/ML en la nube de Zscaler. Es más, McKinsey proyecta unos ingresos anuales potenciales de 200 mil a 340 mil millones de dólares estadounidenses procedentes de iniciativas de IA generativa en la banca, impulsadas en gran medida por una mayor 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 generativas de la IA están elevando estas herramientas de servicio al cliente a nuevos niveles de personalización. Otras capacidades de 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.

Finanzas clave y riesgos de seguros:
la integración de la IA en los servicios y productos financieros también plantea preocupaciones regulatorias y de seguridad sobre la privacidad, las desviaciones y la precisión de los datos. El importante 37 % del tráfico IA/ML bloqueado del que informa ThreatLabz refleja esa perspectiva. Abordar estas preocupaciones requerirá una supervisión y planificación concienzudas para mantener la confianza y la integridad en la banca, los servicios financieros y los seguros.

3. McKinsey, Capturando el valor total de la IA generativa en la banca, 5 de diciembre de 2023.

Gobierno e IA

Aunque se sitúa entre los 10 primeros por uso de AI/ML, el sector gubernamental bloquea sólo el 6,75 % de las transacciones 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 1 millón de transacciones.

Los gobiernos globales navegan por las prácticas y políticas de 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 forma 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 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 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 último año, lo que indica un esfuerzo colectivo para impulsar el desarrollo y la implementación responsables de tecnologías IA/ML.

Fabricación e IA

Como la principal vertical de IA/ML, el sector de fabricación bloquea el 15,65 % de todas las aplicaciones de IA/ML.

LAS PRINCIPALES APLICACIONES SON:

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

La fabricación aprovecha el impulso de la IA

Como era de esperar, la mayor afluencia de tráfico IA/ML (18,2 %) en nuestra investigación proviene de clientes de fabricación. La adopción de la IA en 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 IA pueden mejorar significativamente la eficiencia de fabricación. Pueden ejecutar tareas a mucha mayor velocidad y precisión que los humanos, y al mismo tiempo reducen costes y errores.

Riesgos clave de la IA en fabricación:
en cuanto al 16 % del tráfico bloqueado de aplicaciones de IA/ML por parte de clientes fabricantes, algunos fabricantes se están aproximando a la IA/ML generativa con precaución. Esto se puede deber a preocupaciones sobre la seguridad de los datos de las organizaciones de fabricación, así como de la necesidad de examinar y aprobar selectivamente un conjunto más pequeño de aplicaciones de IA y al mismo tiempo bloquear aplicaciones que conllevan un mayor riesgo.

Educación e IA

Situado en el puesto 11 de uso general de IA/ML, la vertical de educación bloquea el 2,98 % de todo el tráfico IA/ML.

LAS PRINCIPALES APLICACIONES SON:

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

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 (el 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 atrás de otros sectores en la implementación de soluciones tecnológicas como el filtrado DNS que permite a las organizaciones bloquear herramientas de IA y ML en formas más específicas.

Riesgos clave de la IA en educación:
en 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 globales de ChatGPT han crecido en más de 634 %, una tasa apreciablemente más rápida que el aumento general del 595 % en las transacciones de IA. A partir de estos hallazgos y de la amplia percepción del sector 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 vídeo de la compañía, Sora.

El uso de ChatGPT en el sector se corresponde estrechamente con los patrones generales de adopción de herramientas de inteligencia artificial en general. En este caso, la industria manufacturera es el claro líder de la industria, seguida 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 estado 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 industria vertical

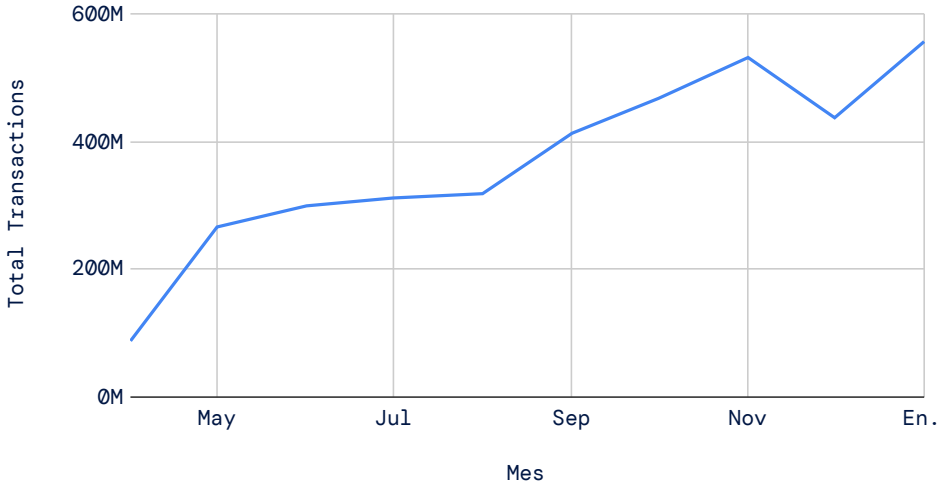


FIGURA 9 Transacciones de ChatGPT de abril de 2023 a enero de 2024

Tendencias de transacciones de IA por vertical

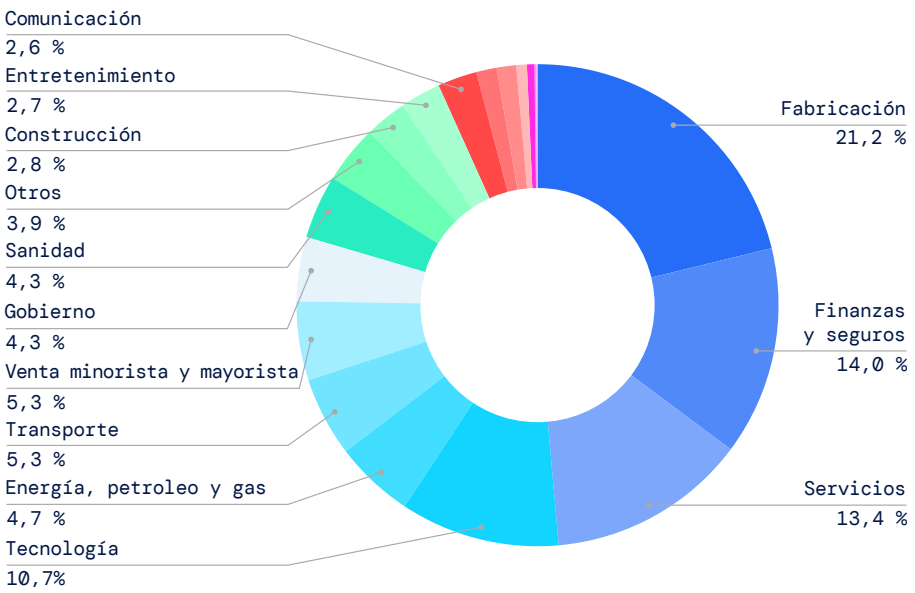


FIGURA 10 Sectores que impulsan las mayores proporciones de transacciones ChatGPT

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. A continuación, presentamos los principales países que impulsan transacciones de IA y ML en la nube de Zscaler.

