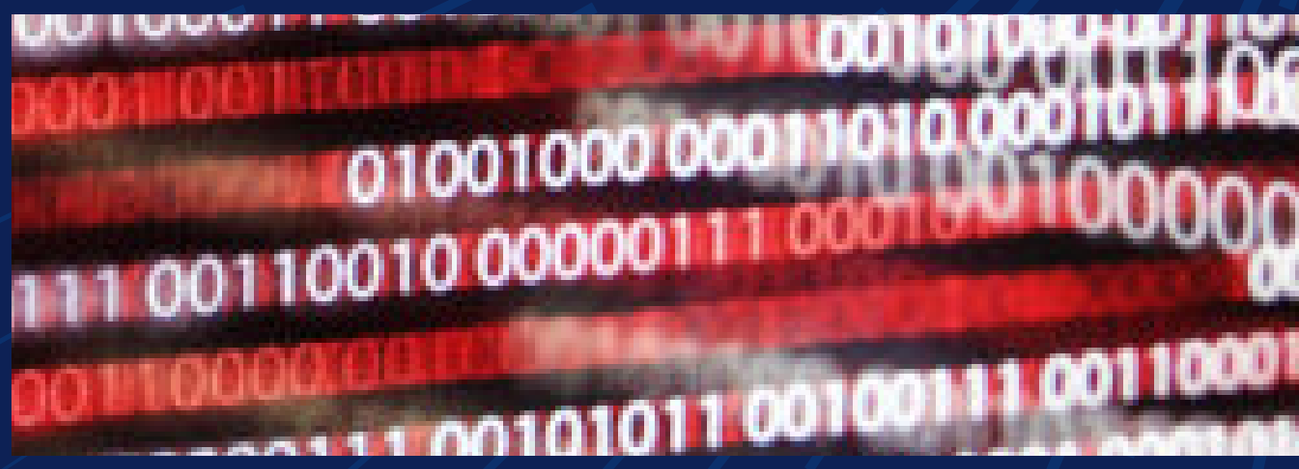
Los mitos de ciberseguridad:

Romper los mitos sobre la seguridad de la IA



La IA está transformando todos los sectores, pero muchas organizaciones son víctimas de mitos que hacen que la seguridad de esta parezca más compleja de lo que realmente es. La realidad es que proteger los sistemas de IA no requiere empezar desde cero: es fundamental aplicar los principios de ciberseguridad que ya se conocen a los retos únicos de la IA.

Desde Dell Technologies, como conocedores de la arquitectura detrás de la IA, podemos ayudarle a adaptar sus soluciones actuales a esta nueva infraestructura. Veamos los mitos más extendidos sobre la seguridad de la IA y descubramos las verdades para proteger sus sistemas de forma eficaz.

Mito 1: "Los sistemas de IA son demasiado complejos para protegerlos".

Realidad: La IA implica nuevos riesgos de ciberseguridad, como la inyección de prompts, la manipulación de datos y la divulgación de información confidencial. Además, los sistemas de IA agéntica también tienen una superficie de ataque más amplia, ya que se pueden aprovechar para manipular resultados o dar privilegios.

Dicho esto, aunque es fundamental reconocer estas vulnerabilidades e implementar medidas de seguridad para proteger los sistemas de IA de las amenazas tradicionales y específicas, se pueden gestionar los riesgos y proteger los modelos de IA. Es importante tener en cuenta que los sistemas de IA requieren y crean grandes cantidades de datos como entradas y salidas, respectivamente. Estos hechos sitúan a la protección de datos como una de las principales estrategias de seguridad junto con las siguientes:

- Principios de confianza cero, como la gestión de identidades, el acceso basado en funciones y la verificación continua.
- Pruebas periódicas de penetraciones y gestión de vulnerabilidades para identificar las debilidades.
- Registro y auditoría para validar las entradas y salidas de datos

Mito 2: "Ninguna de las herramientas que ya tengo protege a la IA".

Realidad: Garantizar la seguridad de la IA no implica comenzar de cero, sino trabajar de forma más inteligente con las herramientas que ya se tienen. La mayoría de las herramientas de ciberseguridad existentes se pueden adaptar para proteger los sistemas de IA de forma eficaz. Esencialmente, la IA es una carga de trabajo más en su arsenal para impulsar su negocio, aunque con características únicas. Las prácticas básicas de ciberseguridad, como la gestión de identidades, la segmentación y supervisión de redes, y la protección de puntos finales y de datos, siguen siendo fundamentales para proteger los entornos de IA. La clave es adaptar estas prácticas a los retos específicos de la IA, entre otros, la protección de los datos de entrenamiento, la seguridad de los algoritmos y la mitigación de riesgos, como las entradas adversarias.

