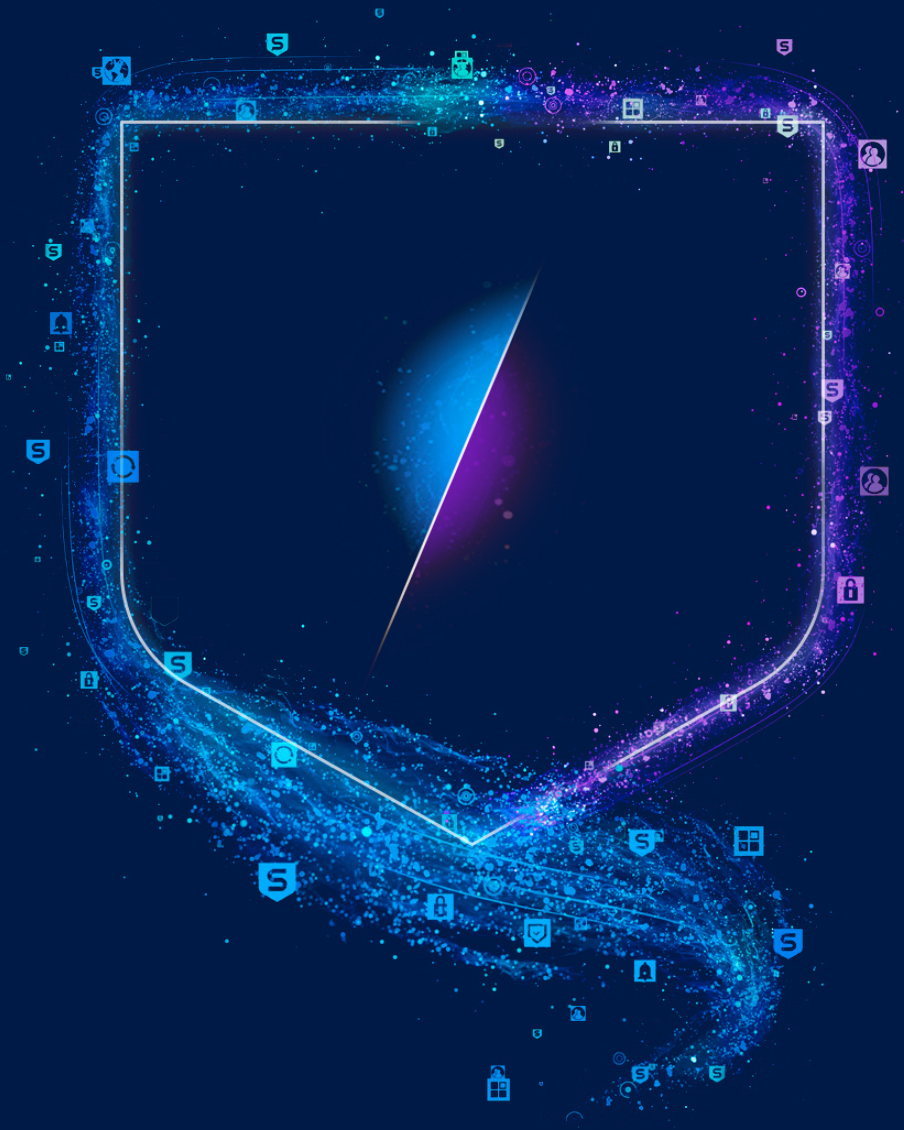
SOPHOS

Descifre el revuelo de la IA en ciberseguridad

Cómo sacar provecho de la IA con seguridad y confianza para reforzar las ciberdefensas de su organización



Contenido

Introducción

La IA ha desatado un enorme revuelo en el ámbito de la ciberseguridad. Las organizaciones se enfrentan a un aluvión de tentadoras promesas sobre la transformación de la ciberseguridad basada en la IA (más protección, costes reducidos, menor necesidad de especialistas) y advertencias alarmantes sobre una nueva era de ciberataques impulsada por esta tecnología.

Esta guía se ha diseñado para ayudar a las organizaciones a desentrañar el revuelo y las confusiones en torno a la IA en el sector de la ciberseguridad. En ella se explica qué puede hacer la IA (y qué no) para mejorar las ciberdefensas de las organizaciones y ahonda en los riesgos de ciberseguridad y operativos que ha introducido esta tecnología. En la guía también se dan recomendaciones para mitigar estos riesgos y sacar partido de la IA de manera segura y fiable con el fin de reforzar la protección frente a las amenazas e incrementar el retorno de la inversión.