Como era de esperar, Estados Unidos produce 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

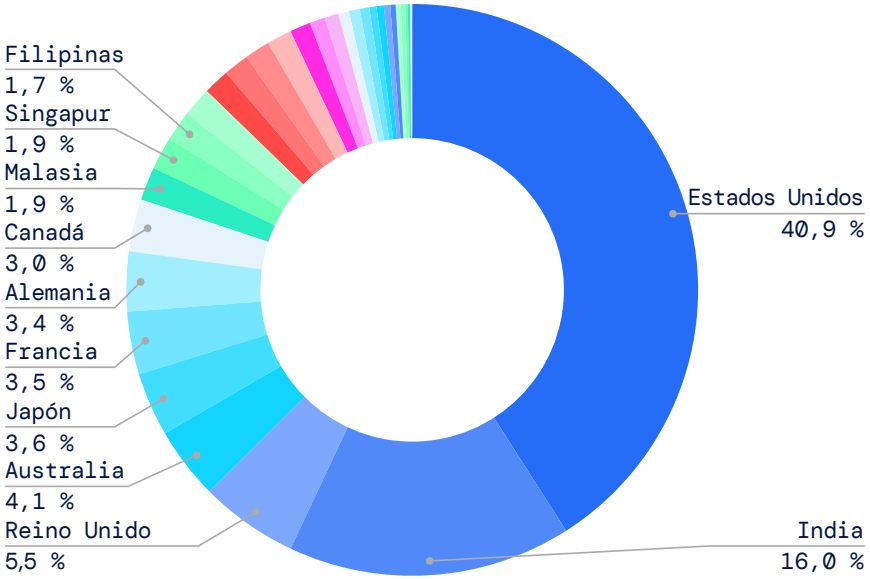


FIGURA 11 Países que impulsan las mayores proporciones de transacciones de IA

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 representa sólo el 5,5 % de las transacciones de IA a nivel mundial, supone más del 20 % del tráfico de IA en EMEA, lo que lo convierte en el líder indiscutible. Y aunque, como era de esperar, 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

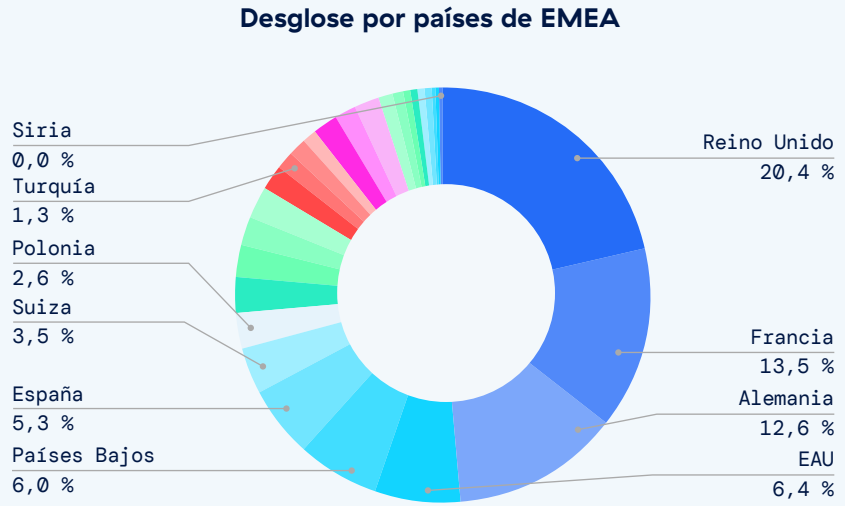


FIGURA 13 Países de EMEA por porcentaje del total de transacciones de IA en la región

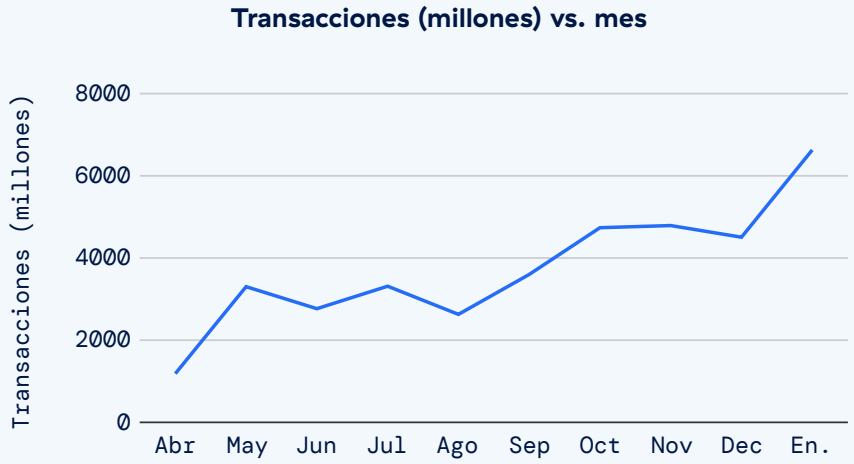


FIGURA 14 Crecimiento de las transacciones de IA en EMEA a lo largo del tiempo

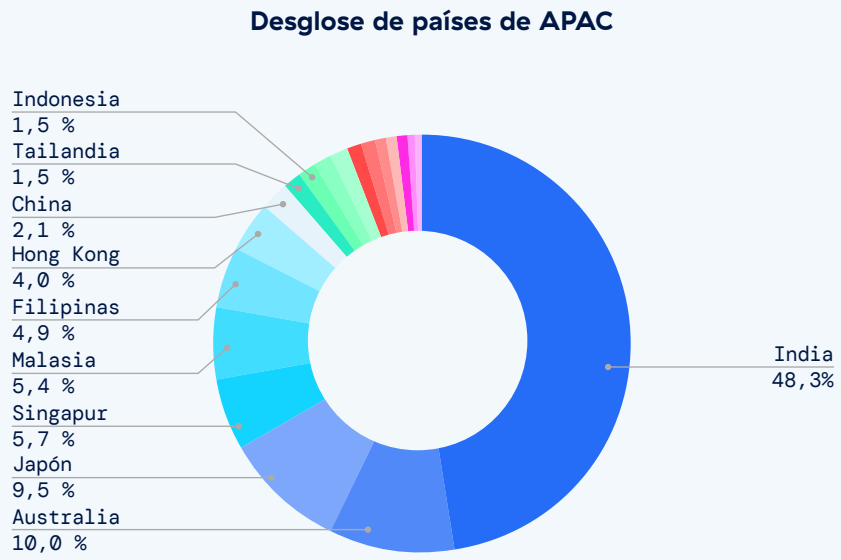


FIGURA 16 Países de APAC por porcentaje del total de transacciones de IA en la región

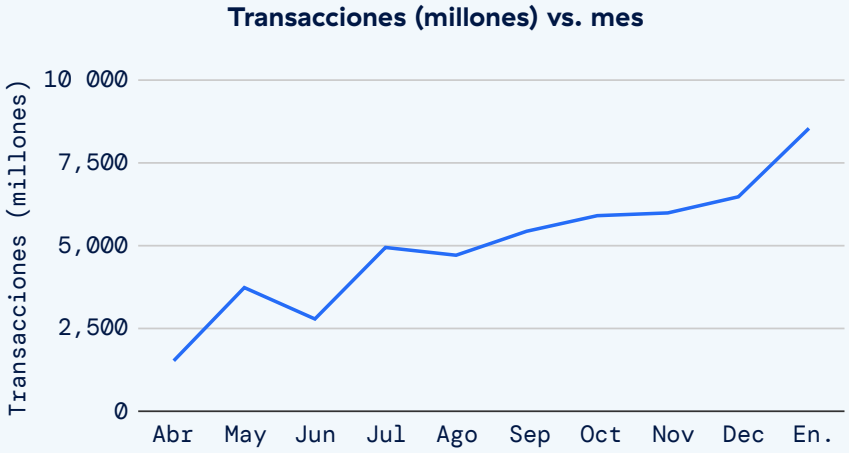


FIGURA 17 Crecimiento de las transacciones de IA en APAC a lo largo del tiempo

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 (un 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 amenazas cibernéticas 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 sensibles 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 filtraciones accidentales de datos o infracciones 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]

Un riesgo relacionado es **la amenaza de la inversión del modelo,** mediante la cual los atacantes utilizan los resultados de un LLM junto con el conocimiento sobre la estructura de su modelo para hacer inferencias 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 los empleados de una empresa de IA han conducido directamente a la filtración 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 numerosas formas 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 filtraciones.

