DELL Technologies

5 Recomendaciones para sacar el máximo partido a la IA generativa de forma segura

1	2	3	4	5
Proteja las capas de un sistema de IA generativa	**Utilice los principios de confianza cero**	**Mantenga la gobernanza y la supervisión humana**	**Aproveche las herramientas de seguridad con IA generativa según su disponibilidad**	**Innove con confianza**
Infraestructura	Nunca confíe, siempre verifique	Implique a las partes interesadas clave	Contenido	Apueste por la ciberseguridad para conducir la misión, no para frenarla
SO y Kubernetes	Acceso con privilegios mínimos	Establezca políticas para el cumplimiento ético y normativo, y la gestión de los datos	Predicción de riesgos	Deje que la madurez de la ciberseguridad genere confianza para innovar en la organización
Aplicaciones de IA generativa	Reforzamiento del sistema	Supervise e imponga la responsabilidad	Conocimiento y automatización	
Datos	Gestión de las identidades	Formación y educación		
	Segmentación			
	Registro, supervisión y auditoría			

La tecnología de IA generativa promete capacidades transformadoras, pero conlleva desafíos de seguridad únicos.

La IA generativa está revolucionando los negocios como nunca antes, impulsando la innovación y ofreciendo beneficios sin precedentes que proporcionan una ventaja competitiva. Aunque esta tecnología tiene un potencial transformador, también conlleva sus propios retos de seguridad.

Los expertos en la materia de Dell, Steve Brodson (gestor de productos de servicios) y Eitan Lederman (consultor de ciberseguridad), se unieron a Chris Cicotte, del equipo de marketing de IA de APEX, para abordar esas preocupaciones y hablar sobre formas de aprovechar el máximo potencial de la IA generativa de forma segura. Siga leyendo para ver un resumen de la conversación y más información sobre el tema, y vea la conversación completa en dell.com/cybersecuritymonth.

> Se trata de ofrecer formación. Las personas necesitan saber cómo usar el sistema de IA generativa. Qué hacer, pero también qué no hacer".
>
> ***Eitan Lederman***
> *Consultor de ciberseguridad de Dell*

Proteja las capas de un sistema de IA generativa

Aunque la IA generativa es una tecnología relativamente nueva, la mayoría de los protocolos de seguridad son las mismas técnicas de ciberseguridad establecidas que se utilizan para proteger otras cargas de trabajo.

Infraestructura - Céntrese en minimizar la superficie de ataque:

- Pruebas de vulnerabilidades y penetración
- Aplicación de parches
- Reforzamiento
- Gestión de identidades, con contraseñas seguras y autenticación multifactor (MFA)
- Supervisión y auditoría
- Garantizar la seguridad de la cadena de suministro de terceros

SO y Kubernetes - también se centran en la reducción de la superficie de ataque con:

- Exploración de vulnerabilidades
- Aplicación periódica de parches
- Actualización de componentes de Kubernetes
- Limitación del control de acceso basado en la gestión de identidades, el acceso basado en funciones (RBAC) y el acceso con privilegios mínimos
- Protección del plano de control, incluidos el servidor API, secrets, kubelet y otros componentes
- Uso de espacios de nombres

Aplicaciones de IA generativa -Implemente medidas de seguridad dirigidas a las nuevas superficies de ataque creadas por la IA generativa:

- Gestión de identidades para enfrentarse a la inyección rápida, la divulgación de información confidencial, el robo de modelos y el envenenamiento de datos de entrenamiento
- Validación del origen de los datos para protegerse contra el
- envenenamiento de datos de entrenamiento y el sesgo de los modelos

  Supervisión y auditoría para identificar y prevenir el DoS de

  modelos, el robo de modelos, la divulgación de información confidencial, la detección de anomalías y el análisis forense

Datos - Incorpore medidas sólidas de protección de datos para proteger los datos en el modelo de lenguaje y la aplicación:

- Vault virtual aislado
- Cifrado
- Plan de respuesta ante incidentes
- Supervisión y auditoría de los datos y resultados de entrenamiento

Garantice que los principios de protección de datos se apliquen a todos los datos, incluidas las entradas de entrenamiento, las salidas del modelo y cualquier dato implicado en la generación aumentada de recuperación (RAG), si se utilizan. Además, garantice el cumplimiento continuo de todas las normativas de protección de datos aplicables.

