

Experiência de simulação:

O desafio do prompt de IA

Visão geral

As ferramentas de IA aceleram o trabalho, mas seu uso de maneira errada é muito perigoso. Embora uma ferramenta aprovada ofereça mais privacidade do que uma pública, a forma como você interage com ela ainda pode criar vulnerabilidades de segurança significativas. Por isso é importante praticar a segurança no uso de prompts.

Por exemplo, prompts mal-estruturados podem levar a conclusões enganosas devido a resultados imprecisos da IA, além de permitir o acesso não autorizado a informações confidenciais. Esta atividade tem como foco o uso de uma ferramenta oficial aprovada para melhorar seu fluxo de trabalho, seguindo as melhores práticas de segurança no uso de prompts.

Funções

Facilitador:

Controla o ritmo do exercício, orienta as discussões e garante a participação de todos.

Participantes:

Discutem abertamente seu raciocínio e respondem aos cenários de acordo com as regras do exercício.

Introdução

Apresente aos participantes este cenário: Sua organização aprovou uma ferramenta de IA para todos os funcionários. Isso significa que você tem um acesso privado no qual seus dados não são compartilhados com o público. Ainda assim, você deve seguir as políticas de segurança abaixo para criar prompts:

- Nunca insira informações pessoais identificáveis (Personally Identifiable Information, PII) – por exemplo, nomes, números de identidade, datas de nascimento – ou informações de saúde protegidas (Protected Health Information, PHI).
- Não utilize segredos comerciais e documentos confidenciais para criar prompts.
- Não insira dados bancários, logins e senhas na ferramenta de IA.

Ao final do primeiro mês, sua gerência compartilhou um relatório de uso dos prompts e pediu à sua equipe que discutisse os resultados, classificando os prompts como “seguros” ou “não seguros”.

Parte 1: Anotação dos sinais

 **Duração:** 15 minutos

Materiais:

- Temporizador ou cronômetro
- Folhetos de cenários
- Canetas ou lápis
- Folhas de papel ou cadernos

Instruções:

- Divida a sala em pequenos grupos.
- Distribua cenários impressos, canetas ou lápis para cada grupo.
- Leia a lista de prompts no folheto.
- Instrua os grupos a **grifar** os elementos que descrevem práticas de IA seguras. Exemplos incluem o uso de declarações genéricas e a omissão de nomes específicos de projetos e clientes.
- Instrua os grupos a **circular** sinais de alerta comportamentais. Exemplos incluem fornecer dados sensíveis e informações confidenciais aos sistemas de IA.
- Explique a importância de seguir as políticas da organização para assegurar o uso de conteúdo autorizado sempre.
- Discuta em grupo: Como você reescreveria os prompts não seguros para torná-los mais seguros? Quais prompts seguros seriam úteis para suas tarefas diárias?
- Reserve aos grupos 15 minutos para que discutam o cenário e façam anotações.

Parte 2: Debate

 **Duração:** 10 minutos

Perguntas:

- Quais sinais de alerta você identificou?
- Que sinais de alerta enganaram você inicialmente?
- Quais práticas seguras para uso de IA você já conhecia?
- Como você pratica a segurança no uso de prompts no seu trabalho diário e na sua vida pessoal?

Respostas possíveis/sugestões:

- Prompts utilizados para ajudar os funcionários nas tarefas rotineiras não incluem informações confidenciais. Os funcionários estão usando uma ferramenta interna aprovada, portanto esses prompts são seguros. Os dados não são usados para treinar modelos públicos.
- Sinais de alerta: colocar em risco a segurança pessoal, misturar dados pessoais privados com ferramentas de trabalho e divulgar informações exclusivas.

Conclusão

 **Duração:** 5 minutos

Para finalizar o exercício, conecte os conceitos diretamente à realidade da sua organização. Explique que identificar um sinal de alerta é apenas a primeira parte da defesa; a segunda é tomar a atitude correta seguindo as políticas da organização.

Aproveite a oportunidade para responder perguntas dos participantes sobre como criar prompts seguros no local de trabalho para melhorar o uso da IA.

Troque experiências e ideias sobre o uso da IA no local de trabalho e compartilhe as melhores práticas.

Anexo – Gabarito do facilitador

ID	O prompt de IA do funcionário	Categoria
A	Crie 10 títulos chamativos para o nosso novo serviço de rastreamento logístico.	Seguro: usar a IA para brainstorming criativo.
B	Esta é minha lista de tarefas para a semana. Organize as tarefas por complexidade para otimizar meu tempo.	Seguro: usar a ferramenta para organização dentro de um acesso privado seguro, a fim de otimizar seu trabalho.
C	Elabore um resumo deste documento confidencial de fusão para o CEO.	Não seguro: mesmo em ferramentas seguras, documentos altamente confidenciais costumam seguir regras adicionais sobre onde podem ser enviados. De acordo com as políticas de segurança do cenário, este é um caso de violação.
D	Elabore uma resposta para um cliente que está satisfeito com o nosso novo armazém em Berlim.	Seguro: utilizar informações gerais para melhorar a comunicação.
E	Estou usando uma conta de convidado. Aqui está minha senha bancária pessoal. Você pode me ajudar a lembrá-la?	Não seguro: nunca insira senhas em nenhuma ferramenta de IA, nem mesmo nas aprovadas.
F	Escreva um e-mail educado para um colega solicitando uma atualização sobre uma tarefa que deveria ter sido entregue ontem, garantindo que o tom permaneça colaborativo.	Seguro: solicitar melhorias em textos para tarefas rotineiras do trabalho.
G	Esta é uma lista dos nomes completos dos nossos principais clientes, datas de aniversário, localizações e datas em que começaram a trabalhar conosco. Elabore mensagens de parabéns criativas e personalizadas para eles.	Não seguro: inserir informações pessoais identificáveis (PII) em ferramentas de IA é uma prática arriscada. Não especifique informações pessoais do cliente em seus prompts.
H	Explique as principais diferenças entre as funções de “gerente de projeto” e “gerente de produto” em um ambiente de startup de tecnologia.	Seguro: fazer uma pergunta geral sem revelar detalhes confidenciais.

Anexo – Folheto de cenário

Analise os seguintes prompts e classifique-os em “seguros” e “não seguros” de acordo com as políticas de segurança do cenário:

ID	O prompt de IA do funcionário	Categoria
A	Crie 10 títulos chamativos para o nosso novo serviço de rastreamento logístico.	
B	Esta é minha lista de tarefas para a semana. Organize as tarefas por complexidade para otimizar meu tempo.	
C	Elabore um resumo deste documento confidencial de fusão para o CEO.	
D	Elabore uma resposta para um cliente que está satisfeito com o nosso novo armazém em Berlim.	
E	Estou usando uma conta de convidado. Aqui está minha senha bancária pessoal. Você pode me ajudar a lembrá-la?	
F	Escreva um e-mail educado para um colega solicitando uma atualização sobre uma tarefa que deveria ter sido entregue ontem, garantindo que o tom permaneça colaborativo.	
G	Esta é uma lista dos nomes completos dos nossos principais clientes, datas de aniversário, localizações e datas em que começaram a trabalhar conosco. Elabore mensagens de parabéns criativas e personalizadas para eles.	
H	Explique as principais diferenças entre as funções de “gerente de projeto” e “gerente de produto” em um ambiente de startup de tecnologia.	