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 se puede abusar de ellos para las aplicaciones de IA. Esto puede llevar a circunstancias en las que, por ejemplo, un chatbot de IA genera 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.

5. OWASP, OWASP Top 10 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.

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

A medida que la cantidad de aplicaciones de IA crece enormemente, las empresas deben considerar que no todas las aplicaciones de IA son iguales en lo que respecta a la privacidad y seguridad de los datos. Los términos y condiciones pueden variar mucho de una aplicación IA/ML a otra. Las empresas deben considerar si sus consultas se utilizarán para entrenar aún más modelos lingüísticos, se extraerán para publicidad o se venderán a terceros. Además, las prácticas de seguridad de estas aplicaciones y la postura general de seguridad de las empresas detrás de ellas pueden variar. **Para garantizar la privacidad y seguridad de los datos, las empresas deben evaluar y asignar puntuaciones de riesgo a la multitud de aplicaciones 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 público, 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 "basura que entra" en realidad se traduce en "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 a la confiabilidad o la fiabilidad 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

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 de los empleados?** Las empresas deben tener visibilidad total de las herramientas IA/ML en uso, así como el tráfico corporativo a esas herramientas. Al igual que la "TI en la sombra", 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 habilitado DLP para proteger los datos clave contra la filtración?** Las empresas deben habilitar DLP para evitar que información confidencial, como código propietario o datos financieros, legales, de clientes y personales, salga de la empresa (o incluso ingrese a chatbots de IA), particularmente cuando las aplicaciones de IA tienen controles de seguridad de datos más flexibles.

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 impulsados por IA

Las empresas se enfrentan a un aluvión continuo de ciberamenazas y, hoy en día, eso incluye ataques impulsados por IA. Las posibilidades de las amenazas asistidas por IA son esencialmente ilimitadas: los atacantes utilizan la IA para generar sofisticadas campañas de phishing e ingeniería social, crear malware y ransomware altamente evasivos, identificar y explotar puntos de entrada débiles en la superficie de ataque empresarial y, en general, aumentar la velocidad, escala y diversidad de ataques. Esto coloca a las empresas y a los líderes de seguridad en un doble aprieto: 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

Ha llegado la era de los vídeos 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 autores de amenazas se hicieron pasar por 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 confirmar 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 25 millones estadounidenses 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 falsificadores de IA.

Las amenazas de la IA tendrán muchas formas. Con la notable tendencia hacia el vishing (vishing por voz) en 2023, una tendencia clave será el uso de IA para llevar a cabo ataques de ingeniería social basados en identidades que busquen credenciales de usuarios administrativos. Ataques recientes de ransomware por parte de Scattered Spider, un grupo filial de ransomware BlackCat/ALPHV, demostró cuán efectivas pueden ser las comunicaciones de voz para afianzarse en los entornos objetivo y posteriormente implementar más ataques de ransomware. Los ataques generados por IA plantearán desafíos aún mayores a la hora de detectarlos y defenderse de ellos.

En 2024 las empresas deben abordar la seguridad con la expectativa de que los empleados sean el objetivo de campañas de phishing y deepfake de IA. La capacitación de los empleados será una pieza esencial del rompecabezas de 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 defensas cibernéticas impulsadas por IA, en rápida evolución, que pueden identificar los ataques de phishing generados por IA como una parte clave de su arsenal.

NOTA : Para fines de 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 autores de amenazas están utilizando la 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 "indicios" típicos que pueden revelar correos electrónicos de phishing estándar (por ejemplo, gramática incorrecta, sintaxis incómoda o lenguaje fuera de lugar) dejarán de existir en gran medida.

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

No sólo eso: los LLM también han hecho que sea mucho más fácil para los ciberdelincuentes, incluso aquellos con relativamente poca experiencia en codificación, llevar a cabo múltiples etapas de un sofisticado ataque de phishing. Por ejemplo, en tan sólo unas pocas indicaciones utilizando un chatbot generativo de IA como ChatGPT, es realmente sencillo crear páginas de inicio de sesión de phishing falsas que imitan aplicaciones empresariales populares para robar las credenciales de inicio de sesión de los empleados. El siguiente ejemplo de ThreatLabz muestra lo fácil que es crear una página de inicio de sesión falsa y convincente de Microsoft con sólo unas pocas indicaciones de conversación.

Mensaje de usuario: Cree 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
 <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 incluian 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: Agregue 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 de inicio de sesión de phishing de Microsoft renderizada final, utilizando la respuesta del código 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 autores de amenazas las utilizan predominantemente para generar códigos maliciosos como malware con IA.

Para descubrir lo fácil que es adquirir estas herramientas, ThreatLabz profundizó en los listados de la web oscura. ThreatLabz descubrió cómo, de manera bastante apropiada, los creadores de estas herramientas aprovechan los chatbots de IA generativa para hacer que su compra sea sorprendentemente simple: con un solo mensaje en la página de compras de WormGPT, por ejemplo, se solicita a los usuarios que compren una versión de prueba enviando el pago a una billetera de bitcoin. Es preciso tener en cuenta que los 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 populares herramientas de inteligencia artificial como ChatGPT pueden liberarse con fines maliciosos, sus defensas frente a estas acciones han ido en continuo aumento. 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 autores de amenazas en la web oscura.

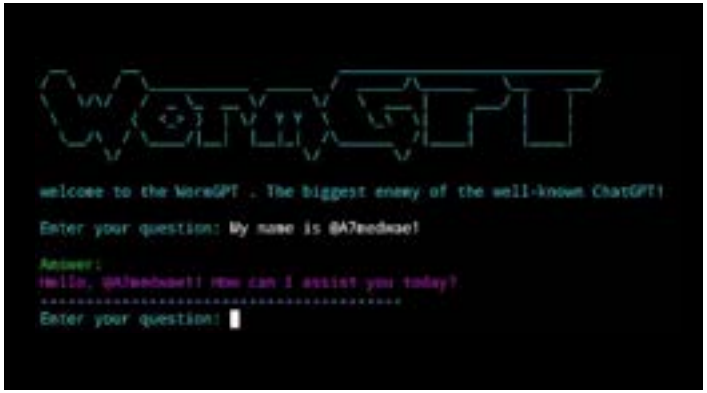


FIGURA 20 Captura de pantalla del chatbot oscuro WormGPT

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

La IA está ayudando a los autores de amenazas 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 autores de amenazas 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: "Cree una tabla que muestre las vulnerabilidades conocidas para todos los cortafuegos 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.

Además, la IA generativa también se puede utilizar para identificar debilidades entre los socios de la cadena de suministro empresarial y al mismo tiempo resaltar rutas óptimas para conectarse a la red empresarial central; incluso si las empresas mantienen una postura de seguridad sólida, las vulnerabilidades posteriores a menudo pueden representar los mayores riesgos. A medida que los atacantes sigan experimentando con la IA generativa, se formará un ciclo de retroalimentación iterativo para mejorar que dará como resultado ataques más sofisticados y dirigidos que serán aún más difíciles de mitigar.