A lo largo del documento también se incluyen valiosas conclusiones sobre la utilización, expectativas e inquietudes relativas a la IA, basadas en los resultados de una encuesta desvinculada de cualquier proveedor a 400 responsables de TI y ciberseguridad realizada a finales de 2024. Estas perspectivas desde la primera línea ofrecen un contexto muy valioso y sirven como punto de referencia para aquellas organizaciones que están evaluando su postura frente a la IA. Para leer todos los resultados, consulte Más allá del revuelo: La realidad empresarial de la IA en ciberseguridad.

En última instancia, ya sea con IA o sin ella, el objetivo sigue siendo el mismo: ofrecer de manera óptima el nivel de ciberresiliencia necesario para garantizar el éxito de su organización al tiempo que se minimizan los gastos totales. O, dicho de otra forma, aprovechar al máximo el (invariablemente limitado) presupuesto de ciberseguridad para impulsar el negocio. Esta guía le proporcionará las claves para lograrlo en la era de la IA.

La ventajas de la IA para la ciberseguridad

"IA" es un acrónimo breve que engloba diversas capacidades que pueden respaldar y agilizar la ciberseguridad de muchas maneras. La buena noticia es que la IA ofrece mayores ventajas incrementales a los defensores que a los adversarios. Existen dos enfoques comunes a la IA en el ámbito de la ciberseguridad: los modelos de Deep Learning y la IA generativa.

Deep Learning

Los modelos de Deep Learning (DL) APLICAN lo aprendido para ejecutar tareas. Estos pueden acelerar la aplicación del conocimiento mucho más allá de lo que los humanos son capaces de lograr. Por ejemplo, los modelos de DL correctamente entrenados pueden identificar si un archivo es malicioso o benigno en una fracción de segundo sin siquiera haber visto nunca antes ese archivo.

El DL es ideal para realizar tareas repetitivas a gran escala. Crea un modelo *estadístico* que analiza nuevos elementos bajo la distribución de todo lo que ha aprendido a partir de su extenso conjunto de datos de entrenamiento. Por ejemplo, los modelos de DL pueden evaluar millones de muestras de archivos con total seguridad para determinar si contienen malware. En consecuencia, el DL se emplea con frecuencia para potenciar las funciones de protección de los productos de ciberseguridad.

Los modelos de DL permiten a los defensores lidiar eficazmente con enormes volúmenes de amenazas generadas por adversarios que se sirven de la automatización y la ciberdelincuencia como servicio. Los modelos de DL también se pueden actualizar y adaptar a medida que evolucionan los ataques, lo que les permite mantenerse al día del panorama de amenazas.

IA generativa

Los modelos de IA generativa (GenAI) asimilan los datos introducidos y los utilizan para CREAR contenido nuevo. Algunos ejemplos de sus aplicaciones son:

- Crear un resumen en lenguaje natural de la actividad de amenazas hasta la fecha y los pasos recomendados que deben seguir los analistas
- Extraer datos clave sobre el comportamiento de los atacantes al analizar comandos que crean detecciones
- Permitir a los analistas utilizar búsquedas en lenguaje natural en lugar de consultas en código para investigar detecciones sospechosas
- Priorizar la aplicación de parches en función de la probabilidad de que se explote una vulnerabilidad

La IA generativa es una potente herramienta para acelerar las operaciones de seguridad. Al asumir una gran parte de la carga de trabajo con los datos, ayuda a los analistas a tomar decisiones inteligentes de forma rápida y les permite dedicar su tiempo a cuestiones de mayor impacto. De esta forma, la IA generativa puede aliviar en muchos casos la presión a la que se ven sometidos los analistas y, por tanto, disminuir el riesgo de agotamiento laboral y la rotación de empleados. La IA generativa también puede ayudar a reducir la barrera tecnológica en las operaciones de seguridad, con lo que los analistas menos experimentados pueden contribuir rápidamente de manera positiva y acelerar el desarrollo de sus habilidades.

El camino hacia la IA generativa

La base de la IA generativa moderna es el transformador, una red neuronal de Deep Learning que aprende el contexto y las relaciones entre los datos de entrada (por ejemplo, las palabras de una frase) y se sirve de este aprendizaje para generar resultados relevantes. Los transformadores suelen utilizarse en tareas de procesamiento del lenguaje natural (PLN), como traducir textos o responder preguntas. De hecho, la "T" de ChatGPT hace referencia a "transformador".

Los transformadores se utilizan mucho en el ámbito de la IA generativa, pero no todos ellos son generativos. Por ejemplo, BERT (las siglas en inglés de "representaciones codificadoras bidireccionales de transformadores") es un modelo de Machine Learning de código abierto para el PLN con la capacidad de leer el texto introducido de forma bidireccional (es decir, de izquierda a derecha y de derecha a izquierda). Esta capacidad le permite mejorar notablemente la comprensión contextual del texto no etiquetado. En Sophos llevamos muchos años utilizando BERT para identificar las estafas por correo electrónico corporativo comprometido y defendernos de ellas.

