IRON MOUNTAIN®

White paper

Cómo la IA generativa está redefiniendo la seguridad

Resumen

Cuando se trata de seguridad, ¿es GenAI una herramienta para el bien o para el mal? Los equipos de seguridad de las empresas están lidiando con lo que esta tecnología emergente representa para sus organizaciones. Cada vez se dan más cuenta de que la IA generativa conlleva tanto oportunidades como riesgos.

Contenido

GenAI: una nueva herramienta para todo

Indudablemente, una de las innovaciones tecnológicas recientes más importantes es la inteligencia artificial generativa (GenAI). Cuando se lanzó ChatGPT a finales de noviembre de 2022, marcó un antes y un después en una nueva era de la informática.

Puedes utilizar GenAI para escribir un artículo o para generar el código de tu última actualización de software. Es capaz como para generar imágenes que parecen fotos realistas o crear vídeos y archivos de audio casi idénticos a los reales. Además, permite buscar en Internet respuestas a casi cualquier pregunta. Y sólo hemos explorado la superficie de lo que esta nueva tecnología puede llegar a hacer.

Es comprensible que las organizaciones estén entusiasmadas con las oportunidades que representa GenAI. Una encuesta de AWS y el MIT reveló que el 80% de los responsables de datos creen que la IA generativa transformará sus organizaciones, y el 45% afirmó que sus empresas ya la han adoptado ampliamente.

«La IA generativa tiene el potencial de ser la tecnología más disruptiva que hemos visto hasta la fecha», afirma Steve Chase, líder de consultoría de KPMG en Estados Unidos. «Cambiará fundamentalmente los modelos de negocio, brindando nuevas oportunidades de crecimiento, eficiencia e innovación, a la vez que emergen riesgos y retos significativos. Para que los líderes aprovechen el enorme potencial de la IA generativa, deben establecer una estrategia clara que haga que su organización pase rápidamente de la fase de experimentación a la de industrialización».

Ante este escenario, los directivos sienten la presión de empezar a emplear GenAI lo antes posible. Sin embargo, es natural que algunos aún lo duden. A pesar de todas sus ventajas potenciales, la IA generativa también presenta algunos riesgos notables.

Muchos líderes de seguridad siguen esforzándose por diseñar un plan sobre cómo manejar la IA generativa. En la encuesta de AWS y el MIT, el 16% de los responsables de datos declaró que sus empresas habían prohibido por completo el uso de la IA generativa, y sólo el 6% señaló que estaban utilizando la IA generativa en la fase de producción.

El resto, por su parte, estaba experimentando a nivel individual, de equipo o de organización, tratando de comprender mejor las capacidades de esta herramienta. En definitiva, eso es GenAI — una herramienta. Como cualquier herramienta, puede usarse para bien o para mal.

Desde la perspectiva de la ciberseguridad, está claro que la IA generativa tiene importantes implicaciones para tres grupos diferentes de personas: los equipos de seguridad, los empleados en general y los agentes malintencionados dentro o fuera de la organización.

Equipos de seguridad + GenAI = Mejor protección

Los proveedores de seguridad llevan años integrando la IA y las funciones del *machine learning* en sus productos, con gran beneficio para sus clientes. La llegada de GenAI parece estar acelerando esa tendencia al mejorar los recursos de los software de seguridad.

Las herramientas de IA generativa son particularmente buenas en detectar amenazas e identificar ataques. Según los analistas de Bain, «la identificación de amenazas tiene el mayor potencial de la IA generativa para mejorar la ciberseguridad...» Asimismo, agregan que «la IA generativa ya está ayudando a los analistas a descubrir más rápidamente un ataque y su posible impacto. Por ejemplo, los puede ayudar a filtrar más eficazmente las alertas de incidentes, rechazando los falsos positivos. La capacidad de la IA generativa para detectar y cazar amenazas será cada vez más dinámica y automatizada».

Un estudio independiente de IBM reveló que «las organizaciones con un amplio uso tanto de la IA como de la automatización registraron un ciclo de vida de filtración de información 108 días más corto en comparación con las organizaciones estudiadas que no las han implementado (214 días frente a 322 días)». Sin embargo, el mismo estudio de IBM descubrió que el 40% de las organizaciones aún no había puesto en marcha la IA y la automatización de la seguridad.