DELL Technologies

Utilice los principios de confianza cero

Ya se ha mencionado el papel de varios principios de confianza cero, como la gestión de identidades, el acceso con privilegios mínimos, el reforzamiento del sistema y la aplicación de parches, lo que indica el valor de los principios de confianza cero para proteger una carga de trabajo de IA generativa. Las arquitecturas de confianza cero también requieren labores de registro, supervisión y auditoría continuas de la actividad de la red, lo que puede evitar riesgos específicos de la IA generativa, como la manipulación de los resultados y el envenenamiento de datos.

Además, la confianza cero también fomenta la microsegmentación, lo que reduce el impacto de una vulneración. También requiere el cifrado de datos, tanto en tránsito como en reposo, que es una parte importante de la estrategia general de protección de datos.

Aunque estas son solo algunas de las formas en las que la confianza cero puede proteger una carga de trabajo de IA generativa, la adopción de principios de confianza cero debe considerarse una práctica recomendada.

Mantenga la gobernanza y la supervisión humana

Gran parte del valor de la IA generativa radica en automatizar tareas que normalmente ejecutarían los humanos, pero la gobernanza humana es fundamental para garantizar la seguridad y el correcto funcionamiento de las aplicaciones. Un modelo de gobernanza suele implicar a las partes interesadas clave de toda la organización, que establecen directrices y requisitos para el cumplimiento ético y normativo, las políticas y procedimientos de gestión de datos y, en última instancia, exigen responsabilidades.

La gobernanza y la supervisión adecuadas pueden ayudar a abordar problemas como el exceso de confianza en los modelos, el sesgo, la manipulación de resultados, la divulgación de información confidencial y el envenenamiento de los datos.

Lederman también señaló la importancia de la formación: "Se trata de ofrecer formación. Las personas necesitan saber cómo usar el sistema de IA generativa; qué hacer, pero también qué no hacer".

Además del riesgo que suponen las aplicaciones de IA generativa de una organización, también existe la proliferación de ciberataques basados en la IA generativa que a menudo requieren intervención humana. Algunos ejemplos son los actores maliciosos que utilizan deepfakes para implementar el comportamiento humano y los ataques de phishing, que son mucho más eficaces al imitar con mayor precisión el estilo de escritura o habla de un humano. La formación y la educación continuas son algunas de las formas más eficaces de abordar estos riesgos, reforzando una vez más el elemento humano.

Aproveche la IA generativa en las herramientas de seguridad según su disponibilidad

Aunque gran parte de la atención se centra en los riesgos, la IA generativa también tiene el potencial de reforzar la seguridad. Aunque estas capacidades están en sus inicios, ofrecerán beneficios en tres áreas clave:

- **Contenido:** Generación de políticas de seguridad, formación personalizada, clasificación de datos y elaboración de informes
- **Predicción:** de riesgo y actividad de ataque, sugerencia de medidas correctivas
- **Conocimiento:** Consulta del entorno (hablar con el sistema), análisis forense, automatización

La contribución de la IA generativa a las herramientas de seguridad podría ayudar a aprovechar al máximo la capacidad de los equipos de seguridad, reducir los costes y mejorar los sistemas de defensa. Aproveche estas soluciones a medida que crecen y maduran.

Innove con confianza

Y lo que es más importante, no permita que los riesgos de seguridad le impidan aprovechar una tecnología potencialmente revolucionaria. La eficiencia, la automatización, la reducción de costes, la resolución de problemas y el impulso de la creatividad son solo algunas de las formas en las que la IA generativa puede transformar las empresas.

Si bien la IA generativa requiere medidas de ciberseguridad sólidas y, a veces, nuevas, el objetivo debe ser impulsar la misión de la organización, no frenarla. Desarrollar una estrategia de ciberseguridad adecuada debería aportar a las organizaciones la confianza necesaria para crecer e innovar.

Descubra cómo afrontar algunos de los principales retos de ciberseguridad actuales en **dell.com/cybersecuritymonth**

DELL Technologies