No existe una solución universal

El tamaño de los modelos de IA varía considerablemente de uno a otro. Los **grandes modelos**, como Microsoft Copilot y Google Gemini, son grandes modelos de lenguaje (LLM) entrenados con un conjunto de datos muy extenso que pueden ejecutar una amplia variedad de tareas. En cambio, los modelos pequeños suelen diseñarse y entrenarse con un conjunto de datos muy específico para realizar una única tarea, como detectar URL maliciosas o ejecutables. Aunque su alcance es más limitado, los **modelos pequeños** tienen ventajas en cuanto a coste, velocidad y rendimiento en comparación con los modelos más grandes.

Limitaciones de la IA

La IA por sí sola no es la respuesta, al menos no en un futuro cercano. La IA complementa la experiencia humana, pero no la sustituye por completo. Las amenazas son sumamente complejas, y llevar a cabo operaciones de seguridad efectivas requiere tanto destrezas técnicas como la capacidad de aplicar conocimientos en el contexto de la organización. La IA por sí sola no puede mantener a las organizaciones un paso por delante de los grupos de ciberdelincuentes cualificados y bien financiados de hoy día.

TIPO

IA con Deep Learning *Aplicar*

Utiliza redes neuronales artificiales para reconocer patrones y tomar decisiones de manera similar al cerebro humano. APLICA lo aprendido para ejecutar tareas.

Ejemplo: **Detectar URL maliciosas**

El modelo de IA se entrena para identificar sitios web maliciosos para que los productos de seguridad puedan bloquear el acceso a los mismos

IA generativa *Crear*

Utiliza la estructura y el patrón de los datos existentes para CREAR (generar) contenido totalmente nuevo.

Ejemplo: **Resumen de casos de amenazas**

El modelo de IA crea un resumen de la actividad de amenazas y da a los analistas los pasos de seguimiento recomendados

TAMAÑO

Grandes modelos de IA

Herramientas multifuncionales entrenadas con enormes cantidades de datos de acceso público, diseñadas para ofrecer soporte en una amplia variedad de tareas.

Ejemplo: **Microsoft Copilot, Google Gemini**

Modelos de IA pequeños

Los modelos centrados en los resultados están diseñados, entrenados y desarrollados para casos de uso específicos.

Ejemplo: **Modelo de detección de malware para Android**

Tasas de adopción de la IA

La IA ya está ampliamente integrada en la infraestructura de ciberseguridad de la mayoría de las organizaciones:

- El 73 % afirma que sus soluciones de ciberseguridad incluyen modelos de Deep Learning
- El 65 % afirma que sus soluciones de ciberseguridad incluyen funciones de IA generativa

Las aplicaciones de la IA en ciberseguridad no se limitan a proveedores externos. El 34 % de las organizaciones ya utilizan la IA generativa internamente para mejorar su ciberseguridad, por ejemplo, para generar correos electrónicos de prueba de phishing.

Es probable que la adopción de la IA se vuelva prácticamente universal en poco tiempo, dado que el 99 % (redondeando) de las organizaciones ya incluyen las funciones de IA como requisito al seleccionar una plataforma de ciberseguridad:

- El 57 % afirma que las funciones de IA son esenciales o sumamente importantes
- El 41 % afirma que las funciones de IA son importantes

En vista de estas tasas de adopción de la IA y su futuro uso, comprender los riesgos y las mitigaciones asociadas a la IA en el ámbito de la ciberseguridad es una prioridad para todas las organizaciones, independientemente de su tamaño o sector.

73 %
Usa herramientas de ciberseguridad con modelos de Deep Learning

65 %
Usa herramientas de ciberseguridad con funciones de IA generativa

99 %
Requiere funciones de IA al seleccionar una plataforma de ciberseguridad

IA generativa: grandes expectativas

El revuelvo provocado por la IA generativa ha creado grandes expectativas sobre cómo esta tecnología puede optimizar los resultados en ciberseguridad. En la tabla siguiente detallamos el principal beneficio que, según la encuesta, esperan obtener las organizaciones de las funciones de IA generativa de las herramientas de ciberseguridad.

