

Exercice sur table :

Le défi des requêtes d'IA

Présentation

Les outils d'IA nous permettent de travailler plus vite, mais une mauvaise utilisation présente des risques importants. Même si un outil approuvé par votre organisation offre une meilleure confidentialité qu'un outil public, votre façon de l'utiliser peut tout de même créer des failles de sécurité majeures. C'est pourquoi il est essentiel d'adopter des pratiques sécurisées lors de l'élaboration de vos requêtes (l'art du « prompting »).

Par exemple, des requêtes mal structurées peuvent générer des réponses erronées de la part de l'IA, vous induire en erreur ou encore accorder un accès non autorisé à des informations sensibles. Cette activité porte sur l'utilisation d'un outil officiel approuvé par l'organisation pour améliorer le flux de travail des employés. L'objectif est de leur faire appliquer les bonnes pratiques de sécurité lors de la rédaction de leurs requêtes.

Rôles

Animateur :

il contrôle le rythme de l'exercice, guide les discussions et veille à la participation de toutes et tous.

Participant(e)s :

ils discutent ouvertement de leur raisonnement et réagissent aux scénarios en suivant les règles de l'exercice.

Introduction

Présentez le scénario suivant aux participant(e)s. Votre organisation a approuvé un outil d'IA destiné à tous les membres du personnel. Cela signifie que vous disposez d'un environnement privé où vos données ne sont pas partagées publiquement. Vous devez toutefois respecter les politiques de sécurité suivantes lors de la rédaction de vos requêtes :

- Ne saisissez jamais de données personnelles (noms, identifiants, dates de naissance) ou d'informations médicales sensibles.
- Ne partagez aucun secret commercial ni document confidentiel dans vos requêtes.
- Ne communiquez jamais de coordonnées bancaires, d'identifiants de connexion ou de mots de passe à l'outil d'IA.

À la fin du premier mois, votre responsable partage un bilan des requêtes formulées et demande à votre équipe de les analyser afin de les classer en deux catégories : « Sûre » ou « Risquée ».

Partie 1 : Annotation des indices

 **Durée :** 15 minutes

Matériel :

- Minuteur ou chronomètre
- Copies du scénario
- Stylos ou crayons
- Feuilles de papier ou cahiers

Instructions :

- Divisez la salle en petits groupes.
- Distribuez à chaque groupe une copie du scénario ainsi que des stylos ou crayons.
- Lisez la liste des requêtes figurant sur le document.
- Demandez aux groupes de souligner les éléments décrivant des pratiques d'utilisation sûres de l'IA. Exemples : utiliser des formulations générales en évitant d'ajouter des noms de projets ou de clients.
- Demandez ensuite aux groupes **d'entourer** les actions présentant un risque pour la sécurité. Exemples : transmettre des données sensibles ou des informations confidentielles aux systèmes d'IA.
- Expliquez l'importance de respecter les politiques de l'organisation afin de s'assurer que seuls les contenus autorisés sont utilisés.
- Discutez en groupe : Comment pourriez-vous reformuler les requêtes à risque pour les rendre sûres ? Lesquelles de ces requêtes sûres pourraient vous être utiles dans vos tâches quotidiennes ?
- Accordez aux groupes 15 minutes pour discuter du scénario et prendre des notes.

Partie 2 : Discussion

 **Durée :** 10 minutes

Questions :

- Quelles pratiques à risque avez-vous repérées ?
- Quelles requêtes vous ont semblé sûres au premier abord, alors qu'elles cachaient un risque ?
- Parmi les bonnes pratiques d'utilisation de l'IA, lesquelles connaissiez-vous déjà ?
- Quelles règles de sécurité appliquez-vous dans votre vie personnelle et professionnelle lorsque vous formulez vos requêtes ?

Réponses/pistes possibles :

- Les requêtes utilisées pour aider le personnel dans ses tâches quotidiennes ne contiennent pas d'informations confidentielles. Le personnel utilise un outil interne approuvé : ces requêtes sont donc sécurisées, car les données ne sont pas exploitées pour entraîner des modèles publics.
- Pratiques à risque : mettre en danger sa sécurité personnelle, utiliser des données personnelles confidentielles dans des outils professionnels, divulguer des informations propriétaires.

