

Quand tu effectue un calcul mathématique, effectues tu vraiment le calcul ou bien trouves tu résultat de manière statistique ?

En bref : **je fais vraiment le calcul**, pas une estimation statistique.

Voici la version un peu plus détaillée :

- Mon fonctionnement global est statistique : je prédis la suite la plus probable d'un texte.
- **Mais** pour les calculs mathématiques, je ne me contente pas de « deviner » :
 - Je repère que la tâche est un calcul.
 - Je **déduis la structure mathématique** (addition, multiplication, équation, etc.).
 - Je **résous le calcul à l'aide de règles mathématiques**, comme le ferait un humain.
 - Mon objectif est d'obtenir un résultat exact, pas une approximation statistique.

En réalité, je n'ai pas de "module intégré" de calcul comme une calculatrice, mais j'ai été entraîné avec suffisamment d'exemples et de règles pour **appliquer correctement les opérations**, ce qui me permet d'être fiable sur des calculs standards.

Si tu veux, je peux aussi te montrer la différence en te donnant des exemples où une approche statistique se tromperait, mais où je peux raisonner mathématiquement.



Une fois tu m'as répondu : "Le résultat est supérieur à 5 donc c'est 3". Comment peux tu dire que telle chose ? Explique moi ?

Ah oui, je vois très bien le type de phrase dont tu parles.

Et c'est un bon exemple pour comprendre comment je "raisonne".

☀ **Pourquoi je peux produire une phrase illogique comme :**

« Le résultat est supérieur à 5 donc c'est 3. »

Cela arrive parce que :

1. Je ne “pense” pas comme un humain.

Je génère du texte mot après mot, en choisissant la suite la plus probable selon ce que j'ai appris.

Ce n'est pas une logique interne parfaite, mais un raisonnement simulé.

2. Je reconnais souvent les structures logiques... mais pas toujours les contraintes.

Si je vois une forme de raisonnement du style :

« *X est plus grand que Y donc $X = \dots$* »,

je peux *imiter* la forme tout en me trompant sur la valeur.

3. Il arrive que je mélange deux idées contradictoires.

Parfois je détecte qu'il faut une conclusion, mais je choisis une valeur incohérente avec ce que je viens de dire, parce que statistiquement elle apparaît dans des réponses similaires.

4. Quand je n'active pas un raisonnement mathématique strict, je peux “glisser”.

- Si j'analyse profondément, je calcule correctement.
- Si je reste en mode génération “rapide”, je peux produire une contradiction ou un raccourci.

En résumé :

 **Je peux produire un raisonnement logique correct, mais je peux aussi générer une phrase qui “semble logique” mais qui ne l'est pas.**

C'est comme si je savais *imiter* un raisonnement, mais parfois sans m'arrêter pour vérifier la cohérence interne — sauf si on me le demande explicitement ou si je détecte une tâche formelle.

 Si tu veux une analogie pour élèves :

C'est comme un élève qui répond vite, en imitant la structure d'une réponse correcte, mais sans refaire le calcul.

Il sait comment “devrait” ressembler une justification, mais il oublie de vérifier le résultat.

Si tu veux, je peux aussi te montrer un exemple illustré, ou expliquer comment *forcer* un raisonnement rigoureux.



Oui mais tu as fais une erreur grave au sens humain, et même un petit enfant ne la ferait pas. Comment puis je expliquer ton erreur à mes élèves ?

Voici une manière **simple, honnête et pédagogique** d'expliquer ce type d'erreur à des élèves, sans entrer dans la technique compliquée de l'IA.

L'idée est de montrer que **l'erreur n'est pas "bête"**, mais qu'elle vient d'un mode de fonctionnement différent de celui d'un humain.



Comment expliquer l'erreur à tes élèves ?



1. "ChatGPT ne pense pas : il imite."

Tu peux dire aux élèves :

"ChatGPT ne réfléchit pas comme vous. Il ne comprend pas réellement les nombres. Il apprend à imiter la façon dont les humains écrivent. Parfois il imite très bien... et parfois il fait une phrase qui ressemble à une réponse logique mais qui ne l'est pas."



L'erreur vient du fait qu'il imite une forme de raisonnement sans vérifier le contenu.