Principal beneficio deseado de la IA generativa Respuestas clasificadas en primer lugar	
1=	Mejor protección frente a las ciberamenazas (20 %)
1=	Mejor retorno de la inversión en ciberseguridad (ROI) (20 %)
3	Mayor eficiencia e impacto de los analistas de TI (17 %)
4	Confianza en que estamos al día de las innovaciones en ciberseguridad (15 %)
5=	Mayor tranquilidad al saber que nuestra organización está adecuadamente protegida contra los ataques (14 %)
5=	Menor agotamiento de los empleados (p. ej., automatizar tareas para liberar tiempo a los empleados de ciberseguridad (14 %)

¿Qué beneficios desea obtener de las funciones de IA generativa de las herramientas de ciberseguridad, si los hubiera? Respuestas clasificadas en primer lugar (n=400)

La variedad de respuestas indica que no hay un único beneficio claro que se espere de la IA generativa en ciberseguridad. Al mismo tiempo, las ventajas más deseadas suelen estar relacionadas con la mejora de la ciberprotección o el rendimiento empresarial (tanto financiero como operativo). Los datos también sugieren que la incorporación de funciones de IA generativa en las soluciones de ciberseguridad da a las organizaciones tranquilidad y confianza en que están al día de las capacidades de protección más recientes.

El hecho de que la reducción del agotamiento laboral se encuentre en el último lugar de la clasificación apunta a que las organizaciones prestan menos atención o muestran menos interés en el potencial de la IA generativa para apoyar a los usuarios. Teniendo en cuenta la escasez de personal especializado en ciberseguridad, la reducción de la rotación de empleados es un área importante en la que centrarse y en la que la IA puede resultar de ayuda.

La mejora de la **protección** y el incremento del **ROI** son los principales beneficios que las organizaciones buscan obtener de la IA generativa

Los riesgos de la IA para la ciberseguridad

El uso de la IA en ciberseguridad es un arma de doble filo. Si bien ofrece enormes beneficios a los defensores en la lucha contra los adversarios, también conlleva una serie de riesgos:

1. **Riesgo de amenaza:** el uso de la IA en ciberataques
2. **Riesgo en la defensa:** IA de mala calidad o mal implementada
3. **Riesgo operativo:** dependencia excesiva de la IA
4. **Riesgo financiero:** bajo retorno de la inversión en IA
5. **Riesgo de secuestro:** manipulación de modelos de IA públicos por parte de los adversarios

1. Riesgo de amenaza: el uso de la IA en ciberataques

Aunque se ha generado mucho revuelo sobre cómo la IA está dando lugar a un nuevo panorama de amenazas, la realidad es más comedida. Las conversaciones sobre la IA en los foros de los ciberdelincuentes son pocas, y muchos actores maliciosos continúan siendo escépticos en cuanto a su uso. En los casos observados, los intentos de crear malware, herramientas de ataque y exploits mediante IA son generalmente rudimentarios y de poca calidad.

Al igual que las organizaciones legítimas, los adversarios están utilizando la IA principalmente para mejorar la calidad de su contenido y la eficiencia de sus operaciones, aunque con fines muy distintos. Para obtener más información sobre el panorama de amenazas más reciente o los ataques basados en IA, consulte el blog de Sophos.

Mejora de la calidad del contenido

Una de las aplicaciones más rápidas, sencillas y accesibles de la IA en los ciberataques es mejorar la calidad y la credibilidad de los correos electrónicos de phishing y de las estafas para que las posibles víctimas caigan más fácilmente en la trampa.

Las señales que suelen delatar el phishing, como los errores gramaticales, ortográficos y de formato, pueden eliminarse fácilmente con herramientas de IA. Con un LLM público, se puede crear un correo electrónico bien escrito para una campaña de phishing en menos de un minuto. De forma similar, ahora se puede acceder fácilmente a textos y mensajes bien redactados y convincentes para redes sociales, en cualquier idioma, diseñados para engañar a los destinatarios y conseguir que hagan clic o compartan información personal Los LLM también facilitan a los atacantes incorporar información actualizada en sus ataques, lo que aumenta aún más la probabilidad de que la víctima caiga en la estafa.

Las herramientas de IA generativa también han abierto la puerta a una nueva era de estafas que suplantan a altos cargos para engañar a las víctimas desprevenidas y convencerlas para que realicen transferencias financieras. La tecnología de clonación de voz ha avanzado hasta tal punto que, con suficiente entrenamiento, los atacantes pueden hacer creer a su interlocutor que está hablando con la persona real. En estos ataques de phishing por voz, o "vishing", el ciberdelincuente suele suplantar a un alto directivo y llamar a un miembro del personal para "pedirle" que realice una compra ilícita de tarjetas regalo, un pago bancario o una transferencia de archivos.

Los adversarios también están utilizando la tecnología deepfake con IA para suplantar visualmente a las personas en sus ataques. Los vídeos manipulados con deepfake se han utilizado para engañar a empleados incautos y persuadirlos para que realizaran pagos importantes, y también para burlar sistemas de reconocimiento facial en solicitudes de préstamos y registros de cuentas bancarias.

