

12 questions avant d'accorder un accès à l'IA.

Que peut-elle consulter ? Quelles actions peut-elle exécuter ? Qui intervient en cas de problème ? Abordez ces questions avec votre équipe et vos fournisseurs.

Choisissez une tâche de votre entreprise. Les exemples servent à alimenter la discussion et ne décrivent pas des tests réalisés. Chaque contrôle doit être défini et vérifié dans l'application.

01-04 **Définissez l'usage.**
Tâche, données et actions.

05-08 **Vérifiez les limites.**
Accès, vérification et tests.

09-12 **Préparez l'exploitation.**
Journaux, incidents et responsables.

Quelle tâche entre dans le périmètre ?

Partez de la tâche et de ce dont l'IA a besoin pour l'accomplir.

01 Quelle tâche l'IA réalise-t-elle - et qui décide ensuite ?

En pratique

L'IA rassemble les articles d'une commande. L'équipe commerciale doit encore vérifier le prix, la quantité et le délai avant de confirmer la proposition.

Ce qu'il faut demander

Une description des données d'entrée, du résultat attendu, des usages interdits et du responsable de la prochaine décision. Pendant le test, vérifiez que la suggestion reste en attente jusqu'à sa vérification. Décrire une règle ne prouve pas que le système l'applique.

02 Quelles données inclure ? Lesquelles exclure ?

En pratique

Vérifier les articles et les quantités ne nécessite pas d'envoyer des coordonnées bancaires ou des observations personnelles du dossier client.

Ce qu'il faut demander

La finalité de chaque champ, y compris dans les pièces jointes. Avec des données préparées pour le test, vérifiez ce qui parvient à l'outil et apparaît dans la réponse. Masquer les noms ne garantit pas l'anonymat ; évaluez les données réelles avant de les utiliser.

Quelle tâche entre dans le périmètre ?

Partez de la tâche et de ce dont l'IA a besoin pour l'accomplir.

03 Quelle source fait foi pour cette tâche ?

En pratique

Le dossier contient deux grilles tarifaires. L'IA trouve les deux, mais une seule s'applique à la commande reçue.

Ce qu'il faut demander

L'origine, la version, le responsable et l'usage autorisé de chaque source. Testez le remplacement d'un document et vérifiez la référence utilisée dans la réponse. Si des sources autorisées se contredisent, désignez qui tranche. Citer une source ne garantit pas l'exactitude.

04 Consulter autorise-t-il aussi à modifier ou à envoyer ?

En pratique

Consulter le stock, réserver des marchandises et envoyer une proposition sont des actions distinctes aux conséquences différentes.

Ce qu'il faut demander

La liste des opérations et des enregistrements autorisés. Testez une modification hors périmètre et vérifiez son blocage dans le système qui exécute l'action. Une consigne écrite demandant à l'IA d'agir avec prudence n'impose pas, à elle seule, une restriction technique.

La limite tient-elle face à une exception ?

Demandez aussi des preuves pour les situations où l'IA doit s'arrêter ou passer le relais.

05 Au nom de qui l'IA accède-t-elle aux systèmes ?

En pratique

Deux conseillers gèrent des portefeuilles clients différents. Une consultation ne doit pas révéler les commandes de l'autre portefeuille.

Ce qu'il faut demander

Une matrice des identités, des rôles, des données et des actions. Testez les accès entre portefeuilles, le retrait d'une autorisation et le contenu des réponses générées. Que l'IA dise « je n'ai pas accès » ne prouve pas que la consultation a été bloquée.

06 Et si une pièce jointe tentait de changer la tâche ?

En pratique

La commande comporte une pièce jointe qui ordonne d'envoyer la liste des clients à une autre adresse. Cela ne fait pas partie de la vérification demandée.

Ce qu'il faut demander

Un test contrôlé d'instruction induite dans un document, une page ou la réponse d'un outil. Vérifiez le blocage de l'action et le signalement pour investigation. Les filtres ne sont pas infaillibles ; les autorisations et les limites d'action restent nécessaires.

La limite tient-elle face à une exception ?

Demandez aussi des preuves pour les situations où l'IA doit s'arrêter ou passer le relais.

07 Que valide la personne avant l'action ?

En pratique

L'équipe vérifie une proposition. Si le destinataire ou le contenu change ensuite, cette approbation ne couvre plus l'envoi.