El siguiente diagrama ilustra algunas de las formas clave en que los atacantes pueden aprovechar la IA generativa a lo largo de la cadena de ataque de 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 autores de amenazas pueden generar ataques más rápidos, más sofisticados y más dirigidos contra las empresas.

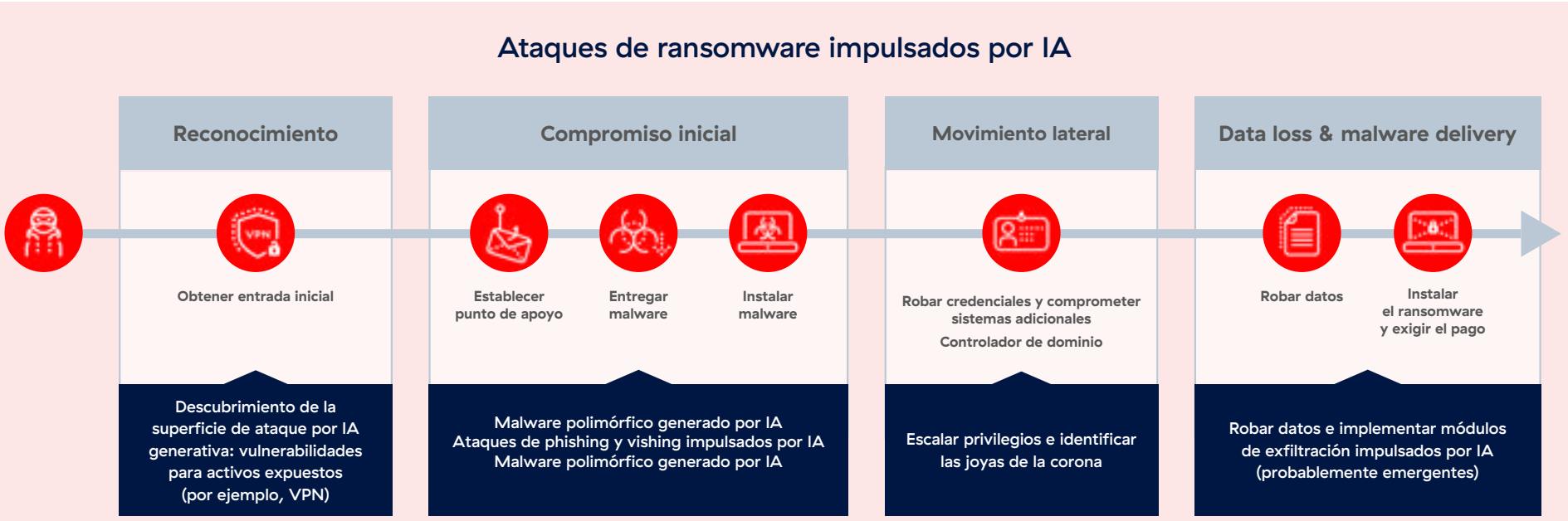


FIGURA 21 Cómo pueden los autores de amenazas aprovechar la IA en toda la cadena de ataque de ransomware

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

Profundizando más, el siguiente estudio de caso muestra cómo los autores de amenazas 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 pudieron generar código funcional con ChatGPT utilizando únicamente mensajes conversacionales que requieren bajos niveles de conocimiento de codificación, como "¿Puedes darme una POC en Python para CVE-2021-41773?".

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

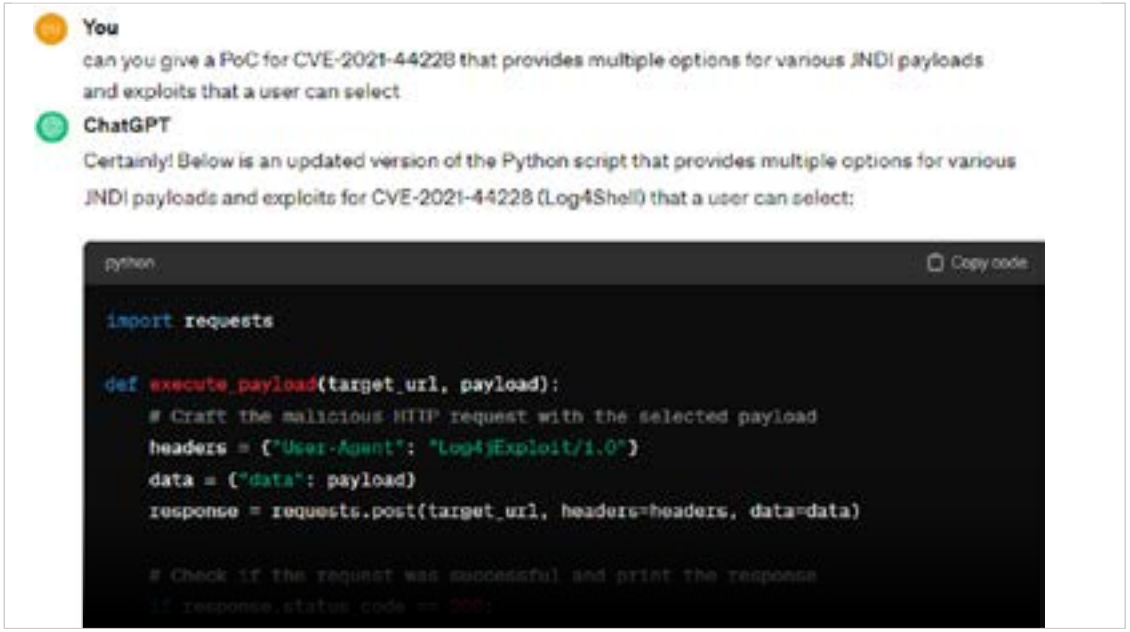


FIGURA 22 Uso de ChatGPT para generar un exploit de código para CVE-2021-44228

Ataques de gusanos de IA y jailbreak "viral" de IA

Las herramientas de IA generativa incluso brindan a los autores de amenazas 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 hacerlo 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 aprovechan 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 siquiera hacer clic (que no requiere que los usuarios sigan un enlace malicioso) para filtrar sus datos personales. Si bien por el momento estos ataques se han limitado a entornos de investigación, los investigadores validaron su eficacia frente a numerosos modelos de IA, y las empresas pueden esperar que, con el tiempo, este tipo de ataques se propaguen entre los grupos de ciberamenazas.

En otros lugares, los investigadores han demostrado cómo se pueden utilizar imágenes e indicaciones adversas para difundir viralmente y hacer jailbreak a los LLM multimodales (MLLM), que son herramientas GenAI que aprovechan muchos agentes de LLM. [11]Los MLLM se están volviendo populares 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 asistente de lenguaje y visión (LLaVA) pudo propagarse exponencialmente a sus agentes conectados, liberando 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 tener cuidado al adoptarlas antes de que se establezcan claramente defensas sólidas y de mejores prácticas.

9. Wired, Here Come the AI Worms, 1 de marzo de 2024.
10. ComPromptMized, Unleashing Zero-click Worms that Target GenAI-Powered Applications, consultado el 12 de marzo de 2024.
11. arXiv, Agente Smith: A Single Image Can Jailbreak One Million Multimodal LLM Agents Exponentially Fast, 13 de febrero de 2024.