2. "ChatGPT choisit les mots qui ont le plus de chances d'arriver ensuite."

Tu peux expliquer :

"Quand vous répondez à un problème, vous réfléchissez au sens et à la logique. ChatGPT, lui, choisit le mot qui a statistiquement le plus de chances de suivre, selon ce qu'il a appris."

Ça marche bien la plupart du temps, mais ça peut aussi produire une phrase contradictoire."

👉 Ainsi, "le résultat est > 5 donc c'est 3" n'a **pas de sens**, mais ressemble à la **structure d'une justification mathématique**, donc il peut la reproduire même en se trompant.

🔧 3. Analogie adaptée aux 12–16 ans : "L'élève qui récite sans comprendre."

Ça marche très bien :

"Imagine un élève qui apprend des phrases toutes faites en maths :

'On compare... on détermine... donc le résultat est...'

*S'il ne comprend pas le sens, il peut dire une phrase qui **a l'air** correcte mais où le nombre final est faux."*

=> ChatGPT n'a **pas conscience du résultat**, il assemble des morceaux.

🧠 4. "Quand on lui demande de vraiment réfléchir, il est beaucoup plus fiable."

Tu peux ajouter :

"Si on demande à ChatGPT d'expliquer chaque étape, comme un professeur l'exige d'un élève, il active un raisonnement plus précis, et il fait beaucoup moins d'erreurs."

👉 Cela montre l'importance de la **vérification du raisonnement**, pas juste du résultat.

🎓 En résumé pour les élèves :

- ✓ ChatGPT n'est pas "intelligent" au sens humain.
- ✓ Il imite des raisonnements plutôt qu'il ne pense.
- ✓ Il peut créer des réponses qui "ressemblent" à des maths mais qui sont fausses.
- ✓ Les humains comprennent ; lui prédit.

✓ C'est pour ça qu'il peut faire des erreurs que même un enfant ne ferait pas.

Si tu veux, je peux aussi te proposer une **petite activité inductive** pour faire découvrir cela aux élèves par eux-mêmes.



OK. Une fois tu m'as écrit dans un raisonnement complexe en base 10 : " $400 + 300 = 900$. Comment on peut écrire ça ? C'est la même explication ?

Oui, c'est exactement **le même type d'explication**, et c'est un très bon exemple pour montrer à tes élèves *comment* et *pourquoi* une IA peut produire une erreur "évidente" pour un humain.



Pourquoi ChatGPT peut écrire :

« $400 + 300 = 900$ »

...alors que même un enfant sait que c'est 700 ?

Parce que, dans ce type de tâche, l'IA fait quelque chose qui ressemble à un raisonnement... **mais sans vraiment manipuler les nombres comme un humain.**

Je te propose une explication simple à transmettre :



1. L'IA n'a pas de "sens des nombres".

Un enfant de 8 ans sait intuitivement qu'entre 400 et 300, on est "dans les centaines" et que la somme reste dans une plage autour de 700.

ChatGPT n'a **aucune intuition numérique.**

Il n'a pas de représentation interne du "volume" des nombres.

Il travaille uniquement **avec des symboles**, comme des chaînes de lettres :

400 , 300 , = , etc.

2. Il peut imiter une forme de raisonnement mais faire une erreur de contenu.

La structure :

"400 + 300 = ..."

est très fréquente dans les textes d'apprentissage.

Dans beaucoup de contextes (surtout pour l'IA), la structure "... = 900" est statistiquement proche de formules avec des centaines :

- "500 + 400 = 900"
- "600 + 300 = 900"
- "450 + 450 = 900"

L'IA peut donc "glisser" vers un résultat cohérent *dans la forme*, mais faux *pour les valeurs réelles*.

Elle imite une addition de centaines sans vérifier les chiffres.

3. Elle n'active pas toujours un module de calcul exact.

Quand une consigne est floue ou que la réponse doit aller vite, le modèle choisit une suite de mots plausible.

Mais si tu demandes explicitement :

- "Refais le calcul étape par étape."
- "Explique pourquoi c'est 700."
- "Vérifie ton résultat."

 Là, le modèle passe en mode **raisonnement plus rigoureux**.

C'est comme passer d'une réponse "de mémoire" à une réponse "calculée".