Mejora de la eficiencia operativa

Al igual que muchas empresas utilizan chatbots basados en IA para mejorar la experiencia de sus usuarios, los atacantes también los emplean. Algunos ciberdelincuentes utilizan los LLM para mejorar los foros que frecuentan mediante la creación de chatbots y respuestas automáticas. En un ejemplo compartido por Sophos X-Ops, el foro XSS creó un chatbot específico para responder a las consultas de los usuarios. El administrador anunció lo siguiente (traducción aproximada del ruso):

> *"En esta sección, puedes chatear con la IA (inteligencia artificial). Haz una pregunta y nuestro bot de IA te responderá. Esta sección y el bot de IA se han diseñado para solucionar problemas técnicos sencillos, para entretener a nuestros usuarios y para que estos se familiaricen con las posibilidades que ofrece la IA".*

Crear y entrenar modelos personalizados requiere mucha experiencia y conocimientos sobre IA, lo que cuesta dinero y no abunda precisamente. Si bien muchas bandas de ciberdelincuentes cuentan con sus propios expertos en IA, suelen aprovechar los LLM existentes en sus ataques en lugar de crearlos.

Contextualización de los atacantes

Es importante poner en contexto el uso que hacen de la IA los adversarios. La IA es tan solo una de las muchas herramientas que tienen los atacantes en su arsenal. Los agentes maliciosos llevan varios años utilizando la automatización y los modelos de ciberdelincuencia como servicio para ampliar el alcance y la frecuencia de sus ataques. Para muchas organizaciones, estas capacidades tendrán una mayor repercusión en la exposición al riesgo que la IA.

2. Riesgo en la defensa: IA de mala calidad o mal implementada

Como hemos visto, los modelos de IA ya están ampliamente integrados en las ciberdefensas de las organizaciones. Aunque no cabe duda de que las intenciones son buenas, los modelos de IA de baja calidad y mal implementados pueden introducir riesgos importantes para la ciberseguridad. La tendencia de los modelos de IA a provocar riesgos depende de diversos factores, entre ellos:

- **Calidad de los datos con los que se entrenan los modelos**. En el contexto de la IA, es innegable que se cosecha lo que se siembra. Utilizar datos de baja calidad para entrenar modelos implica el riesgo de introducir errores, mientras que el uso de conjuntos de datos desequilibrados puede distorsionar los resultados debido a la sobrerrepresentación o subrepresentación de ciertas variables. Cuanto mayor sea la cantidad de datos de alta calidad para el entrenamiento, mejor será el resultado obtenido.
- **Experiencia de los equipos que crean los modelos**. Para crear modelos de IA efectivos para la ciberseguridad, se requiere un conocimiento profundo de dos áreas distintas pero complementarias:
  - **Amenazas:** para determinar qué queremos que haga el modelo de IA, lo primero que debemos entender es cómo funcionan el malware y los adversarios.
  - **IA:** cuando ya sabemos qué queremos que haga la IA, debemos identificar y crear el modelo correcto para lograr el objetivo.

  Para generar modelos de IA eficaces que tengan un impacto tangible en la ciberseguridad, es fundamental integrar estas dos áreas de conocimiento de manera que se complementen mutuamente.
- **Calidad del proceso de desarrollo e implementación del producto**. A mediados de 2024, la distribución de una actualización de contenido con errores en un producto de ciberseguridad provocó una interrupción inmediata en empresas de todo el mundo. Si las funciones de IA no se prueban, evalúan e implementan correctamente, pueden generar un daño incluso mayor, con el riesgo añadido de que el problema no pueda identificarse o rectificarse fácilmente.

Falsa sensación de (ciber)seguridad

Por lo general, las organizaciones son conscientes del riesgo que implica una IA mal desarrollada e implementada en las soluciones de ciberseguridad. La gran mayoría (89 %) de los profesionales de TI/ciberseguridad encuestados afirman estar preocupados por la posibilidad de que existan fallos en las funciones de IA generativa de las herramientas de ciberseguridad que puedan perjudicar a su organización; de estos, el 43 % aseguran estar extremadamente preocupados y el 46 %, algo preocupados.

Por tanto, no es de extrañar que el 99 % (redondeando) de las organizaciones afirmen que, al evaluar las funciones de IA generativa de las soluciones de ciberseguridad, tienen en cuenta la calidad de los procesos y los controles de ciberseguridad utilizados en su desarrollo:

- El 73 % afirma que evalúa exhaustivamente la calidad de los procesos y los controles de ciberseguridad
- El 27 % afirma que evalúa parcialmente la calidad de los procesos y los controles de ciberseguridad

Aunque el elevado porcentaje que afirma realizar una evaluación completa pueda parecer alentador de entrada, en realidad pone de manifiesto que muchas organizaciones presentan una significativa carencia en este ámbito.