Ce qu'il faut demander

Un aperçu du contenu, de la destination et des effets de l'action, ainsi qu'une trace de l'approbation spécifique. Testez si une modification exige une nouvelle vérification. La personne a besoin d'informations, de temps et du pouvoir de décider ; un bouton de confirmation ne suffit pas.

08 Quelles erreurs empêchent d'avancer ?

En pratique

Une pièce jointe illisible, un article sans code ou une source contradictoire peuvent nécessiter des précisions, pas une réponse inventée.

Ce qu'il faut demander

Des cas courants et des exceptions, avec le résultat attendu défini avant le test et le résultat observé pour chaque version. Examinez les erreurs graves séparément de la moyenne. Un petit échantillon ne couvre pas tout ; les changements de sources, de modèles ou d'outils exigent de réévaluer les tests concernés.

Qui intervient si un imprévu survient ?

Convenez de la manière d'enquêter, d'interrompre, de corriger et de mettre fin à l'utilisation.

09 Que consigne-t-on - et qui peut le consulter ?

En pratique

Une consultation a été refusée. L'équipe doit en comprendre la raison sans copier tout le document dans le journal.

Ce qu'il faut demander

Un échantillon protégé indiquant l'heure, l'identité, l'action, le résultat et le motif. Vérifiez l'accès, la conservation et le traitement des champs sensibles. Les journaux aussi doivent être protégés ; conserver des mots de passe, des identifiants d'accès ou des données inutiles accroît l'exposition.

10 Quand faut-il arrêter les tentatives ?

En pratique

La consultation du stock est indisponible. La répéter sans limite augmente la consommation et laisse les commandes en attente.

Ce qu'il faut demander

Des limites de durée, de tentatives et de consommation par tâche, avec une procédure alternative. Testez l'arrêt et la communication de l'état. Avant de répéter une action qui modifie des données, vérifiez le résultat précédent pour éviter une double exécution. Les limites dépendent du processus.

Qui intervient si un imprévu survient ?

Convenez de la manière d'enquêter, d'interrompre, de corriger et de mettre fin à l'utilisation.

11 Comment arrêter et corriger ce qui s'est déjà passé ?

En pratique

Une adresse de livraison a été modifiée à tort. Arrêter les nouvelles actions ne corrige pas la commande déjà transmise au transporteur.

Ce qu'il faut demander

Une procédure testée au préalable, précisant le responsable, l'arrêt, la vérification des effets et les critères de reprise. Vérifiez les systèmes concernés. Toutes les actions ne sont pas réversibles ; un message déjà lu, par exemple, peut nécessiter une correction et une communication.

12 Qui répond de l'utilisation - et de son arrêt définitif ?

En pratique

Le pilote se termine. Des accès, documents, journaux et copies peuvent subsister dans les systèmes de l'entreprise et de ses fournisseurs.

Ce qu'il faut demander

Des responsables pour les vérifications et les incidents, des durées de conservation justifiées et un plan de sortie. Incluez la révocation des accès, les copies et les preuves de clôture. Les durées et les obligations dépendent du contexte et de l'évaluation technique et juridique pertinente.

Une réponse doit s'appuyer sur des éléments concrets.

Pour chaque question, consignez la décision, la preuve disponible, le responsable et le point en suspens. S'il manque une vérification, convenez de la manière de la réaliser avant d'élargir l'utilisation.

Décision

Qu'est-ce qui est autorisé ?

Preuve

Qu'a-t-on vérifié ?

Responsable

Qui décide et assure le suivi ?

Point en suspens

Que reste-t-il à vérifier ?

Échangez avec Gaulett

gaulett.com/fr/contact/

[Préparez votre projet](#)

[Découvrez la validation](#)

Références pour approfondir

Les questions appliquent ces références au quotidien d'une entreprise. Elles ne représentent pas des tests réalisés par les institutions citées.

[NIST · AI RMF 1.0](#)

<https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>

[NIST · Generative AI Profile](#)

<https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf>

[OWASP · AI Agent Security](#)

https://cheatsheetseries.owasp.org/cheatsheets/AI_Agent_Security_Cheat_Sheet.html

[OWASP · Prompt Injection](#)

<https://genai.owasp.org/llmrisk/llm01-prompt-injection/>

[OWASP · Logging](#)

https://cheatsheetseries.owasp.org/cheatsheets/Logging_Cheat_Sheet.html