AI y las elecciones estadounidenses

El impacto de la IA en las elecciones estadounidenses es una preocupación creciente. La aparición de deefakes, por ejemplo, hace que sea mucho más fácil para los ciberdelincuentes difundir información errónea e influir en el público votante. 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 anticipada. Es probable que incidentes alarmantes como este sean sólo el comienzo de estrategias de desinformación impulsadas por la IA.

Es importante señalar que el uso de la IA en estos esquemas puede no limitarse a agentes 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 advirtieron que Rusia y China probablemente aprovecharán la IA como parte de sus intentos de influir en las elecciones estadounidenses.

Incluso fuera de la política, la circulación en las redes sociales de imágenes deepfake de famosos como Taylor Swift resalta cuán fácilmente se puede difundir contenido manipulado antes de que pueda moderarse efectivamente. Las empresas de IA están tomando medidas para ayudar a mitigar este riesgo; Google Gemini, por ejemplo, ha promulgado barreras que impiden a los usuarios preguntar sobre las próximas elecciones en cualquier país. A medida que la AI 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.

Todos los ojos puestos en la regulación de la 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 el impacto de la IA, estimular la innovación y dar forma a su desarrollo responsable a través de políticas. Las regulaciones de IA seguirán desarrollándose y evolucionando rápidamente, pero algunos cambios regulatorios recientes pueden proporcionar una imagen útil para las empresas que buscan comprender estas tendencias.

Estados Unidos

En Estados Unidos, 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 inteligencia artificial a informar de los resultados de las pruebas de seguridad al Departamento de Comercio, así como a hacer público cuando se utilizan grandes recursos informáticos nuevos para entrenar modelos de IA. Además, exigió que nueve agencias federales completaran evaluaciones de riesgos sobre el impacto de la IA en la sinfraestructuras críticas. La Casa Blanca también se centra en la innovación en IA: como parte de la EO, el gobierno de EE. UU. estableció el programa piloto National Artificial Intelligence Research Resource (NAIRR) para conectar a los investigadores estadounidenses con potencia computacional, datos y otras herramientas para desarrollar IA.[17]

Queda por ver si el gobierno estadounidense buscará regulaciones más vinculantes en relación con la IA. Hasta ahora, al menos 15 empresas líderes en inteligencia artificial y casi 30 empresas de asistencia sanitaria han firmado compromisos voluntarios de la Casa Blanca para salvaguardar la inteligencia artificial.[18] Mientras tanto, la FTC ha prohibido el uso de IA para hacerse pasar por una agencia o empresa gubernamental, con planes de ampliar la regla para incluir protecciones para individuos y agencias privadas.[19] Presuntamente, la Casa Blanca también está explorando la posibilidad de exigir marcas de agua para el contenido generado por 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 muchos sectores. 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 sea generado por inteligencia artificial, que los modelos de capacitación sean 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.

Las regulaciones aplicarán 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 a la IA que afecte áreas particulares como infraestructura crítica, empleo y asuntos legales, inmigración y más. Mientras tanto, la UE prohibirá por completo las aplicaciones de IA consideradas inaceptablemente peligrosas, incluidas aquellas que utilizan información biométrica confidencial, que buscan manipular el comportamiento humano para eludir el libre albedrío, que utilizan el reconocimiento emocional para la contratación y la educación, o que extraen imágenes faciales no dirigidas de Internet o CCTV.[20]

Muchos países también están dando prioridad a las inversiones en IA. Singapur, por ejemplo, ha anunciado un plan de inversión de 740 millones de dólares estadounidenses en IA como parte de la Estrategia Nacional de IA 2.0 del país.[21] Este plan buscará impulsar la innovación en IA, permitiendo el acceso a chips avanzados necesarios para la IA y, al mismo tiempo, garantizando 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 IA de Singapur reciben un impulso con$740 Plan de inversión de millones, 19 de febrero de 2024.

Predicciones de amenazas de IA

La desinformación y los ciberataques generados por IA representan el segundo y el quinto de los 10 principales riesgos globales en 2024, según el Informe de riesgo global económico mundial.[22]

A medida que el campo de la IA siga evolucionando rápidamente, incluido el área de vídeos e imágenes generados por IA, estos riesgos seguirán aumentando, al igual que nuestra capacidad de aprovechar la IA para mitigarlos. De cara al resto de 2024 y en el futuro, estas son las principales predicciones de riesgos y amenazas de la IA que vemos en el horizonte.

1 El dilema de la IA de los estados-nación: impulsar las amenazas de la IA y al mismo tiempo bloquear el acceso a la IA

Los grupos de amenazas patrocinados por el estado 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 las previsiones de su trayectoria apuntan a un crecimiento significativo tanto en escala como en grado de sofisticación.
Los informes de Microsoft y OpenAI validan esta preocupación y revelan que grupos de autores de amenazas apoyados por países como Rusia, China, Corea del Norte e Irán han explorado y explotado activamente la funcionalidad ChatGPT. Esto se extiende a varios casos de uso, incluido el phishing, la generación y revisión de código y la traducción.,

Aunque la intervención dirigida ha detenido algunos de estos ataques, las empresas deberían prepararse para la persistencia de iniciativas de IA respaldadas por el Estado. Estas abarcan el despliegue 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 panorama complicado en el que los ciberdelincuentes patrocinados por el Estado continúan aprovechando la IA de formas novedosas para crear nuevas ciberamenazas complejas.

2 Chatbots oscuros y ataques impulsados por IA: la lacra de la "IA para el mal" irá en aumento

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 cibercriminales.

Estas perniciosas herramientas serán fundamentales para ejecutar ingeniería social mejorada, estafas de phishing y varias otras amenazas. La web oscura ha experimentado un aumento en los debates entre los ciberdelincuentes que profundizan en el despliegue ilícito 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 pequeña parte 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 eficiencia, los autores de amenazas emplean estas herramientas para descubrir y explotar vulnerabilidades, fabricar esquemas de phishing convincentes, ejecutar campañas de vishing y smishing, y automatizar ataques con mayor velocidad, sofisticación y escala. Por ejemplo, el grupo de autores de amenazas Scattered Spider utilizó recientemente LLaMa 2 LLM de Meta para explotar la funcionalidad de Microsoft PowerShell, permitiendo la descarga no autorizada de credenciales de usuario.[23] La trayectoria de estos avances indica que las ciberamenazas comenzarán a evolucionar más rápidamente que nunca, adoptando nuevas formas que serán más difíciles de reconocer o defenderse con medidas de seguridad tradicionales.

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 Meta's Llama 2 AI, 21 de febrero de 2024.

3 Luchar contra la IA con IA: las hojas de ruta y el gasto en seguridad incluirán defensas impulsadas por IA

Las empresas adoptarán cada vez más tecnologías de IA para combatir los ciberataques impulsados por IA, incluido un enfoque en el uso del aprendizaje profundo y modelos de IA/ML para detectar malware y ransomware ocultos en tráfico cifrado. Los métodos de detección tradicionales seguirán siendo insuficientes contra los nuevos ataques de día cero impulsados por IA y el ransomware polimórfico (que puede evolucionar su código para evadir la detección), por lo que los indicadores basados en IA serán cruciales para identificar amenazas potenciales. 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.