Una defensa sólida comienza con una buena ciberhigiene, con prácticas que incluyen la aplicación de parches a los sistemas, el control de acceso y la gestión de vulnerabilidades. Lo importante es adaptarlas para abordar los riesgos específicos de la IA. Con la integración de estrategias de protección en su enfoque actual y las herramientas adecuadas, la seguridad de la IA es fácil de gestionar y eficaz.

Sin embargo, es importante señalar que la actualización del hardware puede ser clave en la lucha contra los ciberataques. Por ejemplo, los PC modernos con IA son una primera línea de defensa sólida contra un vector de ataque importante: los puntos finales. Con el fin de la asistencia para Windows 10, los PC obsoletos se convierten en un riesgo. Además, Windows 11 requiere la versión 2.0 de Trusted Platform Module (TPM), un chip de seguridad que contribuye al cifrado, al arranque seguro y a la protección contra ataques al firmware. Muchos PC antiguos no tienen TPM o solo admiten una versión anterior. Dell ofrece PC comerciales con IA con estas mejoras de seguridad integradas.

Lo mismo ocurre con la infraestructura de IA, como los servidores y el almacenamiento. Dell AI Factory incluye hardware optimizado para la seguridad de IA y contiene varias características de seguridad integradas, que van desde una cadena de suministro segura hasta la inmutabilidad, el aislamiento y el cifrado de los datos.

Mito 3: "La seguridad de la IA solo consiste en proteger los datos".

Realidad: La seguridad de la IA va más allá de la protección de datos básica: implica proteger todo el ecosistema de IA, incluidos los modelos, las API, las salidas, los sistemas y los dispositivos. A medida que la IA se integra más en las aplicaciones esenciales, aumentan los riesgos asociados a su uso indebido o su explotación. Sin medidas de seguridad sólidas, los modelos de IA se pueden manipular para generar resultados dañinos o engañosos, las API se pueden aprovechar para obtener acceso no autorizado a sistemas confidenciales y los resultados pueden exponer involuntariamente información privada o confidencial.

La seguridad integral de la IA requiere un enfoque multicapa. Esto incluye proteger los modelos de ataques adversarios que intentan manipular los datos de entrada para engañar a los sistemas de IA, proteger las API con

métodos de autenticación sólidos para evitar el uso no autorizado y **supervisar continuamente los resultados** en busca de patrones inusuales o sospechosos que podrían indicar un ataque o un mal funcionamiento. La seguridad eficaz de la IA no solo garantiza la integridad y fiabilidad de los sistemas de IA, sino que también genera confianza entre los usuarios y las partes interesadas al mitigar los riesgos de uso malicioso y las consecuencias indeseadas.

Mito 4: "La IA no necesita supervisión humana".

Realidad: La gobernanza y la supervisión humana son fundamentales para garantizar que los sistemas de IA funcionen de forma ética, predecible y en línea con los valores humanos. Los sistemas avanzados de IA, en particular la IA agéntica con capacidades de toma de decisiones autónomas, presentan desafíos únicos que exigen medidas de protección sólidas. Sin la supervisión adecuada, estos sistemas podrían desviarse de los objetivos previstos o tener comportamientos no deseados que pueden suponer riesgos.

Para hacer frente a esto, es fundamental establecer límites claros, implementar mecanismos de control por capas y garantizar la constante participación humana en los procesos clave de toma de decisiones. Las auditorías periódicas, la transparencia en las operaciones de IA y las pruebas exhaustivas pueden aumentar aún más la responsabilidad y la confianza, lo que evita el uso indebido y promueve la implementación responsable de las tecnologías de IA.

Procedimientos recomendados para reforzar la seguridad de la IA

Para eliminar las brechas de seguridad específicas de la IA, las organizaciones deben adoptar un enfoque proactivo y estratégico. Descubra 10 procedimientos recomendados para proteger sus sistemas de IA:

Arquitectura de seguridad a capas:
Utilice la segmentación, los firewalls y un sistema de autenticación sólido para proteger su infraestructura, software y datos en todas las capas.

Asegurar la cadena de suministro:
Implemente un programa sólido de gestión de proveedores. Audite a los proveedores y a los componentes de terceros, valide la integridad y confíe en el código firmado para evitar vulnerabilidades en el ciclo de vida de desarrollo de la IA.

Proteger los modelos y datos de entrenamiento:
Protéjase frente a los datos envenenados, entradas adversarias y otras amenazas mediante la supervisión de la integridad de los datos y la aplicación de herramientas de validación sólidas.

