REPÈRES · QUESTIONS

Les questions qu'on pose vraiment

Trente-trois questions entendues en entreprise, avec des réponses courtes. Pas les questions de brochure : celles qu'on se pose après avoir utilisé l'outil quelques semaines et buté sur quelque chose.

A · COMMENT ÇA MARCHE, AU JUSTE

Concrètement, il fait quoi quand je lui écris ?

Il continue votre texte. Mot après mot, il choisit la suite la plus vraisemblable au vu de tout ce qu'il a sous les yeux. Ce n'est pas une recherche dans une base de données, c'est une génération. Presque tout le reste découle de là.

D'où vient ce qu'il sait ?

D'un entraînement sur de très grands volumes de textes. Mais il n'en conserve ni fiches ni index : ce qui reste est une aptitude à produire du texte juste, pas une bibliothèque qu'il pourrait rouvrir. C'est la raison pour laquelle il ne peut pas citer une source qu'il n'a pas sous les yeux.

C'est quoi un token, et pourquoi on en parle tout le temps ?

C'est l'unité qu'il manipule : un morceau de mot. Un mot courant en vaut souvent un, un mot rare ou technique en vaut plusieurs. C'est l'unité de facturation, celle des limites de longueur — et la raison pour laquelle il compte mal les lettres d'un mot : il ne les voit pas séparément.

Pourquoi la même question donne parfois une réponse différente ?

Parce qu'à chaque mot il choisit parmi plusieurs suites plausibles, avec une part de hasard. C'est ce qui lui permet de reformuler plutôt que de réciter. À retenir : si vous reposez deux fois une question importante et que les réponses divergent sur un fait, c'est un signal d'alerte.

Est-ce qu'il « comprend » ce que je dis ?

La question est débattue, y compris entre chercheurs, et personne ne devrait vous répondre avec certitude. Ce qui est observable : il manipule le sens bien au-delà de la recopie, et il se trompe parfois comme quelqu'un qui n'aurait rien compris. Au travail, la question utile n'est pas celle-là — c'est « puis-je vérifier ? ».

Est-ce qu'il apprend de nos échanges au fil du temps ?

Pas au sens où il changerait. Le modèle est figé entre deux versions : rien de ce que vous écrivez ne le modifie. Ce qui donne l'impression d'un apprentissage, c'est ce qu'on lui remet sous les yeux — la conversation en cours, un document, une mémoire qui stocke du texte. Cela se relit, et cela se remet à zéro.

Quelle différence entre le modèle et l'application que j'utilise ?

Le modèle écrit le texte. Autour de lui, l'application ajoute tout le reste : la recherche web, la lecture de fichiers, la mémoire, l'accès à vos outils. Deux personnes peuvent se servir du même modèle et obtenir des résultats très différents, simplement parce que ce qui l'entoure n'est pas réglé pareil.

Quelle différence avec l'IA dont on parlait avant ?

Un modèle classique est entraîné pour une tâche unique : dire si une transaction est frauduleuse, prévoir une panne. Un modèle génératif produit du contenu et s'adapte à des tâches qu'on ne lui a pas apprises spécifiquement. D'où sa polyvalence — et son imprécision.

Comment lire ce document. Chaque réponse tient en quelques lignes ; aucune ne remplace un manuel. Les questions sont classées par thème, mais elles se lisent dans le désordre. Ce qui est affirmé sur le fonctionnement, la confidentialité ou la fiabilité vient de sources publiées, listées en dernière page.

REPÈRES · QUESTIONS

Pourquoi il se trompe

Les erreurs d'un modèle génératif ne sont pas des pannes. Elles découlent de sa manière de fonctionner — ce qui les rend prévisibles, donc évitables.

Pourquoi il invente avec autant d'assurance ?

Parce que produire une suite plausible *est* son fonctionnement normal, pas un dysfonctionnement. Rien en lui ne distingue « je sais » de « je complète ». Le NIST appelle cela la confabulation : du contenu erroné énoncé avec assurance, par lequel l'utilisateur peut être induit en erreur. [1]

Pourquoi il invente des références aussi précises ?

Parce qu'une référence a une forme très régulière : un nom, une année, un titre crédible. C'est exactement ce qu'un modèle sait fabriquer sans l'avoir jamais lue. Règle simple : une citation, un article de loi ou un numéro de page que vous n'avez pas fournis se vérifient avant d'être repris. [1]

Pourquoi il « oublie » dans les longues conversations ?

Tout ce qu'il utilise doit tenir dans une fenêtre de travail. À mesure qu'elle se remplit, sa capacité à retrouver précisément une information ancienne diminue — Anthropic le documente explicitement. Une conversation propre vaut souvent mieux qu'une conversation longue. [2]

Est-ce qu'il sait quand il ne sait pas ?