Algunos equipos de seguridad también están utilizando IA generativa para mejorar las estrategias de confianza cero. GenAI puede ayudar a crear perfiles de riesgo para diferentes *endpoints*. Y sus funciones de correspondencia de patrones pueden ayudar a detectar cualquier evento anómalo.

GenAI también permite aumentar las funciones del personal de otras maneras. Por ejemplo, puede ayudar a investigar amenazas emergentes o proveedores potenciales. También analiza historiales para encontrar patrones, ayuda en la redacción de informes y elabora políticas diseñadas para prevenir o mitigar futuros incidentes de seguridad.

En resumen, la IA generativa puede mejorar las competencias del personal existente, ayudándolos a ser más eficientes y eficaces, y, en última instancia, mejorar la seguridad de la organización.

Aunque, no todos los efectos potenciales de la IA generativa son completamente positivos.

Empleados + GenAI = Mayor riesgo

Gran parte de las pruebas con GenAI se lleva a cabo fuera del departamento de seguridad. Es probable que los colaboradores de casi todos los equipos de una organización experimenten con la tecnología. Lamentablemente, nadie sabe realmente en qué medida utilizan los empleados la IA generativa, ni siquiera qué herramientas emplean. Del mismo modo que las organizaciones han tenido durante mucho tiempo «TI en la sombra», ahora tienen «IA en la sombra».

Eso supone un problema porque la IA generativa no sólo ofrece enormes ventajas, sino que también plantea riesgos significativos. Y uno de los posibles problemas es la fuga de información.

Las herramientas de IA generativa se basan en grandes modelos de lenguaje (LLM – por sus siglas en inglés). Estos LLM absorben un gran volumen de texto, que luego utilizan para predecir la siguiente palabra al generar un nuevo texto. Uno de sus casos de uso más útiles es la incorporación de texto existente y su optimización.

Ahora bien, si los empleados introducen información personal identificable, código privado u otro secreto de la empresa en el LLM, la herramienta puede almacenar esos datos y enviarlos a otra persona.

Los analistas de PwC explican que «las aplicaciones de GenAI podrían exacerbar los riesgos de privacidad y protección de datos. Después de todo, la promesa de los grandes modelos de lenguaje es que utilizan una gran cantidad de datos y crean aún más datos nuevos, que son vulnerables a lo tendencioso, la mala calidad, el acceso no autorizado y la pérdida».

Otro posible problema es que los desarrolladores principiantes a veces pueden usar chatbots para escribir un código inseguro. «En el contexto de la ciberseguridad, deberíamos esperar que programadores inexpertos recurran a herramientas de modelos de lenguaje predictivo para ayudarlos en sus proyectos cuando se enfrenten a un problema complejo de codificación», escribe InfoWorld.

Este añade que «aunque no es intrínsecamente negativo, pueden surgir problemas cuando las organizaciones no tienen procesos de revisión de código adecuadamente establecidos y el código se despliega sin ser examinado».

Los humanos podrían filtrar fácilmente secretos de una empresa o escribir un código inseguro sin utilizar GenAI, por supuesto. No obstante, esta nueva herramienta representa un nuevo vector de filtración de datos – especialmente uno difícil de proteger.

Esto significa que el personal necesitará más formación sobre cómo utilizarla de forma segura. Los equipos de seguridad y gestión de riesgos deberán encontrar modos de supervisar el uso que hacen los empleados de las herramientas para garantizar que no exponen innecesariamente a la organización a un mayor riesgo.

Agentes malintencionados + GenAI = Posibles desastres

La implicación más significativa de GenAI para la seguridad empresarial es la posibilidad de que los agentes malintencionados utilicen la IA generativa para cometer delitos. La misma herramienta que facilita a los equipos de seguridad a encontrar ataques también hace más fácil para los ciberdelincuentes encontrar nuevas formas de ejecutar esos ataques.

De hecho, ya están apareciendo pruebas de que los delincuentes utilizan la IA generativa. Según Bain, «las menciones a la IA generativa en la *dark web* proliferaron en 2023. Es común ver a hackers presumiendo el uso de ChatGPT».

Para algunos aspirantes a *hackers*, la IA generativa hace que sea más fácil empezar como ciberdelincuente. En lugar de aprender a escribir su propio código, pueden hacer que una de las herramientas de codificación de IA lo haga por ellos.