Evaluar los procesos y controles empleados para desarrollar funciones de IA generativa exige transparencia por parte del proveedor y un nivel razonable de conocimientos de IA por parte del evaluador. Desafortunadamente, ambos escasean. Los proveedores de soluciones rara vez permiten acceder con facilidad a todos los detalles de sus procesos de desarrollo e implementación de IA generativa, y los equipos de TI suelen tener un conocimiento limitado de las prácticas recomendadas en desarrollo de IA. En el caso de muchas organizaciones, esta conclusión sugiere que "no saben lo que no saben".

3. Riesgo operativo: dependencia excesiva de la IA

La inteligencia artificial influye en casi todos los aspectos de nuestra vida cotidiana, desde buscar la mejor ruta al supermercado hasta recibir recomendaciones de programas de televisión. Su naturaleza omnipresente hace que sea fácil recurrir automáticamente a la IA y dar por sentado que esta puede realizar ciertas tareas mejor que las personas. Afortunadamente, la mayoría de las organizaciones son conscientes de las consecuencias de una dependencia excesiva de la IA y se preocupan por ellas.

- El 84 % está preocupado por la presión resultante para reducir la plantilla de profesionales de ciberseguridad.
- El 87 % está preocupado por la falta de responsabilidad en ciberseguridad que podría derivarse de ello.

Permanecer alerta ante estos riesgos es el primer paso hacia su mitigación. Es importante recordar que la IA es solo una de las muchas herramientas de defensa que utilizan las organizaciones y, si bien constituye una valiosa parte de su pila de seguridad, no siempre ofrece la mejor estrategia y rara vez es la única solución necesaria. Cada organización es distinta, y el uso de la IA debe adaptarse a la configuración específica y a las necesidades más amplias del negocio.

4. Riesgo financiero: bajo retorno de la inversión en IA

Las funciones avanzadas de IA generativa de las soluciones de ciberseguridad requieren una inversión considerable tanto en desarrollo como en mantenimiento. Los responsables de TI y ciberseguridad son conscientes del impacto de este gasto, y un 80 % cree que la IA generativa incrementará considerablemente el coste de sus productos de ciberseguridad.

Pese a las expectativas de aumento de los precios, la mayoría de las organizaciones ven la IA generativa como una vía para disminuir su gasto general en ciberseguridad: el 87 % de los encuestados están seguros de que los costes asociados a la IA generativa en las herramientas de ciberseguridad quedarán completamente compensados por los ahorros que esta genere.

Al mismo tiempo, las organizaciones reconocen que cuantificar estos costes es un desafío. Los gastos asociados a la IA generativa suelen incluirse en el precio global de los productos y servicios de ciberseguridad, con lo que resulta difícil determinar cuánto están destinando las organizaciones a la IA generativa para la ciberseguridad. En línea con esta falta de visibilidad, el 75 % coincide en que estos costes son difíciles de medir (39 % totalmente de acuerdo, 36 % algo de acuerdo).

Sin unos informes claros al respecto, las organizaciones corren el riesgo de no obtener el retorno esperado de sus inversiones en IA para la ciberseguridad o, lo que es peor, de destinar inversiones a la IA que podrían dar más frutos en otros ámbitos.

5. Riesgo de secuestro: grandes modelos de lenguaje (LLM) comprometidos

Los riesgos de ciberseguridad de la IA van más allá de las herramientas y aplicaciones de ciberseguridad. La rápida expansión global del uso de LLM públicos abre la puerta a que atacantes sofisticados y bien financiados manipulen estos modelos para sus propios objetivos. Esto podría concretarse de varias formas, por ejemplo:

- **Contaminación de los datos**. En su artículo de 2023 Poisoning Web-Scale Training Datasets is Practical (La contaminación de los conjuntos de datos de entrenamiento a escala web es posible), Carlini *et. al.* demostraron que la contaminación de los datos (es decir, la manipulación de los datos con los que se entrena un modelo para influir en los resultados) es un riesgo de amenaza viable.
- **Puertas traseras de actores estatales**. Muchos estados nacionales cuentan con los recursos necesarios para crear poderosos LLM. Al incorporar puertas traseras secretas y luego poner los modelos a disposición pública de manera gratuita, los actores estatales podrían manipular el LLM a su favor en caso necesario.
- **Suplantación de LLM**. Los ciberdelincuentes pueden comprometer LLM legítimos (por ejemplo, añadiendo puertas traseras) y luego presentar esos cambios como "mejoras". Para engañar a los usuarios y conseguir que utilicen la herramienta manipulada, suplantan el nombre del proveedor de confianza, por ejemplo, omitiendo una letra o sustituyendo la letra O por el número 0.

