

Tabletop-Übung:

Die KI-Prompt-Challenge

Überblick

KI-Tools können unsere Arbeit beschleunigen, aber eine falsche Verwendung ist gefährlich. Auch wenn ein genehmigtes Tool mehr Datenschutz bietet als ein öffentliches, kann die Art und Weise, wie Sie damit interagieren, enorme Sicherheitsschwachstellen verursachen. Deshalb ist es entscheidend, die Verwendung sicherer Prompts zu üben.

Ein schlecht strukturierter Prompt kann beispielsweise zu irreführenden Schlussfolgerungen durch falsche KI-Ausgaben führen und den unbefugten Zugriff auf geschützte Informationen ermöglichen. Bei dieser Aktivität liegt der Schwerpunkt auf der Verwendung eines offiziell genehmigten Tools zur Verbesserung Ihres Workflows unter Berücksichtigung sicherer Best Practices für Prompts.

Rollen

Moderation:

Steuert den Ablauf der Übung, leitet Diskussionen und bezieht, für eine aktive Mitarbeit, alle mit ein.

Teilnehmende:

Besprechen offen ihre Gedankengänge und reagieren im Rahmen der Regeln der Übung auf die verschiedenen Szenarien.

Einführung

Stellen Sie den Teilnehmenden dieses Szenario vor: Ihre Organisation hat ein KI-Tool für alle Mitarbeitenden genehmigt. Das bedeutet, sie haben einen privaten Rahmen zur Verfügung, in dem die Daten nicht öffentlich zugänglich gemacht werden. Dennoch müssen Sie folgende Sicherheitsmaßnahmen für die Erstellung von Prompts befolgen:

- Niemals persönlich identifizierbare Informationen (Personally Identifiable Information, PII) wie Namen, Ausweisdaten, Geburtsdaten oder geschützte Gesundheitsinformationen (Protected Health Information, PHI) eingeben
- Keine Betriebsgeheimnisse oder vertraulichen Dokumente für die Erstellung von Prompts verwenden
- Keine Bankdaten, Anmeldedaten und Passwörter in das KI-Tool eingeben

Am Ende des ersten Monats stellte Ihre Führungskraft einen Bericht zur Prompt-Nutzung vor und bat das Team, die Prompts zu besprechen und sie in die Kategorien „sicher“ und „unsicher“ einzuordnen.

Teil 1: Identifizieren von Warnsignalen

 **Dauer:** 15 Minuten

Materialien:

- Timer oder Stoppuhr
- Szenario-Handouts
- Stifte
- Blätter oder Notizbücher

Anweisungen:

- Teilen Sie die Anwesenden in kleine Gruppen ein.
- Händigen Sie jeder Gruppe die Ausdrucke der Szenarien und Stifte aus.
- Lesen Sie die Liste der Prompts im Handout vor.
- Weisen Sie die Gruppen an, die Elemente zu unterstreichen, die sichere KI-Praktiken beschreiben. Beispiel: allgemeine Aussagen ohne konkrete Projekt- oder Kundennamen.
- Weisen Sie die Gruppen an, Verhaltenswarnsignale **einzukreisen**. Beispiel: Eingabe sensibler Daten oder vertraulicher Informationen in KI-Systeme.
- Erläutern Sie, wie wichtig es ist, die Richtlinien der Organisation zu befolgen, damit Sie immer autorisierten Content nutzen.
- Besprechen Sie Folgendes in der Gruppe: Wie könnten Sie die unsicheren Prompts umformulieren, um sie sicherer zu gestalten? Welche sicheren Prompts wären für Ihre alltäglichen Aufgaben hilfreich?
- Geben Sie den Gruppen 15 Minuten Zeit, das Szenario zu besprechen und sich Notizen zu machen.

Teil 2: Diskussion

 **Dauer:** 10 Minuten

Fragen:

- Welche Warnsignale haben Sie bemerkt?
- Welche Warnsignale schienen anfangs harmlos?
- Welche sicheren Praktiken für die KI-Nutzung waren Ihnen bekannt?
- Wie praktizieren Sie in Ihrem Arbeitsalltag und im Privatleben Prompt-Sicherheit?

Mögliche Antworten/Prompts:

- Prompts, die Mitarbeitenden bei Routineaufgaben helfen sollen, beinhalten keine vertraulichen Informationen. Die Mitarbeitenden verwenden ein genehmigtes internes Tool, daher sind diese Prompts sicher. Die Daten werden nicht verwendet, um öffentliche Modelle zu trainieren.
- Warnsignale: Risiko für die persönliche Sicherheit, Vermischung privater personenbezogener Daten mit beruflich genutzten Tools und Offenlegung geschützter Informationen.

