

Experiencia de simulación:

el desafío de las instrucciones de IA

Descripción general

Las herramientas de IA nos ayudan a trabajar más rápido, pero utilizarlas de forma incorrecta es peligroso. Si bien una herramienta autorizada ofrece mayor privacidad que una pública, la forma en que interactúa con ella igualmente puede crear importantes vulnerabilidades de seguridad. Por eso es fundamental practicar la seguridad de las instrucciones.

Por ejemplo, las instrucciones mal estructuradas pueden conducir a conclusiones erróneas a partir de resultados inexactos de la IA y permitir el acceso no autorizado a información de alto nivel. Esta actividad se enfoca en el uso de una herramienta autorizada para mejorar el flujo de trabajo, todo ello siguiendo las prácticas recomendadas de seguridad en las instrucciones.

Roles

Persona facilitadora:

controla el ritmo del ejercicio, modera los debates y garantiza la participación de todas las personas.

Participantes:

hablan abiertamente sobre sus procesos de razonamiento y responden a las situaciones presentadas siguiendo las reglas del ejercicio.

Introducción

Presente a quienes participan la siguiente situación: Su organización autorizó el uso de una herramienta de IA para todo el personal. Esto significa que usted cuenta con un espacio de trabajo privado en el que sus datos no se comparten con el público. De todos modos, debe seguir las políticas de seguridad mencionadas a continuación para crear instrucciones:

- Nunca ingrese información de identificación personal (PII) (p. ej., nombres, números de identidad, fechas de nacimiento) ni información de salud protegida (PHI)
- No utilice secretos comerciales ni documentos confidenciales para crear instrucciones.
- No ingrese información bancaria, información de inicio de sesión y contraseñas en la herramienta de IA.

Hacia el final del primer mes, la gerencia compartió un informe de uso de instrucciones y le pidió a su equipo que las analice y clasifique en dos categorías: seguras e inseguras.

Parte 1: anotación de señales

 **Duración:** 15 minutos

Materiales:

- Cronómetro o temporizador.
- Hojas impresas con las situaciones.
- Bolígrafos o lápices.
- Hojas de papel o anotadores.

Instrucciones:

- Divida al salón en grupos pequeños.
- Entregue las hojas impresas con las situaciones y los bolígrafos o lápices a cada grupo.
- Lea la lista de instrucciones de la hoja impresa.
- Indíqueles a los grupos que subrayen los elementos que describen prácticas seguras de IA. Algunos ejemplos son el uso de enunciados generales y el hecho de no mencionar nombres específicos de proyectos y de la clientela.
- Indíqueles a los grupos que **marquen con un círculo** las señales de advertencia en el comportamiento. Algunos ejemplos son la mención de datos delicados e información confidencial en los sistemas de IA.
- Explique la importancia de seguir las políticas de la organización a fin de asegurarse de usar siempre el contenido autorizado.
- Debate en grupo: ¿Cómo se podrían volver a escribir las instrucciones inseguras para volverlas seguras? ¿Qué instrucciones seguras les serían útiles para sus tareas diarias?
- Deles a los grupos 15 minutos para hablar sobre la situación y tomar notas.

Parte 2: debate

 **Duración:** 10 minutos

Preguntas:

- ¿Qué señales de advertencia observó?
- ¿Que señales de advertencia le causaron confusión al principio?
- ¿Qué prácticas seguras del uso de la IA le resultaron familiares?
- ¿Cómo practica la seguridad de las instrucciones en su trabajo diario y en su vida personal?

Posibles respuestas/instrucciones:

- Las instrucciones que se utilizan para ayudar al personal en sus tareas de rutina no incluyen información confidencial. El personal está utilizando una herramienta interna autorizada, por lo que estas instrucciones son seguras. Los datos no se usan para entrenar modelos públicos.
- Señales de advertencia: Poner en riesgo la seguridad personal, mezclar datos personales con las herramientas de trabajo y divulgar información de dominio privado.