Para profundizar en el tema de la manipulación de los LLM, consulte la investigación más reciente del equipo de Sophos AI.

Pasos concretos para afrontar el revuelo de la IA

Aunque la IA conlleva riesgos, con una estrategia cuidadosa, las organizaciones pueden gestionarlos adecuadamente y sacar partido a la IA de forma segura para mejorar sus ciberdefensas. Muchas de las siguientes recomendaciones también son útiles para implementar la IA correctamente en otro ámbitos.

Riesgo de amenaza: refuerce sus ciberdefensas para la era de la IA

Un aspecto clave sería el de incrementar la resiliencia frente a las amenazas impulsadas por IA. Puesto que los adversarios están utilizando la IA principalmente para mejorar la calidad y la credibilidad de los correos electrónicos de phishing y las estafas, lo más lógico es centrarse en estas cuestiones. He aquí algunas sugerencias:

- **Refuerce la protección del correo electrónico**. Busque soluciones que puedan detectar correos electrónicos de phishing y estafas generados con IA para impedir que lleguen a los buzones de sus usuarios.
- **Despliegue protección frente a estafas por correo electrónico corporativo comprometido y protección de usuarios VIP**. Opte por soluciones de seguridad del correo electrónico que incluyan protección frente a BEC y de usuarios VIP; por ejemplo, que permitan escanear el contenido según el tono y el estilo utilizados a fin de detectar estafas.
- **Tenga especial precaución con las redes sociales**. En general, los usuarios no están muy atentos cuando utilizan las redes sociales, lo que aumenta la posibilidad de caer en una estafa.
- **Implemente procesos para mitigar el riesgo de la clonación de voz**. Por ejemplo, procedimientos que deben seguirse si se recibe una solicitud inesperada de pago o de uso compartido de datos. Estos podrían ser:
  - Llamar al solicitante para verificar la petición
  - Implementar el uso de códigos de acceso o frases de seguridad

Riesgo en la defensa: evalúe la calidad de la IA utilizada en los productos de ciberseguridad

Esté alerta a los riesgos y al impacto de una IA de baja calidad en sus inversiones en seguridad. Pregunte a los proveedores sobre lo siguiente:

- **Datos de entrenamiento**. ¿Cuál es la calidad, la cantidad y la fuente de los datos con los que se entrenan los modelos? Cuanto mejores sean los datos introducidos, mejores serán los resultados.
- **Equipo de desarrollo**. Pregunte por las personas que trabajan en los modelos. ¿Qué experiencia en IA tienen? ¿Cuáles son sus conocimientos sobre amenazas, comportamientos de adversarios y operaciones de seguridad?
- **Procesos de ingeniería e implementación de productos**. ¿Qué pasos sigue el proveedor al desarrollar y desplegar funciones de IA en sus soluciones? ¿Qué medidas de control y supervisión se han establecido?

En última instancia, hágase esta pregunta: ¿hasta qué punto confío en que esta organización esté trabajando correctamente con la IA y haya implementado los controles rigurosos de calidad y de despliegue que se necesitan?

Riesgo operativo: aborde la IA priorizando la intervención humana

A la IA no le importará que sufra una filtración, pero a su personal, sí. Si se da el peor de los casos y su organización se ve expuesta, necesitará a un equipo de expertos capaces de entender y solucionar la situación en el contexto de su negocio.

- **Mantenga la perspectiva**. La IA es tan solo una herramienta más en el arsenal del defensor. Puede utilizarla, pero debe subrayar que la responsabilidad última de la ciberseguridad recae en las personas.
- **No reemplace: acelere**. La actual escasez global de profesionales especializados en ciberseguridad es ampliamente reconocida. Los graves problemas de agotamiento laboral empeoran el desafío. En lugar de utilizar la IA para reducir personal, céntrese primero en cómo puede ayudar la IA al que ya tiene. La IA puede gestionar numerosas tareas operativas de seguridad repetitivas de bajo nivel y proporcionar información guiada, por ejemplo:
  - Liberar tiempo para tareas de mayor valor que generen un impacto en el negocio
  - Reducir el exceso de alertas y ayudar a disminuir la fatiga
  - Acelerar el desarrollo profesional de los analistas cualificados
  - Permitir que los analistas con menos experiencia realicen operaciones de seguridad y construyan un flujo de recursos

Riesgo financiero: imponga el rigor empresarial en las decisiones de inversión en IA

Esta es una de las áreas más fáciles de mitigar para las organizaciones, ya que muchos factores están totalmente bajo su control.

