

Édition Future.

Concevoir un agent IA contrôlable

Mission, outils, permissions et scénarios d'essai.



GUIDE PRATIQUE

EXEMPLES ET EXERCICES CORRIGÉS

Votre parcours

Lisez les repères, puis réalisez les exercices avec vos propres données non confidentielles.

Avant de commencer

1. Workflow ou agent ?
2. Définir le contrat de mission
3. Concevoir les outils
4. Sources et instructions malveillantes
5. Atelier : répondre sans inventer
6. Tester les limites, pas seulement le cas facile
7. Définir une sortie vérifiable
8. Reprise, surveillance et livraison

Source

Avant de commencer

Ce guide vous aide à concevoir un assistant qui prépare une réponse à une demande client.

L'exercice utilise des données fictives et des outils simulés. Il ne crée pas une connexion réelle à une messagerie et n'envoie aucun message. Vous pourrez ensuite transposer le contrat dans un outil adapté, après avoir vérifié ses capacités et ses permissions.

Le résultat attendu est un dossier clair : mission, données, outils, règles de décision, scénarios d'essai et procédure de reprise. Un long prompt qui se dit « autonome » ne suffit pas à constituer un système fiable.

1. Workflow ou agent ?

Dans un workflow, les étapes sont principalement prévues à l'avance. Dans un système agentique, le modèle peut choisir certaines étapes ou certains outils selon ce qu'il observe. Cette distinction de conception est décrite notamment par Anthropic ; elle n'est pas une promesse de performance.

[Construire des agents efficaces.](#)

Workflow pour Nora : recevoir une demande, vérifier que l'adresse existe, lire le type de besoin, préparer un brouillon. Si ces étapes suffisent, il n'est pas nécessaire d'ajouter davantage d'autonomie.

Agent possible : après avoir lu la demande, décider s'il faut consulter la politique de retour, les caractéristiques d'un produit ou demander une précision. Ce choix dynamique peut être utile, mais augmente les possibilités d'erreur et la difficulté de contrôle.

Commencez par la plus petite autonomie nécessaire. Un agent qui prépare une proposition sans l'envoyer peut déjà faire gagner du temps. Son utilité ne dépend pas du nombre d'outils auxquels il a accès.

2. Définir le contrat de mission

Une mission exploitable nomme le résultat, les données permises, les actions autorisées et les conditions d'arrêt. Évitez « Fais tout ce qu'il faut », qui ne précise ni coût ni conséquence acceptable.

Contrat fictif : préparer une réponse française à une question sur un carnet. Utiliser uniquement la fiche produit et la politique commerciale fournies. Ne pas modifier une commande, proposer de remboursement ou envoyer un message. Si une condition manque, préparer une question de clarification. Terminer par un brouillon et les faits sur lesquels il s'appuie.

Définissez une limite de temps, d'appels d'outils et de coût adaptée à votre environnement. Ces limites doivent être appliquées par le système qui exécute les actions, pas seulement écrites dans le prompt. Le modèle peut oublier une consigne ; le programme doit refuser une action interdite.