Fazit

 **Dauer:** 5 Minuten

Schließen Sie die Übung ab, indem Sie die Konzepte in direkten Zusammenhang mit der Realität Ihrer Organisation stellen. Erklären Sie, dass das Erkennen von Warnsignalen nur der erste Teil der Abwehr ist. Der zweite Teil besteht darin, die richtigen Maßnahmen in Einklang mit den Richtlinien der Organisation zu ergreifen.

Beantworten Sie mögliche Fragen von Teilnehmenden zur Erstellung sicherer Prompts am Arbeitsplatz für eine bessere KI-Nutzung.

Tauschen Sie Erfahrungen und Gedanken zur KI-Nutzung am Arbeitsplatz sowie Best Practices aus.

Anhang – Antwortschlüssel für die Moderation

ID	KI-Prompt der Person	Kategorie
A	Generiere 10 ansprechende Überschriften für unseren neuen Logistik-Tracking-Dienst.	Sicher: Verwendung von KI für kreatives Brainstorming.
B	Das ist meine To-do-Liste für diese Woche. Organisiere die Aufgaben nach Komplexität, damit ich meine Zeit bestmöglich nutzen kann.	Sicher: Verwendung des Tools für die Organisation innerhalb des privaten Rahmens zur Optimierung der Arbeit.
C	Fasse dieses vertrauliche Dokument bezüglich einer Fusion für den CEO zusammen.	Unsicher: Selbst in sicheren Tools gelten besondere Regeln dazu, wo vertrauliche Dokumente hochgeladen werden dürfen. Nach den Sicherheitsrichtlinien des Szenarios handelt es sich um einen Verstoß.
D	Verfasse eine Antwort an eine Kundin, die mit unserem neuen Lager in Berlin zufrieden ist.	Sicher: Beschreibung allgemeiner Informationen zur Verbesserung der Kommunikation.
E	Ich bin mit einem Gastzugang angemeldet. Hier ist mein persönliches Bank-Passwort. Wie kann ich es mir am besten merken?	Unsicher: Geben Sie niemals Passwörter in KI-Tools ein, auch nicht in genehmigte.
F	Verfasse eine höfliche E-Mail an einen Kollegen und bitte um ein Update zu einer Aufgabe, die gestern fällig war. Der Tonfall soll konstruktiv bleiben.	Sicher: Bitte um Textverbesserungen für Routineaufgaben am Arbeitsplatz.
G	Hier ist eine Liste mit den vollständigen Namen, Geburtsdaten und Standorten unserer wichtigsten Kunden und Kundinnen sowie jeweils dem Datum, seit dem sie mit uns zusammenarbeiten. Erstelle jeweils einen Entwurf für personalisierte Geburtstagsgrüße.	Unsicher: Die Eingabe persönlich identifizierbarer Informationen (PII) in KI-Tools ist eine unsichere Praxis. Geben Sie in Ihren Prompts niemals personenbezogene Kundendaten an.
H	Erkläre die Hauptunterschiede zwischen den Rollen als „Projektverantwortliche“ und „Produktverantwortliche“ in einer Tech-Startup-Umgebung.	Sicher: Allgemeine Frage ohne Offenlegung geschützter Informationen.

Anhang – Szenario-Handout

Analysieren Sie die folgenden Prompts und ordnen Sie sie nach den Sicherheitsrichtlinien für das Szenario in die Kategorien „sicher“ und „unsicher“ ein:

ID	KI-Prompt der Person	Kategorie
A	Generiere 10 ansprechende Überschriften für unseren neuen Logistik-Tracking-Dienst.	
B	Das ist meine To-do-Liste für diese Woche. Organisiere die Aufgaben nach Komplexität, damit ich meine Zeit bestmöglich nutzen kann.	
C	Fasse dieses vertrauliche Dokument bezüglich einer Fusion für den CEO zusammen.	
D	Verfasse eine Antwort an eine Kundin, die mit unserem neuen Lager in Berlin zufrieden ist.	
E	Ich bin mit einem Gastzugang angemeldet. Hier ist mein persönliches Bank-Passwort. Wie kann ich es mir am besten merken?	
F	Verfasse eine höfliche E-Mail an einen Kollegen und bitte um ein Update zu einer Aufgabe, die gestern fällig war. Der Tonfall soll konstruktiv bleiben.	
G	Hier ist eine Liste mit den vollständigen Namen, Geburtsdaten und Standorten unserer wichtigsten Kunden und Kundinnen sowie jeweils dem Datum, seit dem sie mit uns zusammenarbeiten. Erstelle jeweils einen Entwurf für personalisierte Geburtstagsgrüße.	
H	Erkläre die Hauptunterschiede zwischen den Rollen als „Projektverantwortliche“ und „Produktverantwortliche“ in einer Tech-Startup-Umgebung.	