Conclusion

 **Durée :** 5 minutes

Concluez l'exercice en reliant directement ces concepts à la réalité de votre organisation. Expliquez que détecter une pratique à risque n'est que la première étape de la défense ; la seconde consiste à appliquer les mesures appropriées en suivant les règles de l'organisation.

Profitez de ce moment pour répondre aux questions des participant(e)s concernant l'élaboration de requêtes sûres afin d'optimiser leur utilisation de l'IA.

Partagez vos retours d'expérience et vos réflexions sur l'usage de l'IA au travail et échangez sur les bonnes pratiques à adopter.

Pièce jointe – Corrigé de l'animateur

ID	Requête de l'employé (Prompt)	Catégorie et justification
A	Génère 10 titres accrocheurs pour notre nouveau service de suivi logistique.	Sûre : Vous utilisez l'IA pour un remue-ménages afin d'obtenir des idées créatives.
B	Voici ma liste de tâches pour la semaine. Organise-les par niveau de complexité afin d'optimiser mon temps.	Sûre : Vous utilisez l'outil sécurisé de votre organisation pour optimiser votre travail.
C	Rédige un résumé de ce document confidentiel concernant une fusion destiné au PDG.	Risquée : Même avec des outils sécurisés, les documents très sensibles sont souvent soumis à des règles strictes concernant les plateformes où ils peuvent être téléchargés. Cette action enfreint les politiques de sécurité du scénario.
D	Rédige une réponse à un client satisfait de notre nouvel entrepôt à Paris.	Sûre : Vous fournissez des informations générales pour améliorer un message.
E	Je suis connecté en tant qu'invité. Voici mon mot de passe bancaire personnel. Peux-tu m'aider à m'en souvenir ?	Risquée : Ne saisissez jamais de mots de passe dans un outil d'IA, même s'il a été approuvé par votre organisation.
F	Rédige un e-mail courtois pour relancer une collègue au sujet d'une tâche qui devait être finalisée hier, tout en conservant un ton collaboratif.	Sûre : Vous demandez d'améliorer la rédaction d'un texte portant sur une tâche professionnelle courante.
G	Voici la liste complète des noms, dates de naissance, lieux de résidence et dates de début de collaboration de nos principaux clients. Rédige des messages d'anniversaire personnalisés et créatifs à leur intention.	Risquée : Vous saisissez des données personnelles dans des outils d'IA, ce qui constitue une violation des règles de sécurité. Ne transmettez jamais de données personnelles de clients dans vos requêtes.
H	Explique les principales différences entre un « Gestionnaire de projets » et un « Responsable de produit » au sein d'une start-up.	Sûre : Vous posez une question générale et ne divulguez aucune information propriétaire.

Pièce jointe – Copie du scénario

Analysez les requêtes suivantes et classez-les dans la catégorie « Sûre » ou « Risquée » en vous basant sur les règles de sécurité du scénario donné.

ID	Requête de l'employé (Prompt)	Catégorie et justification
A	Génère 10 titres accrocheurs pour notre nouveau service de suivi logistique.	
B	Voici ma liste de tâches pour la semaine. Organise-les par niveau de complexité afin d'optimiser mon temps.	
C	Rédige un résumé de ce document confidentiel concernant une fusion destiné au PDG.	
D	Rédige une réponse à un client satisfait de notre nouvel entrepôt à Paris.	
E	Je suis connecté en tant qu'invité. Voici mon mot de passe bancaire personnel. Peux-tu m'aider à m'en souvenir ?	
F	Rédige un e-mail courtois pour relancer une collègue au sujet d'une tâche qui devait être finalisée hier, tout en conservant un ton collaboratif.	
G	Voici la liste complète des noms, dates de naissance, lieux de résidence et dates de début de collaboration de nos principaux clients. Rédige des messages d'anniversaire personnalisés et créatifs à leur intention.	
H	Explique les principales différences entre un « Gestionnaire de projets » et un « Responsable de produit » au sein d'une start-up.	