Las herramientas tienen algunas salvaguardas — no puedes pedirle a ChatGPT que te escriba un malware, por ejemplo. Aunque no hay que ser demasiado listo para encontrar soluciones a algunas de estas barreras. Incluso, los ciberdelincuentes comparten sus métodos con los demás.

Sin embargo, estos *hackers* novatos no son la mayor amenaza para las redes corporativas. Las herramientas de seguridad de las empresas deberían ser capaces de detener la mayoría de los ataques procedentes de amateurs. Una amenaza mucho mayor es el posible ataque de *phishing*.

La mayoría sabe que no debe fiarse de un correo con muchas palabras mal escritas. Además, las iniciativas educativas de las empresas han hecho un buen trabajo formando a la gente para que compruebe dos veces correos electrónicos inusuales.

Ahora, ¿qué ocurre cuando esos correos suenan exactamente igual que todos los demás? ¿Y si se adjunta un video o una nota de voz que pareciera y sonara exactamente igual a la de tu jefe?

Herramientas GenAI

Las empresas están añadiendo funcionalidades de IA generativa a gran variedad de productos, y cada día surgen nuevas. Algunas de las aplicaciones GenAI más conocidas son las siguientes:

ChatGPT – el revolucionario modelo de lenguaje de grande de OpenAI, que responde a preguntas y mantiene conversaciones.

GitHub Copilot – asistente de codificación de IA que se autodenomina «la herramienta para desarrolladores de IA más adoptada del mundo».

Copy.ai – herramienta de escritura diseñada para tareas como blogs y contenidos de marketing.

Scribe – asistente de escritura especializado en la creación de documentación y guías.

Bing – el motor de búsqueda de Microsoft que ahora incorpora respuestas de GPT-4.

Gemini – la alternativa de Google a ChatGPT y Bing.

Dall-E2 – herramienta de OpenAI para crear imágenes fotorrealistas a partir de textos

Synthesia – plataforma de IA que transforma texto en videos realistas.

Rephrase.ai – plataforma de prueba a video con avatares de stock y personalizados.

Bardeen – herramienta de automatización de flujos de trabajo que se encarga de tareas tediosas.

Murf.ai – generador de audio para crear locuciones basadas en voces humanas reales.

Designs.ai – herramienta de diseño gráfico para crear logotipos, videos, anuncios, y mucho más.

PwC señala: «¿El riesgo más inmediato del que preocuparse? Un *phishing* más sofisticado. Señuelos más convincentes y personalizados en chats, videos, o audios o videos falsos generados en *lives*, haciéndose pasar por alguien conocido o en posición de autoridad».

Esas mismas herramientas también podrían utilizarse para perjudicar la reputación de tu empresa en Internet. Los agentes malintencionados pueden crear imágenes, audios o videos falsos y publicarlos en las redes sociales. También, pueden amenazar con publicarlo en la web con el objetivo de exigir un rescate.

Más allá de las falsificaciones profundas *(deep fakes)*, la GenAI también puede usarse para generar códigos maliciosos. Por ejemplo, los investigadores de seguridad de HYAS utilizaron GenAI para originar un nuevo tipo de malware polimórfico que apodaron *«Black Mamba»*. Aunque parece un código benigno, *Black Mamba* en realidad se reescribe a sí mismo en tiempo de ejecución para convertirse en un *keylogger* malicioso que extrae datos a través de Microsoft Teams. Como el malware se reescribe a sí mismo continuamente, evade incluso las defensas de ciberseguridad más sólidas.

«Usando estas nuevas técnicas, un agente de amenazas puede combinar una serie de comportamientos altamente detectables de forma inusual y eludir tal detección al aprovechar la incapacidad del modelo para reconocerlo como un patrón malicioso», explicó HYAS. Además, añade, *«este problema se agrava cuando la IA es la que conduce los ciberataques, ya que los métodos que elige pueden ser muy atípicos en comparación con los utilizados por sus homólogos humanos. Asimismo, la velocidad a la que pueden ejecutarse estos ataques complica exponencialmente la amenaza»*.

Pese a lo aterradora que es *Black Mamba*, los investigadores de seguridad afirman que probablemente no sea la peor amenaza planteada por la IA generativa. Ese título tal vez pertenece a la incorporación indirecta de prompts, un tipo de ataque que subvierte las populares herramientas de GenAI, alimentándolas con datos maliciosos mediante websites aparentemente normales.