Conclusión

 **Duración:** 5 minutos

Concluya el ejercicio relacionando los conceptos directamente con la realidad de su organización. Explique que detectar una señal de advertencia es solo la primera parte de la defensa; la segunda consiste en tomar las medidas adecuadas siguiendo las políticas de la organización.

Aproveche la oportunidad para responder cualquier pregunta que puedan tener las personas que participan sobre cómo crear instrucciones seguras en el lugar de trabajo para mejorar el uso de la IA.

Intercambie experiencias y opiniones sobre el uso de la IA en el lugar de trabajo y comparta las prácticas recomendadas.

Adjunto: respuestas para la persona facilitadora

ID	Instrucción de IA del personal	Categoría
A	Genera 10 titulares llamativos para nuestro nuevo servicio de seguimiento logístico.	Segura: la IA se usa para una lluvia de ideas creativa.
B	Esta es mi lista de tareas pendientes para esta semana. Organizar las tareas por nivel de complejidad para aprovechar mejor mi tiempo.	Segura: uso de la herramienta para organizarse dentro del “espacio de trabajo privado” seguro para optimizar su trabajo.
C	Escribe un resumen de este documento confidencial sobre la fusión para la dirección ejecutiva.	Insegura: aunque se trate de una herramienta segura, los documentos muy delicados con frecuencia deben seguir reglas adicionales acerca de dónde pueden cargarse. Conforme a las políticas de seguridad de la situación, esta acción constituye una infracción.
D	Redacta una respuesta a una persona de la clientela que está feliz con nuestro nuevo almacén en Ciudad de México.	Segura: descripción de información general para mejorar la comunicación.
E	Inicié sesión como invitado. Esta es mi contraseña personal del banco. ¿Puedes ayudarme a recordarla?	Insegura: nunca se deben ingresar contraseñas en una herramienta de IA, aunque sean herramientas autorizadas.
F	Redacta un correo electrónico para alguien de mis colegas que está pidiendo una actualización sobre una tarea que debió entregarse ayer, y asegúrate de que el tono siga siendo colaborativo.	Segura: solicitar mejoras de redacción para tareas de rutina en el lugar de trabajo.
G	Esta es una lista con nombres, cumpleaños y ubicaciones de nuestra clientela más importante, y la fecha en que comenzaron a trabajar con nosotros. Redacta mensajes de cumpleaños personalizados y creativos para la clientela.	Insegura: ingresar información de identificación personal (PII) en herramientas de IA es una práctica insegura. No especifique información personal de la clientela en sus instrucciones.
H	Explica las diferencias básicas entre el rol de una persona que es “gerente de proyecto” y el de una persona que es “gerente de producto” en una empresa emergente tecnológica.	Segura: se hacen preguntas en general sin divulgar información de dominio privado.

Adjunto: hoja impresa con situaciones

Analice las siguientes instrucciones y clasifíquelas en “seguras” e “inseguras” según las políticas de seguridad de la situación:

ID	Instrucción de IA del personal	Categoría
A	Genera 10 titulares llamativos para nuestro nuevo servicio de seguimiento logístico.	
B	Esta es mi lista de tareas pendientes para esta semana. Organizar las tareas por nivel de complejidad para aprovechar mejor mi tiempo.	
C	Escribe un resumen de este documento confidencial sobre la fusión para la dirección ejecutiva.	
D	Redacta una respuesta a una persona de la clientela que está feliz con nuestro nuevo almacén en Ciudad de México.	
E	Inicié sesión como invitado. Esta es mi contraseña personal del banco. ¿Puedes ayudarme a recordarla?	
F	Redacta un correo electrónico para alguien de mis colegas que está pidiendo una actualización sobre una tarea que debió entregarse ayer, y asegúrate de que el tono siga siendo colaborativo.	
G	Esta es una lista con nombres, cumpleaños y ubicaciones de nuestra clientela más importante, y la fecha en que comenzaron a trabajar con nosotros. Redacta mensajes de cumpleaños personalizados y creativos para la clientela.	
H	Explica las diferencias básicas entre el rol de una persona que es “gerente de proyecto” y el de una persona que es “gerente de producto” en una empresa emergente tecnológica.	