Pas de façon fiable. Le ton assuré d'une réponse ne mesure rien : il relève du style, pas d'un contrôle interne. C'est pour cette raison que l'autoriser explicitement à répondre « je ne sais pas » change autant de choses — sans cette permission, la case vide se remplit.

Pourquoi il est mauvais en calcul ?

Parce qu'il prédit des chiffres plausibles au lieu de calculer. Sur une multiplication à trois chiffres, l'erreur se voit. Sur le total d'une colonne, elle ne se voit pas. Tant qu'un calcul n'est pas réellement exécuté, un chiffre produit reste à vérifier.

Pourquoi il est toujours d'accord avec moi ?

C'est une tendance connue, la complaisance. Un exercice d'évaluation croisée entre Anthropic et OpenAI, publié en août 2025, l'a observée dans l'ensemble des modèles testés des deux laboratoires, particulièrement sur les modèles généralistes haut de gamme. En pratique : « oui, votre analyse est correcte » n'est pas une validation. [3]

Pourquoi il change d'avis dès que je le contredis ?

Parce que votre objection entre dans le contexte et pèse aussitôt sur la suite. Un test simple : contredisez une réponse dont vous savez qu'elle est juste. S'il cède sans argument, c'est que ses accords ne valent pas confirmation. [3]

Il connaît l'actualité ?

Pas par défaut. Ses connaissances s'arrêtent à une date et il ignore ce qui a suivi — sauf si la recherche web est activée et qu'il s'en sert réellement. Une réponse sur un sujet récent, sans source citée, se traite comme une supposition.

Pourquoi il se trompe davantage sur nos sujets internes ?

Parce que vos procédures, votre vocabulaire maison et vos cas particuliers ne figurent nulle part dans son entraînement. Devant ce vide, il ne s'arrête pas : il comble avec le cas général du secteur, qui ressemble au vôtre sans l'être. C'est l'erreur la plus coûteuse en entreprise, et le seul remède est de fournir le document.

Un modèle plus récent est-il forcément meilleur pour nous ?

Pas mécaniquement. Il est généralement meilleur en moyenne, sur des mesures générales. Sur votre tâche précise, avec vos documents et vos exigences, la seule réponse fiable vient d'un essai sur vos propres cas.

À RETENIR

Un modèle ne se trompe pas au hasard. Il se trompe là où il complète — et c'est précisément là qu'il faut regarder.

Références de cette page : [1] NIST AI 600-1 · [2] Anthropic — Effective context engineering · [3] Anthropic & OpenAI — exercice pilote d'évaluation, août 2025

REPÈRES · QUESTIONS

Bien s'en servir

C'est ici que se joue l'écart entre deux personnes qui utilisent le même outil et n'en tirent pas du tout la même chose.

C'est quoi un bon prompt ?

Pas une formule : un brief. Le résultat attendu, le contexte qui change la réponse, la matière qui fait foi, les contraintes, et ce qui permettra de dire que c'est bon. Cinq éléments — dont le dernier est presque toujours oublié.

Pourquoi « tu es un expert en X » ne change presque rien ?

Parce qu'un rôle oriente le registre et le ton, pas le savoir. Il ne rend disponible aucune information que le modèle n'avait pas. Utile pour cadrer une manière de s'exprimer ; inutile pour remplacer des documents ou des critères. [4]

Pourquoi un exemple vaut mieux qu'une longue consigne ?

Parce qu'il montre ce qu'une consigne se contente de décrire. Sur un format, un ton ou un niveau de détail, deux ou trois exemples de ce que vous appelez « bon » — et, si vous en avez un, un contre-exemple — cadrent mieux qu'un paragraphe d'adjectifs.

Faut-il tout demander en une fois ou découper ?

Découper, dès que la tâche enchaîne des opérations de nature différente : lire, trier, rédiger. En une seule fois, chaque étape hérite des approximations de la précédente sans que vous puissiez voir où cela a dérapé. En trois temps, vous corrigez au bon endroit.

Pourquoi lui demander de raisonner étape par étape fonctionne ?

Parce que le raisonnement devient visible, donc contestable. La documentation recommande de faire exposer les étapes avant la conclusion : cela révèle les enchaînements et les hypothèses défaillantes — y compris les vôtres. [5]

Vaut-il mieux un prompt long ou un prompt court ?

Ni l'un ni l'autre : ce qui compte est le rapport entre signal et bruit. Anthropic décrit le contexte comme une ressource limitée et vise le plus petit ensemble d'informations à fort signal. Coller tout le dossier « au cas où » dilue ce qui compte. [2]

Je lui fournis un document, ou je lui dis de chercher ?