Wired informó: «En un experimento realizado en febrero, los investigadores de seguridad forzaron al chatbot Bing de Microsoft a comportarse como un estafador». Este agrega, «En las instrucciones ocultas en una página web, los investigadores le decían al chatbot que pidiera a la persona que lo usara para darle los datos de su cuenta bancaria. Este tipo de ataque, en el que la información oculta puede hacer que la IA se comporte de forma no deseada, es solo el principio».

Casi con toda seguridad, los ciberdelincuentes están — en este mismo momento — pensando en otras formas de utilizar GenAI como método de ataque. Los responsables de seguridad y TI de las empresas son conscientes de la amenaza, pero hasta ahora han hecho poco para combatirla.

Un estudio de McKinsey reveló que el 53% de los encuestados creían que GenAI suponía un riesgo para la seguridad, pero sólo el 38% está trabajando para mitigar ese riesgo. Esto nos lleva a preguntarnos: ¿qué deberían estar haciendo?

Alternativas de actuación

Si GenAI fuera sólo una amenaza, la respuesta sería obvia: las organizaciones bloquearían sus sistemas para impedir que los colaboradores accedan a ella. Emplearían los métodos más estrictos posibles para prevenir y mitigar los ataques que incorporasen funciones de GenAI. Ahora, ella no es sólo una amenaza. También es una oportunidad.

Los equipos de seguridad inteligente están buscando formas de integrar esta herramienta en sus organizaciones para ayudarlos a alcanzar sus objetivos, a la vez que combaten la amenaza. Con esto en mente, considera las siguientes alternativas de actuación

1. Sigue las medidas de seguridad existentes. La buena noticia es que las herramientas de ciberseguridad actuales ofrecen estas medidas contra las amenazas de la GenAI. Si ya dispones de medidas de seguridad sólidas, estás preparado para el nuevo mundo de la IA generativa.

2. Mejora la protección de tus modelos de IA. Conforme tu organización amplía el uso de la IA, tus modelos representan un objetivo muy atractivo. PwC señala: «GenAI añade un valioso activo para los agentes de amenazas, y para que tu organización los controle. Podrían manipular los sistemas de IA para hacer predicciones incorrectas o negar el servicio a los clientes». Por ello, la empresa recomienda: «Tu lenguaje propietario y modelos fundacionales, datos y nuevos contenidos necesitarán protecciones de ciberdefensa más fuertes».

3. Añade GenAI y automatización a tu arsenal de seguridad. Las organizaciones que despliegan tanto IA como la automatización como parte de sus defensas encuentran un malware mucho más rápido que las que no lo hacen. Además, GenAI también puede hacer que tu equipo de seguridad sea más productivo de muchas otras formas. Al incorporar estas nuevas herramientas a sus esfuerzos en curso, estarás mejor preparado para frustrar los ataques que intenten utilizar la IA generativa contra ti.

4. Educa a tu personal. Como la IA generativa es tan nueva, su contenido cambia constantemente. Insta a tus equipos a que se mantengan al día sobre la información más reciente. InfoWorld aconseja: «Ahora, lo más prudente sería replantearse la formación de tus empleados para incorporar directrices para el uso responsable de las herramientas de IA en el entorno laboral». Incluso, menciona: «La formación de los profesionales también debe tener en cuenta la sofisticación de la IA en las nuevas técnicas de ingeniería social».

5. Supervisa la normativa. No sólo debes monitorear investigaciones o información — también tienes que seguir de cerca las respuestas de los gobiernos a las nuevas herramientas. Gartner apunta: «La Ley de Inteligencia Artificial de la UE y otros marcos normativos de Norteamérica, China e India ya están estableciendo regulaciones para gestionar los riesgos de las aplicaciones de IA». Y añade: «Prepárate para el cumplimiento, más allá de lo que ya exigen normativas como las de protección de la privacidad».

6. Redacta y aplica políticas. Tus colaboradores ya utilizan IA generativa. Ahora, si eres como la mayoría de las organizaciones, probablemente aún no has implementado ninguna norma en torno a lo que crees que es apropiado para tu lugar de trabajo. Según McKinsey, «sólo el 21% de los encuestados que informan de la adopción de IA afirman que sus organizaciones han establecido políticas que regulan el uso que hacen los empleados de estas tecnologías en su trabajo». Eso es un problema porque, como destaca PwC, «sin gobernanza y supervisión adecuadas, el uso de GenAI puede crear o agudizar los riesgos legales».