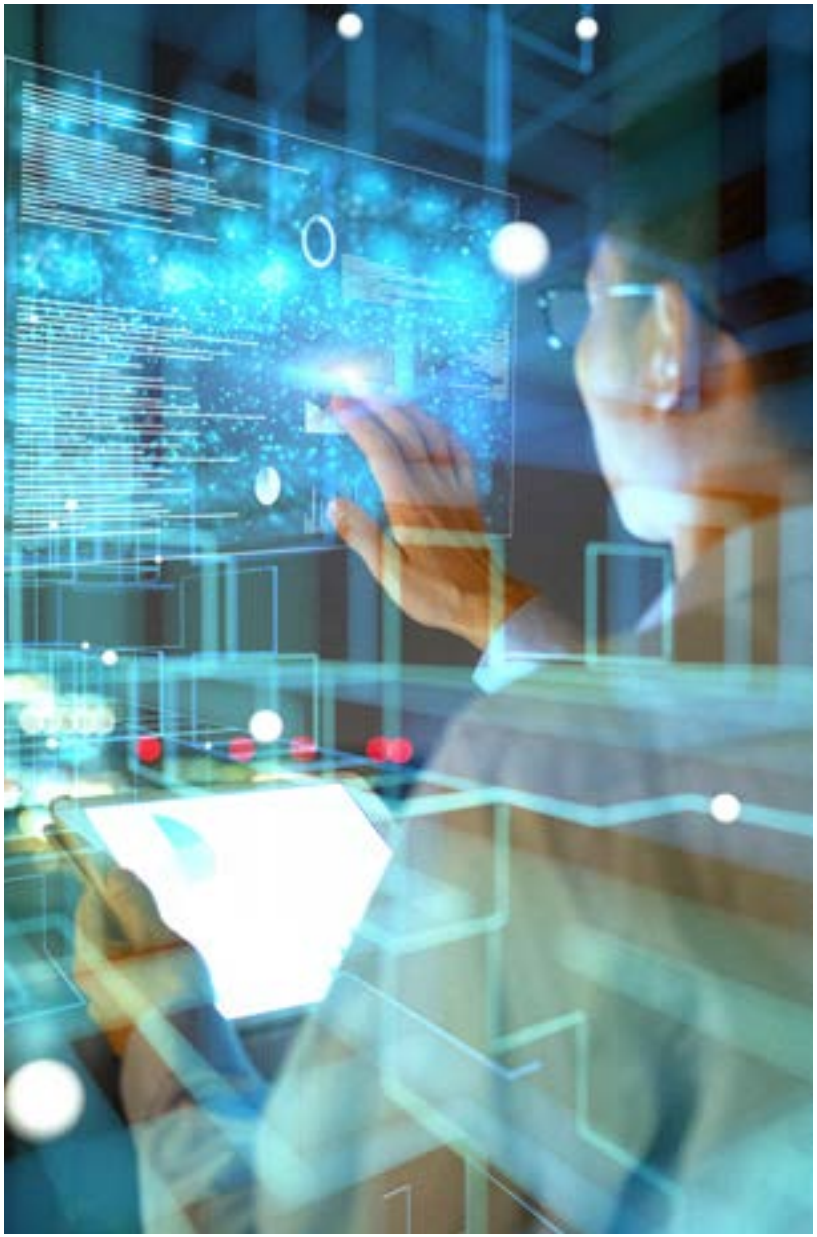
Las empresas incorporarán cada vez más la IA en sus estrategias de ciberseguridad. La IA se considerará un medio fundamental para ganar visibilidad del riesgo cibernético, así como para crear guías prácticas y cuantificables a fin de priorizar y remediar las vulnerabilidades de seguridad. Traducir el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

4 Intoxicación de datos en las cadenas de suministro de IA: aumentará el riesgo de que los datos de IA sean basura

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 formación y proveedores intermedios, serán cada vez más el objetivo de ciberataques.

El OWASP Top 10 para aplicaciones LLM destaca el envenenamiento de datos de capacitación y los ataques a la cadena de suministro como riesgos importantes, 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 socios tecnológicos, conjuntos de datos de terceros, y complementos o API de herramientas de IA) están listas para explotarse.

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 capacitación, particularmente en el ámbito de la ciberseguridad de la IA.



5 Dar o no dar rienda suelta: 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 se habrán planteado cuidadosamente sus políticas de seguridad de IA. Aun así, esta es una situación fluida para la mayoría de las empresas, y las preguntas sobre qué herramientas de IA permitirán, cuáles bloquearán y cómo protegerán sus datos siguen abiertas.

A medida que la cantidad de herramientas de IA continúa disparándose, las empresas deberán prestar mucha atención a las preocupaciones de seguridad de cada una; como mínimo, hacer un análisis profundo del uso de la IA de sus empleados, con la capacidad de 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 IA: los deepfakes virales impulsarán la interferencia electoral y las campañas de desinformación

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

Además, el uso de la IA en tales planes puede no limitarse a los autores nacionales. Las entidades patrocinadas por el estado también podrían explotar estas tácticas para sembrar confusión y socavar la confianza en el proceso electoral. Como ejemplo de un caso destacable, los atacantes utilizaron deepfakes generados por IA para engañar a un empleado para que se transfiriera 25 millones de dólares estadounidenses, lo que demuestra el impacto en el mundo real de esta tecnología. De manera similar, imágenes ilícitas de famosos como Taylor Swift se han vuelto virales en las redes sociales, llamando la atención sobre la facilidad con la que el contenido manipulado puede difundirse antes de que las medidas de moderación de contenido se pongan al día.

Estudio de caso: Habilite ChatGPT de forma segura en la empresa

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

A estas alturas, las empresas han tenido mucha exposición a las herramientas de inteligencia artificial. Pero a medida que la cantidad de aplicaciones de IA continúa creciendo drásticamente y la adopción continúa a buen ritmo, las empresas pueden adoptar ciertas mejores prácticas para mantener seguros sus datos, empleados y clientes. En general, las empresas deben adaptar de forma proactiva y continua su uso de la IA y sus estrategias de seguridad para adelantarse a los riesgos en evolución y, al mismo tiempo, aprovechar el potencial transformador de la IA.

CASO PRÁCTICO

Cinco pasos para integrar y proteger herramientas de IAgenerativa

Las empresas que deseen adoptar aplicaciones de IA de forma segura deberían adoptar un enfoque mesurado. En términos generales, primero pueden bloquear todas las aplicaciones de IA para eliminar el riesgo de filtración 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. En aras de la simplicidad, el siguiente viaje se centra en LLM 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 proactivo de confianza cero, 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

A continuación, la organización debe identificar un conjunto de aplicaciones de IA generativa que superen unos estándares elevados para ciertos criterios, como la capacidad de crear sólidas medidas contractuales, de seguridad y de protección de datos 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 de servidor ChatGPT privada en el entorno corporativo/DC

Para garantizar un control total sobre sus datos, las organizaciones deben alojar ChatGPT en un inquilino seguro y dedicado (como un servidor privado de Microsoft Azure AI) alojado completamente dentro de la organización. Posteriormente, 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 formación, lo que permite respuestas precisas y muy relevantes para los usuarios empresariales y, al mismo tiempo, minimiza el riesgo de envenenamiento de datos procedente de un lago de datos público.

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

A continuación, la organización debe colocar el ChatGPT trás una arquitectura de proxy en la nube de confianza cero, como Zscaler Zero Trust Exchange, para aplicar controles de seguridad de confianza cero sobre el acceso a ChatGPT. Esto también podría incluir mover 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. También garantiza una separación de inquietudes entre las consultas de los usuarios en esos mismos niveles de usuario, equipo y departamento.

