

## IA, pourquoi tu hallucines ?

Je voudrais savoir comment gérer une conversation, ou un chat, quand l'IA "se dérègle" c'est à dire perd le fil de l'entretien, commence à divaguer voire halluciner comme cela arrive quelquefois.

Bien sûr, clore la session est une manière radicale de se débarrasser du problème, mais il y a-t-il une façon de procéder pour remettre l'IA sur le bon chemin afin de continuer ensemble notre route ?

Réponse de l'IA :

C'est une excellente question. En effet, fermer la session (ouvrir un nouveau chat) est le réflexe le plus simple, mais ce n'est pas toujours souhaitable, surtout si l'on a construit un long contexte ou un travail complexe.

Quand une IA « déraile » ou « hallucine », c'est souvent parce que sa **fenêtre de contexte** est saturée, qu'elle a « glissé » sur une association d'idées erronée, ou que le niveau de bruit (les tokens précédents) l'empêche de voir clair. Voici une boîte à outils pratique pour **recadrer** l'IA et la remettre sur les rails sans perdre le fil de votre conversation :

**Le recadrage factuel (Le "Pattern Interrupt")** Ne dites pas simplement "Tu te trompes" ou "C'est faux", car l'IA risque de s'enfermer dans sa logique (par complaisance ou biais de confirmation). Soyez très direct, factuel et autoritaire (dans le bon sens du terme). \* **À faire** : \*\* "Stop. Reprenons depuis le début. Tu fais fausse route sur [tel point]. La réalité est [tel fait]. Oublie tes trois dernières réponses." Pourquoi ça marche Utiliser des mots comme "Stop", "Oublie" ou "Reprenons" agit comme un signal fort dans le flux textuel.

**Le résumé de resynchronisation** L'IA fonctionne avec ce qui est immédiatement sous ses "yeux" (le texte récent). Si elle est perdue, remettez le contexte en perspective en lui dictant ce qui a été fait. \* **À faire** : \*\* "faisons le point. Jusqu'à présent, nous avons validé A et B. Notre objectif actuel est C. Tu es parti sur une tangent en parlant de X, ce qui n'a rien à voir. Revenons à C : comment fait-on ?" Pourquoi ça marche : Vous lui ré insérez un contexte propre et synthétique, ce qui réduit le bruit ambiant du chat.

**Le changement de rôle ou de perspective (Meta-prompting)** Prenez de la hauteur et adressez-vous à l'IA non plus comme à un exécutant, mais en cassant le personnage ou le ton en cours. \* **À faire** : \*\* "Mettons les choses au clair : tu t'es mis à divaguer. Sortons un instant de notre tâche. Analyse pourquoi tu viens de me donner une réponse incohérente, puis repropose une solution valable pour [votre problème]." Pourquoi ça marche : Cela force l'algorithme à basculer dans un mode d'analyse critique de son propre processus (quand le modèle en est capable).

**La technique de la « Table rase » partielle** Si vous ne voulez pas fermer le chat, demandez explicitement à l'IA d'effacer sa mémoire récente (symboliquement). **À faire** : "Considère que tout ce qu'on s'est dit depuis [tel message] est nul et non avenu. On repart sur les bases du message [tel numéro]." Ce qu'il faut **éviter** de faire (qui empire les choses) : Argumenter longuement : Plus vous écrivez de longs messages pour lui prouver qu'elle a tort, plus vous alimentez le contexte avec de la confusion. Restez court et percutant. Être trop polonais ou ambigu : "Euh, je crois que tu te trompes un peu, tu es sûr que c'est ça ?" L'IA risque de répondre "Oui, vous avez raison, mais en fait non..." et continuer à divaguer. Soyez affirmatif. La limite à connaître : Parfois, le dérèglement est dû à une **corruption de la mémoire du chat** (le modèle s'est

enfoncé dans une boucle de tokens aberrants dont il ne peut mathématiquement plus sortir). Si après 2 recadrages fermes, l'IA continue de halluciner ou tourne en rond, c'est qu'il est effectivement trop tard : **\*\*le coût en temps pour la recadrer est supérieur à celui de rouvrir un nouveau chat\*\*** en lui copiant-collant simplement les 2 ou 3 éléments clés de votre progression.