7. Formaliza la gestión de riesgos. El riesgo es inherente a todo empresa. Sin embargo, las empresas inteligentes eligen cuidadosamente qué riesgos están dispuestas a asumir y cómo mitigar los posibles peligros. Ese proceso es muy difícil a menos que cuentes con un proceso formal de gestión de riesgos. Gartner pronostica, «para 2026, los modelos de IA de las organizaciones que pongan en práctica la transparencia, la confianza y la seguridad de la IA lograrán una mejora del 50% en términos de adopción de objetivos empresariales y aceptación de los usuarios». Si tu organización no tiene un proceso formal de gestión de riesgos o si no crees que tus procesos actuales sean adecuados para el reto de la IA generativa, busca ayuda de un socio externo, como los asesores de gestión de riesgos de Iron Mountain.

8. Mejora la gestión de la información. También puedes proteger tu organización contra los peligros de la GenAI al gobernar y respaldar eficazmente tus datos. Al seguir buenas prácticas de gestión de información es menos probable que sufras una filtración de datos. Y si se infiltran en tus sistemas, disponer de copias de seguridad y de recuperación ante desastres puede protegerte contra la pérdida de datos. Una vez más, lo recomendable es buscar ayuda de un proveedor de gestión de la información como Iron Mountain, para asegurarte de que estás preparado para las amenazas de GenAI.

9. Analiza cuidadosamente a los proveedores. Cuando busques herramientas de IA generativa o ayuda relacionada con la ciberseguridad y protección de datos, asegúrate de evaluar cualquier empresa externa con la que trabajes. Con las nuevas tecnologías, es tentador apresurarse a desplegar nuevas tecnologías. Si bien no debes demorarte, sí conviene que te tomes el tiempo suficiente para estar seguro de que puedes confiar en los socios que elijas para que te ayuden a integrar GenAI a tus flujos de trabajo, y a proteger tu empresa de los riesgos asociados a la GenAI.

Sobre Iron Mountain

Iron Mountain Incorporated (NYSE: IRM), fundada en 1951, es líder mundial en servicios de almacenamiento y gestión de la información. Con la confianza de más de 225.000 organizaciones en todo el mundo, y con una red de propiedades de más de 85 millones de pies cuadrados en más de 1.400 instalaciones ubicadas en más de 50 países, Iron Mountain almacena y protege miles de millones de activos de información, incluyendo información crítica para el negocio, datos altamente confidenciales y artefactos culturales e históricos. Al ofrecer soluciones que incluyen almacenamiento seguro, gestión de la información, transformación digital, destrucción segura, así como centros de datos, almacenamiento de arte y logística, y servicios en la nube, Iron Mountain ayuda a las organizaciones a reducir costos y riesgos, cumplir la normativa, recuperarse de desastres y permitir una forma de trabajar más digital.

IRON MOUNTAIN®

ironmountain.com/es-co | ironmountain.com/es-ar | ironmountain.com/es-cl | ironmountain.com/es-pe | ironmountain.com/es-mx

© 2023 Iron Mountain, Incorporated y/o sus sucursales «Iron Mountain». Todos los derechos reservados. La información proporcionada en este documento es propiedad y confidencial de Iron Mountain y/o sus licenciadores y no representa ni implica una invitación u oferta, y no puede utilizarse para el análisis competitivo o la construcción de un producto competitivo o reproducirse de otro modo sin el permiso por escrito de Iron Mountain. Iron Mountain no se compromete a ninguna disponibilidad regional o futura y no representa una afiliación con ninguna otra parte ni el respaldo de la misma. Iron Mountain no será responsable de ningún daño directo, indirecto, consecuente, punitivo, especial o fortuito derivado del uso o de la imposibilidad de uso de la información, proporcionada TAL CUAL, y no ofrece ninguna declaración ni garantía con respecto a la exactitud o integridad de la información proporcionada, ni a su idoneidad para un fin determinado. «Iron Mountain» es una marca registrada de Iron Mountain en Estados Unidos y otros países, y Iron Mountain, el logotipo de Iron Mountain y sus combinaciones, y otras marcas identificadas con ® o TM son nombres comerciales de Iron Mountain. Todas las demás marcas comerciales siguen siendo marcas comerciales de sus respectivos propietarios.