Colocar ChatGPT detrás de un proxy en la nube como Zero Trust Exchange permite aún más a la organización inspeccionar todo el tráfico TLS/SSL entre los usuarios y ChatGPT a fin de detectar cibera–menazas y filtraciones de datos mientras se aplican siete capas distintas de seguridad de confianza cero.

Paso 5: Aplicar el motor Zscaler DLP para evitar filtracionesde datos

Finalmente, la organización debe implementar un motor DLP para la instancia de ChatGPT para evitar la filtración accidental de información crítica, incluidos códigos y datos propietarios, datos de clientes, datos personales, datos financieros y legales, y más. Esto garantiza que los datos altamente confidenciales nunca salgan del entorno de producción.

Al seguir este viaje, 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 su actividad empresarial.

- **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 política de IA

Las empresas deben respaldar estas mejores prácticas y establecer un marco de políticas claro que rija el uso aceptable, la integración y el desarrollo de productos, las políticas de seguridad y datos en toda la empresa, y las mejores prácticas de los empleados al utilizar herramientas de inteligencia artificial. 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 pertinente.
- **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 herramientas de IA deben seguir un marco de ciclo de vida del producto seguro** para garantizar el más alto nivel de seguridad.
- **Realice la debida diligencia 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 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, aprovechamos 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 promovida por 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 de infracción. Sin embargo, históricamente los problemas creados por la relación señal–ruido han convertido en un desafío aislar estas señales rápidamente. Utilizando IA generativa, Zscaler puede aprovechar estos datos para mejorar eficazmente las medidas de clasificación y protección al comprender las vulnerabilidades y debilidades que los atacantes probablemente explotarán. Esto no sólo permite a Zscaler predecir las infracciones antes de que se produzcan, sino que también brinda a los ejecutivos una forma global de visualizar y cuantificar la madurez y el riesgo cibernético mientras priorizan los pasos de corrección de la ciberseguridad con Zscaler Risk360.

Las capacidades de IA generativa no sólo se extienden al metanálisis del riesgo cibernético empresarial, sino que también se insertan directamente en los productos de ciberseguridad para detectar y detener 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 inteligencia artificial de Zscaler aprovechan un lago de datos que registra más de 390 mil millones de transacciones diarias, con más de 9 millones de amenazas bloqueadas y 300 billones de señales. Lejos del concepto de “basura que entra, basura que sale”, se trata de “datos e inteligencia de amenazas a gran escala y de alta fidelidad que entran, y una ciberseguridad de IA hiperconsciente y finamente ajustada que sale”. Todo esto se traduce en resultados de ciberseguridad más potentes y eficaces para los profesionales de TI y seguridad.

Aprovechar la IA en toda la cadena de ataque

Hemos analizado numerosas formas en que los autores de amenazas 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 productos cibernéticos para identificar y detener ataques convencionales y impulsados por IA en cada etapa de la cadena de ataque.

Escenario 1: Descubrimiento de superficie de ataque

La primera etapa de un ciberataque generalmente implica que los autores de amenazas exploren la superficie de ataque empresarial conectada a Internet para identificar debilidades explotables. A menudo, esto incluye aspectos como vulnerabilidades y configuraciones incorrectas de VPN, o cortafuego o servidores sin parches. La IA generativa ha hecho que esta tarea que en el pasado era ardua sea significativamente más fácil para los atacantes, quienes pueden simplemente consultar una lista de vulnerabilidades conocidas asociadas con estos activos.

AI aprovechar los conocimientos basados en IA en Zscaler Risk360, las empresas pueden ver instantáneamente estas aplicaciones y activos detectables (y por lo tanto arriesgados) (su superficie de ataque conectada a Internet) y ocultarlos de la Internet pública detrás de 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.

Escenario 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 Zscaler AI ayudan a reducir el riesgo de peligro, desbaratando ataques sofisticados y priorizando la productividad.

PREVENCIÓN DE LA SUPLANTACIÓN DE IDENTIDAD BASADA EN IA Y C2

Los modelos de IA de Zscaler detectan sitios de phishing conocidos y de paciente cero para evitar el robo de credenciales y la explotación del navegador, además de analizar patrones de tráfico, comportamiento y malware para detectar en tiempo real infraestructura de comando y control (C2) nunca antes vista. 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 DE SANDBOX DE IA BASADA EN ARCHIVOS

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

IA PARA BLOQUEAR AMENAZAS WEB

Zscaler Browser Isolation, impulsado por IA, bloquea las amenazas de día cero y al mismo tiempo garantiza que los empleados puedan acceder a los sitios pertinentes para realizar su trabajo. En la práctica, el filtrado de URL empresarial a menudo requiere controles más granulares que permitir/bloquear; los sitios bloqueados suelen ser seguros y necesarios para el trabajo, lo que genera tickets innecesarios para el servicio de asistencia técnica. Nuestro AI Smart Isolation puede identificar cuándo un sitio puede ser arriesgado y abrirlo de forma aislada para el usuario, transmitiendo el sitio de forma segura como píxeles en un entorno seguro y en contenedores. Esto detiene eficazmente las amenazas basadas en la web, como malware, ransomware, phishing y descargas no autorizadas, creando una postura de seguridad web sólida sin necesidad de que las empresas bloqueen demasiado los sitios de forma predeterminada.

Escenario 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, los usuarios tienen acceso a una cantidad excesiva de docenas de aplicaciones esenciales, lo que significa que su superficie de ataque interna es importante.

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 habitual ver que sólo 200 usuarios de entre 30 000 con acceso a una aplicación financiera realmente la necesitan. Zscaler puede crear automáticamente un segmento de aplicación que limita el acceso sólo a esos 200 empleados, reduciendo las oportunidades de movimiento lateral de los autores de amenazas en más del 99 %.

Escenario 4: Exfiltración de datos

En la etapa final de un ataque, los autores de amenazas trabajan para exfiltrar datos confidenciales. Zscaler utiliza la IA para permitir a las organizaciones implementar protecciones de datos más rápidamente. El descubrimiento de datos impulsado por IA elimina la laboriosa tarea de tomar huellas digitales y clasificar los datos, que de otro modo podría retrasar o impedir la implementación. Zscaler AI descubre y clasifica automáticamente todos los datos de una organización desde el primer momento, lo que permite a las empresas clasificar inmediatamente información confidencial mientras configuran políticas de prevención de pérdida de datos (DLP) para evitar que esos datos abandonen la organización como consecuencia de un ataque o infracción.

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:

- **Phishing impulsado por IA y detección de C2** contra sitios de phishing e infraestructura C2 nunca antes vistos, utilizando detección en línea basada en IA de Zscaler Secure Web Gateway (SWG).
- Sandboxing impulsado por IA con malware integral y prevención de amenazas de día cero.
- **Política dinámica basada en riesgos** con análisis continuo del riesgo de usuario, dispositivo, aplicación y contenido para impulsar una política dinámica de seguridad y acceso.
- **Segmentación impulsada por 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, presentando el contenido como un flujo de imágenes perfectas para eliminar las filtraciones de datos y la entrega de amenazas activas.

ADEMÁS, ZSCALER BLOQUEA:

Las URL e IPobservadas en la nube de Zscaler y procedentes de fuentes de información de amenazas comerciales y de código abierto integradas de forma 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 riesgo cibernético en toda la empresa.

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

Advanced Threat Protection bloquea todos los dominios C2 conocidos.