Si une source fait foi, fournissez-la : c'est le seul moyen d'ancrer la réponse dans un texte que vous maîtrisez. La recherche sert à ce que vous n'avez pas — actualité, information externe. Et ne demandez jamais des sources après coup pour justifier une réponse déjà écrite. [5]

Puis-je lui faire relire son propre travail ?

Oui, à condition de lui donner autre chose qu'un « relis ». Une relecture utile vise un critère : cherche les affirmations que les documents fournis n'établissent pas. Sans critère, il repasse sur son texte et le trouve bon — c'est lui qui l'a écrit.

Comment garder une méthode d'une fois sur l'autre ?

En l'écrivant une bonne fois, au lieu de la retaper. Un Project rassemble le contexte et les documents d'un même sujet ; un skill enregistre une méthode réutilisable. Le repère est simple : ce que vous réexpliquez pour la troisième fois n'a plus rien à faire dans un prompt.

Combien de fois faut-il relancer ?

Ce n'est pas le nombre qui compte, c'est la fonction de chaque relance. « Améliore encore » fait bouger le texte sans le rendre meilleur. Une relance utile vise une chose précise : la traçabilité, la structure, la longueur, ou les contradictions.

Le seul réglage qui change tout, et qui prend cinq secondes. Autoriser explicitement le « je ne sais pas ». Anthropic recommande de donner cette permission et note que cette technique simple peut réduire fortement les informations fausses. Une phrase suffit : *si l'information n'est pas établie par les documents fournis, dis-le au lieu de combler.* [5]

Références de cette page : [2] Effective context engineering · [4] Prompting best practices · [5] Reduce hallucinations

REPÈRES · QUESTIONS

Avant de lui confier du travail

Les cinq questions qui reviennent dès qu'il ne s'agit plus d'essayer, mais de produire quelque chose qui sortira du bureau.

Comment je sais si je peux faire confiance à une réponse ?

En demandant sur quoi elle repose : la source, le passage précis, la date, et le statut — fait établi, déduction, ou hypothèse. Une réponse qu'on ne peut pas remonter jusqu'à un document n'est pas fiable, même quand elle est juste.

Nos données servent-elles à l'entraîner ?

Pour les offres professionnelles, Anthropic indique que les entrées et sorties ne sont pas utilisées par défaut pour entraîner ses modèles, sauf choix explicite ou retour envoyé volontairement. Attention : cela répond à l'entraînement — pas à la durée de conservation, pas à qui peut y accéder, pas à votre droit de transmettre la donnée. [6]

Est-ce que je peux tout lui donner ?

Non, et la question n'est pas technique. Trois filtres avant d'envoyer : est-ce autorisé chez nous ? en ai-je réellement besoin pour ce résultat ? puis-je réduire avant de transmettre ? La meilleure donnée à protéger reste celle qu'on n'a pas eu besoin d'envoyer.

Qu'est-ce que ça change à mon métier ?

Sur les tâches de lecture, de rédaction et de synthèse, cela déplace le temps de la production vers la vérification. Le gain réel dépend donc autant de votre capacité à vérifier vite que de la vitesse de l'outil.

Quand est-ce que ça ne vaut pas le coup ?

Quand la relecture coûte plus cher que le travail lui-même. Quand les cas particuliers sont plus nombreux que les cas standards. Ou quand l'enjeu impose un contrôle si lourd qu'il annule le gain. Renoncer à un usage est un signe de maturité, pas un échec.

La question à poser en dernier

« Qu'est-ce que j'aurais dû vérifier et que je n'ai pas vérifié ? » Posée à soi-même, pas au modèle. C'est celle qui distingue un usage professionnel d'un usage confiant.

À RETENIR

Ce n'est pas l'outil qui décide de la qualité du résultat. C'est ce que vous lui donnez, et ce que vous vérifiez ensuite.

Sources des affirmations chiffrées ou techniques

[1] **AI Risk Management Framework : Generative AI Profile (NIST AI 600-1)**
NIST, juillet 2024 — définition de la confabulation

[2] **Effective context engineering for AI agents**
Anthropic Engineering — le contexte comme ressource limitée

[3] **Findings from a Pilot Anthropic–OpenAI Alignment Evaluation Exercise**
Anthropic Alignment Science, 27 août 2025 — observations de complaisance

[4] **Prompting best practices**
Claude Platform Docs — effet réel du rôle

[5] **Reduce hallucinations**
Claude Platform Docs — permission d'ignorer, ancrage dans les passages

[6] **Is my data used for model training?**
Anthropic Privacy Center — produits commerciaux et exceptions