- **Establezca objetivos**. Defina de manera clara, específica y detallada los resultados que desea obtener de la IA.
  - Identifique qué necesita. ¿Qué carencias tiene? ¿En qué puede ayudarle la IA?
  - Considere las ventajas en términos financieros, de tiempo y de protección.
- **Cuantifique los beneficios**. Entienda qué impacto tendrán las inversiones en IA.
  - Si el objetivo es reducir los costes generales y el TCO de la ciberseguridad, cuantifique los ahorros que generará como resultado.
  - Si lo que desea es reducir la rotación de empleados de TI y ciberseguridad, defina de forma clara cómo afectará la herramienta de IA al equipo. ¿Qué tareas les evitará? ¿Cuántas horas liberará?
- **Priorice inversiones**. La IA puede ayudar de muchas maneras, pero unas tendrán mayor repercusión que otras. Identifique las métricas más relevantes para su organización: ahorros financieros, efecto en la rotación de personal y reducción de la exposición, por ejemplo, y compare las distintas opciones.
- **Mida el impacto**. Las decisiones de inversión se toman con las mejores intenciones. Asegúrese de comprobar que el rendimiento obtenido se ajuste a las expectativas iniciales. ¿Está viendo las ventajas que esperaba? ¿Hay ganancias inesperadas? ¿Hay áreas en las que no esté obteniendo los resultados esperados? Utilice las respuestas a estas preguntas para realizar los cambios necesarios.

Pregúntese si la IA es la mejor forma de conseguir su objetivo, o si existen otras tecnologías o estrategias que hubieran podido generar un mayor impacto.

Riesgo de secuestro: esté alerta al peligro

Este riesgo es el más difícil de mitigar para las organizaciones. El simple hecho de estar alerta a este riesgo ayuda a reducir sus posibles repercusiones. Dicho esto, si opta por LLM públicos, fíjese en lo siguiente:

- **Modelos de proveedores reconocidos y con buena reputación**. Aunque no son inmunes a los ataques por contaminación de datos, es más probable que se publique y comparta cualquier problema que se produzca con los resultados de los datos.
- **Nombres de proveedor correctos**. Los atacantes suplantan los nombres de proveedores reconocidos para engañar a los usuarios y hacerles creer que sus modelos comprometidos son legítimos.

Los especialistas en IA para la ciberseguridad ya están trabajando en estrategias para neutralizar este riesgo.

Conclusión

La IA ofrece ventajas fenomenales en el sector de la ciberseguridad. Si las organizaciones no se dejan llevar por el revuelo generado por la IA y adoptan una estrategia bien pensada y centrada en los resultados, podrán aprovechar esta tecnología para reforzar sus ciberdefensas y potenciar a su valioso equipo de TI y ciberseguridad.

Acerca de la encuesta

Fuente: Más allá del revuelo: La realidad empresarial de la IA en ciberseguridad

Sophos encargó al especialista en investigación independiente Vanson Bourne una encuesta a 400 responsables de TI y ciberseguridad de organizaciones de entre 50 y 3000 empleados. La encuesta se llevó a cabo en noviembre de 2024 y los encuestados procedían de 13 sectores. Con el objetivo de garantizar una representación amplia de la industria, la encuesta se desvinculó de cualquier proveedor, y las organizaciones de los encuestados utilizaban soluciones de seguridad para endpoints de 19 proveedores diferentes.

Acerca de Sophos

Sophos es una empresa líder global en ciberseguridad con un completo catálogo de productos y servicios galardonados, que van desde protección para firewalls y endpoints y herramientas de EDR/XDR hasta servicios de detección y respuesta gestionadas (MDR) y de respuesta a incidentes (IR)

Sophos ha estado reforzando la ciberseguridad con inteligencia artificial desde 2017, combinando la IA con la experiencia humana para detener la más amplia variedad de ciberamenazas en cualquier entorno. Las capacidades de Deep Learning e IA generativa que resuelven los problemas más cruciales de los clientes están integradas en nuestros productos y servicios y se ofrecen a través de la plataforma de seguridad nativa de IA más amplia del sector. Gracias al entrenamiento con datos de ataques en más de 600 000 entornos de clientes distintos, nuestra plataforma adaptativa con IA ofrece una protección inigualable frente a las amenazas avanzadas y potencia el trabajo de los defensores.

Para obtener más información y explorar las soluciones de Sophos, visite es.sophos.com

2025-01-15 (WP-MP)

SOPHOS