Reforzar los controles de acceso:
Aplique los principios de privilegios mínimos, implemente el control de acceso basado en funciones (RBAC), rote las credenciales con frecuencia y audite los permisos para evitar el acceso no autorizado.

API seguras:
Utilice protocolos de autenticación seguros (como OAuth 2.0), aplique el cifrado HTTPS y actualice las API periódicamente para acabar con posibles vulnerabilidades.

Supervisar y validar los resultados de IA:
Utilice la detección de anomalías, el registro y las alertas para supervisar patrones inusuales o comportamientos perjudiciales en los resultados de IA.

Plan de resiliencia:
Haga copias de seguridad periódicas y pruebe los planes de recuperación ante desastres para minimizar el tiempo de inactividad y garantizar una recuperación rápida en caso de vulneración.

Implementar un cifrado sólido:
Cifre los datos confidenciales en reposo y en tránsito mediante algoritmos sólidos, y gestione y rote las claves de cifrado de forma segura y regular.

Hacer auditorías de seguridad y pruebas de penetración periódicas:
Evalúe con frecuencia los sistemas en busca de vulnerabilidades y utilice pruebas de penetración para detectar riesgos antes de que puedan darse.

Formar al personal sobre los procedimientos recomendados de seguridad de la IA:
Dé formación periódica a su equipo sobre el desarrollo seguro, el reconocimiento de amenazas y el mantenimiento de prácticas de seguridad sólidas para evitar vulneraciones.

DELL Technologies

Propuesta de valor de Dell: soluciones prácticas de seguridad de IA.

Aunque la seguridad de la IA puede parecer compleja, no es tan abrumadora como parece. La realidad es que proteger la IA no es tan distinto a proteger las cargas de trabajo existentes. El secreto está en conocer la arquitectura y aplicar las estrategias adecuadas. Ahí es donde interviene Dell Technologies.

Desmitificamos la seguridad de la IA aprovechando sus soluciones actuales e integrándolas fácilmente en arquitecturas centradas en la IA. Abordamos retos como la inyección de prompts, el abuso de las API y los ataques adversarios sin necesidad de una revisión completa de la infraestructura.

La experiencia de Dell permite desmitificar la seguridad de la IA y demostrar su factibilidad. Tanto si acaba de comenzar su transición a la IA como si desea mejorar sus defensas, le ayudaremos a proteger sus inversiones, asegurar sus sistemas y crear un futuro digital resiliente, con confianza y eficacia. Simplifiquemos juntos la seguridad de la IA.

Productos y soluciones Dell destacados

Solución Dell destacada	Descripción
Dell AI Factory	Dell AI Factory protege las cargas de trabajo de IA a través de una cadena de suministro segura para garantizar una infraestructura de confianza, desde el desarrollo hasta la implementación. Con funciones como la inmutabilidad, el aislamiento y el cifrado de datos, protege los modelos y conjuntos de datos confidenciales, garantiza la seguridad contra ciberamenazas y permite operaciones de IA ampliables, eficientes y fluidas en entornos dinámicos y basados en datos.
Ciberresiliencia	PowerProtect protege las cargas de trabajo de IA con funciones avanzadas, como la inmutabilidad y el aislamiento, para garantizar la integridad de los datos y la protección frente a ciberamenazas. Ofrece cifrado integral y detección de anomalías, y permite una recuperación rápida para minimizar el tiempo de inactividad.
Dell Trusted Workspace (seguridad de puntos finales)	Una combinación de funciones complementarias integradas y opcionales diseñadas para proteger los PC comerciales con IA y las cargas de trabajo de IA que se ejecutan en ellos. La solución está desarrollada con procedimientos seguros de la cadena de suministro y funciones integradas como SafeBIOS y SafeID con TPM. Los complementos opcionales incluyen la verificación de componentes seguros, SafeID con ControlVault y software de socios CrowdStrike y Absolute para maximizar la seguridad del espacio de trabajo.
Servicios de asesoramiento para la seguridad de la IA	Un conjunto de herramientas para ayudarle a desarrollar e implementar una estrategia de seguridad de IA. La oferta incluye servicios de asesoría, vCISO de IA y planificación de la seguridad de los datos.
Operaciones de seguridad gestionadas para IA	Permite una visibilidad profunda de toda la pila para detectar y responder rápidamente a las amenazas Entre las funciones se incluyen Managed Detection and Response, protección gestionada de IA, pruebas de penetración para IA y servicios de respuesta y recuperación ante incidentes.
Integración de software de seguridad	Diseñe, instale y configure herramientas de seguridad que protejan la gestión de acceso, aplicaciones, redes, clouds y mucho más.

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