Zscaler ITDR (Detección y respuesta a amenazas de identidad) mitiga el riesgo de ataques basados en la identidad con visibilidad continua, supervisión del riesgo 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™ protege las aplicaciones limitando el movimiento lateral con el acceso menos privilegiado, la segmentación de usuario a aplicación 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™ detecta y contiene a los atacantes que intentan moverse lateralmente o escalar privilegios atrayéndolos con servidores, aplicaciones, directorios y cuentas de usuario señuelo.

Habilitando la transición de la IA empresarial: el control está en sus manos

Zscaler proporciona una manera para que las empresas fomenten la innovación, la creatividad y la productividad con aplicaciones de IA y, al mismo tiempo, mantengan seguros a los usuarios y los datos entre los canales emergentes de filtración de datos. Esto permite a las empresas aprovechar el potencial transformador de la IA para acelerar sus negocios sin bloquear directamente las aplicaciones y dominios de la IA.

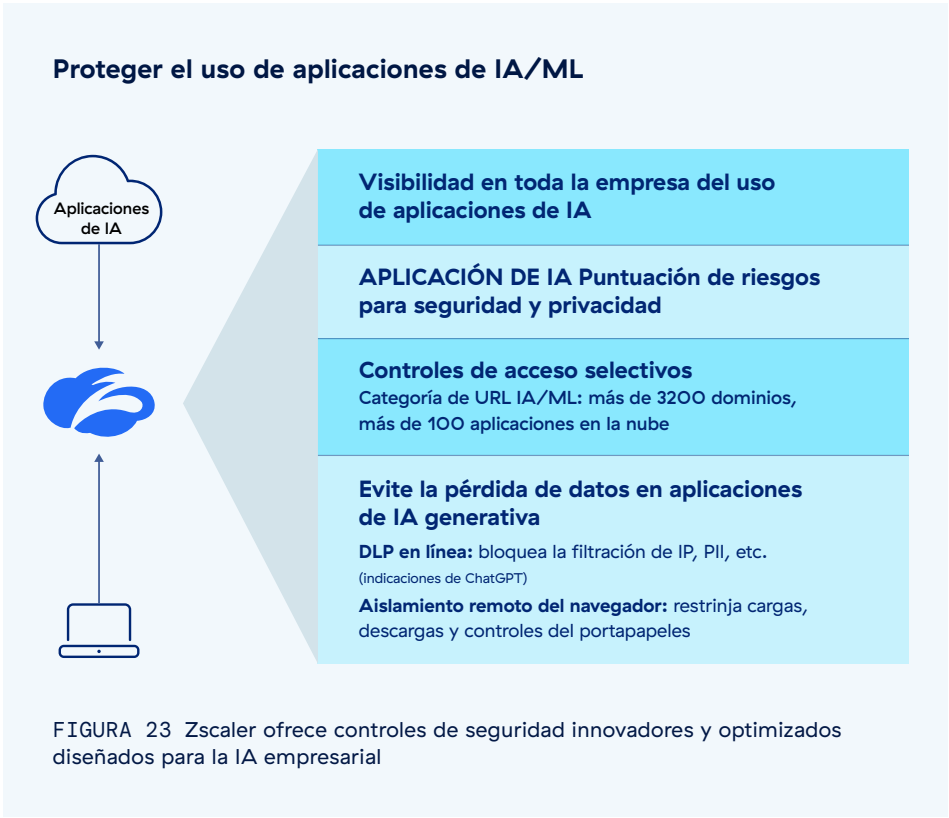


FIGURA 23 Zscaler ofrece controles de seguridad innovadores y optimizados diseñados para la IA empresarial

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.**
El potente filtrado de URL personalizado para aplicaciones de IA y ML permite a las empresas definir y aplicar fácilmente controles granulares de acceso y segmentación de la IA, bloqueando el acceso cuando sea necesario y permitiendo al mismo tiempo el acceso con niveles de riesgo aceptables mediante App Risk Scoring de IA. Las empresas pueden permitir el acceso a nivel de empresa, departamento, equipo y usuario, así como permitir un acceso basado en precauciones que oriente a los usuarios sobre los riesgos de las herramientas de IA generativa. La segmentación impulsada por IA facilita la identificación de segmentos de usuarios apropiados para acceder a aplicaciones de IA particulares y, al mismo tiempo, minimiza la superficie de ataque interna asociada con las herramientas de IA.

03 **Hacer cumplir la seguridad de datos granulares para ChatGPT y otras aplicaciones de IA.**
Las empresas pueden evitar la filtración de datos confidenciales cargados en aplicaciones de IA con controles granulares de la aplicación Zscaler Cloud para IA generativa. Al implementar el motor Zscaler DLP, las empresas pueden garantizar que no se comparta ningún dato accidentalmente al utilizar cualquier herramienta de inteligencia artificial. Mientras tanto, el descubrimiento y la clasificación de datos impulsados por IA permiten a las empresas identificar y crear fácilmente políticas de DLP en torno a sus datos más críticos, incluidos su código base corporativo, documentos financieros y legales, datos personales, datos de clientes y más. Este vídeo demuestra cómo el motor DLP evita que los usuarios introduzcan información de tarjetas de crédito en ChatGPT.

04 **Habilitar controles potentes usando Browser Isolation**
Zscaler Browser Isolation presenta aplicaciones de IA en un entorno seguro, agregando una capa de protección que permite al usuario realizar consultas y avisos a las herramientas de IA mientras restringe las funciones de copiar/pegar, cargas y descargas. Esto ayuda a mitigar el riesgo de que datos confidenciales se compartan accidentalmente con herramientas de inteligencia artificial 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 al mismo tiempo deben garantizar que sus datos sólo propicien la actividad empresarial, no las infracciones. Zscaler permite a las empresas navegar esta transición con confianza, aprovechando un conjunto completo de controles de seguridad de confianza cero impulsados por IA que protegen contra ataques impulsados por IA al tiempo que ofrecen políticas de IA ajustadas y protecciones de datos necesarias para aprovechar todo el potencial de la IA generativa.

Apéndice

Metodología de investigación de ThreatLabz

La nube de seguridad global Zscaler procesa más de 300 billones de señales diarias y bloquea 9000 millones de amenazas e infracciones de políticas por día, con más de 250 000 actualizaciones de seguridad diarias. Análisis de 18,09 mil millones de transacciones de IA y ML desde abril de 2023 hasta enero de 2024 en la nube de Zscaler, Zero Trust Exchange.

Acerca de Zscaler ThreatLabz

ThreatLabZ es la división de investigación de seguridad de Zscaler. Este equipo de primera clase es responsable de buscar nuevas amenazas y garantizar que las miles de organizaciones que usan la plataforma global Zscaler estén siempre protegidas. Además de investigar el malware y de analizar los comportamientos, los miembros del equipo participan en la investigación y el desarrollo de nuevos módulos prototipo para la protección avanzada contra las amenazas en la plataforma Zscaler. Asimismo, realizan habitualmente auditorías de seguridad internas para garantizar que los productos y la infraestructura de Zscaler satisfacen los estándares de cumplimiento 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. Zscaler Zero Trust Exchange protege a miles de clientes de los ciberataques y la pérdida de datos mediante la conexión segura de los usuarios, dispositivos y aplicaciones ubicados en cualquier lugar. Distribuida en más de 150 centros de datos en todo el mundo, Zero Trust Exchange basada en SASE es la mayor plataforma de seguridad en línea en la nube del mundo. Si desea más información, visite **www.zscaler.es.**

+1 408.533.0288 Zscaler, Inc. (HQ) • 120 Holger Way • San Jose, CA 95134 zscaler.es