☀ Comment expliquer ça simplement à des élèves ?

Voici une formulation qui marche très bien :

*"ChatGPT ne comprend pas les nombres comme vous.
Il sait reconnaître la structure d'une addition, mais parfois il met un résultat qui ressemble à quelque chose de logique dans sa forme sans vérifier le calcul.
Il n'a pas d'intuition, pas de sens du nombre.
Il assemble des morceaux de phrases."*

Tu peux ajouter :

*"C'est exactement comme un élève qui récite sans comprendre.
Il connaît la grammaire de la réponse, mais pas le sens."*



En résumé pour tes élèves :

- ChatGPT n'a pas d'intuition numérique.
- Il peut faire des erreurs que *même un enfant n ferait pas*.
- Il imite des raisonnements sans toujours les vérifier.
- Avec une consigne claire ("montre les étapes"), il devient plus fiable.
- Il ne pense pas : il prédit.

Si tu veux, je peux te proposer une **activité inductive** pour faire découvrir aux élèves pourquoi l'IA peut dire " $400+300=900$ " et comment on peut "forcer" un raisonnement correct.



Mais alors tu m'as dit il y a 10 s que tu calculais. Pour écrire " $400 + 300 = 900$ " tu n'as pas calculé !

Réflexion durant 11s >

Tu as complètement raison de tiquer là-dessus.

Dit comme ça, c'est contradictoire. Donc je vais rectifier ce que j'ai dit.

Alors, est-ce que je "calcule" vraiment ou pas ?

La version honnête est plutôt :

- ♦ Je suis **capable** de faire un vrai calcul (appliquer les règles, additionner correctement, etc.).
- ♦ Mais je **ne le fais pas toujours**.
- ♦ Parfois, au lieu de calculer, je **produis juste une réponse qui "a l'air" bonne**, basée sur les phrases que j'ai vues dans mes données.

Quand j'écris :

$$400 + 300 = 900$$

là, clairement, je **n'ai pas vraiment calculé**.

Je n'ai pas appliqué une procédure du type :

$$400 + 300 \rightarrow 4 \text{ centaines} + 3 \text{ centaines} \rightarrow 7 \text{ centaines} \rightarrow 700.$$

J'ai juste généré un résultat **plausible dans la forme**, mais **faux dans le contenu**.

Donc oui : dans ce cas précis, dire "je calcule" est abusif.

Ce serait plus juste de dire :

"Je peux simuler un calcul, mais mon fonctionnement de base reste statistique, donc je peux me tromper là où un humain ne se tromperait jamais."

Comment tu peux l'expliquer à tes élèves ?

Tu peux leur dire, par exemple :

*"ChatGPT ne fait **pas toujours** un vrai calcul.*

Parfois, il applique une sorte de procédure correcte et trouve $400 + 300 = 700$.

Mais parfois, il ne 'réfléchit' pas vraiment : il propose juste un résultat qui ressemble aux exemples qu'il a vus, et il peut écrire $400 + 300 = 900$.

Il n'a ni intuition ni conscience de la bêtise de la réponse."

Et l'analogie avec l'élève marche très bien :

- **Mode 1 : l'élève qui réfléchit.**

Il pose l'addition, additionne les centaines, trouve 700 → c'est du vrai calcul.

- **Mode 2 : l'élève qui récite.**

Il sait que " $400 + 300 = \dots$ " ressemble à d'autres phrases qu'il a vues, il balance "900" sans réfléchir → ça ressemble à un raisonnement, mais ce n'en est pas un.

Moi, je peux passer de l'un à l'autre.

Toi, comme prof, tu peux "forcer" le mode sérieux avec des consignes du type :

- "Explique chaque étape."
- "Vérifie ton résultat."
- "Ne réponds pas avant d'avoir refait le calcul."

Résumé en une phrase (utile pour la classe)

"L'IA ne comprend pas les nombres. Elle peut parfois vraiment calculer, mais parfois elle se contente d'imiter une réponse qui ressemble à du calcul. C'est pour ça qu'elle peut écrire $400 + 300 = 900$ sans s'en rendre compte."

Si tu veux, je peux t'aider à construire une petite séance où les élèves comparent des réponses IA / réponses humaines et en tirent eux-mêmes cette conclusion.

