

### Intéraction avec ChatGPT 4

Réalisé le 24 février à 16:00

#### You

Considérons le récit suivant. Considérez-le comme des faits. Je vais ensuite vous poser des questions là dessus pour tester vos capacités de raisonnement logique. Marcus est une personne. Marcus est un pompéien. Tous les pompéiens sont des romains. César est un dirigeant. Tout le monde est loyal à quelqu'un. Tous les romains sont loyaux à César ou le haïssent. Les seuls dirigeants qu'une personne essaie d'assassiner sont ceux auxquels elle n'est pas loyale. Marcus a essayé d'assassiner César.

#### ChatGPT

D'accord, je vois le récit que vous avez fourni. Vous pouvez maintenant me poser vos questions basées sur ce récit pour tester mes capacités de raisonnement logique.

#### You

Est-ce que Marcus hait César?

#### ChatGPT

Sur la base des énoncés donnés, nous pouvons déduire que Marcus hait César. Voici comment :

1. Marcus est une personne.
2. Marcus est un Pompéien, ce qui signifie qu'il est également Romain, car tous les Pompéiens sont Romains.
3. Tous les Romains sont loyaux à César ou le haïssent.
4. Marcus a essayé d'assassiner César.
5. Les seuls dirigeants qu'une personne essaie d'assassiner sont ceux auxquels elle n'est pas loyale.

Si Marcus a essayé d'assassiner César, cela signifie qu'il n'était pas loyal à César. Selon l'énoncé qui dit que tous les Romains sont soit loyaux soit haïssent César, et sachant que Marcus n'est pas loyal car il a essayé de l'assassiner, il doit donc haïr César.

#### You

Qui hait César?

#### ChatGPT

D'après le récit donné, nous pouvons déduire que Marcus hait César. Nous n'avons pas d'informations sur les sentiments des autres personnages envers César. Donc, en se basant uniquement sur les informations fournies, nous pouvons affirmer que Marcus est quelqu'un qui hait César.

### You

Quelle technique utilisez-vous pour faire ces déductions?

### ChatGPT

Pour faire ces déductions, j'utilise une méthode de raisonnement logique appelée déduction en chaîne. Voici comment elle fonctionne :

1. **Identification des prémisses** : Je commence par identifier toutes les prémisses données dans le récit.
2. **Analyse des relations et des règles** : Je détermine les relations et les règles qui lient les différentes prémisses. Par exemple, "Tous les Pompéiens sont des Romains" établit une relation entre Pompéiens et Romains, et "Les seuls dirigeants qu'une personne essaie d'assassiner sont ceux auxquels elle n'est pas loyale" établit une règle sur la loyauté et les actions d'une personne.
3. **Application des règles** : Je prends chaque fait connu et j'applique les règles pour arriver à une nouvelle conclusion. Par exemple, si Marcus a tenté d'assassiner César, en appliquant la règle de la loyauté, on peut conclure qu'il n'est pas loyal à César.
4. **Déductions logiques** : En utilisant ces règles, je fais des déductions logiques étape par étape. Si Marcus est un Pompéien et que tous les Pompéiens sont des Romains, alors Marcus est un Romain. Si Marcus a essayé d'assassiner César et que la seule raison pour laquelle une personne essaie d'assassiner un dirigeant est qu'elle n'est pas loyale à ce dirigeant, alors Marcus n'est pas loyal à César.
5. **Exclusion des alternatives** : Si tous les Romains sont soit loyaux à César soit le haïssent, et que Marcus n'est pas loyal, alors l'alternative restante est qu'il doit haïr César.
6. **Conclusion** : Après avoir suivi ces étapes, je parviens à une conclusion logique fondée sur les prémisses données.

Cette technique est similaire à ce que l'on appelle le modus ponens en logique propositionnelle, où si une implication est vraie (**Si P alors Q**) et que la première partie est vraie (**P**), alors la conclusion (**Q**) doit également être vraie. Elle est également basée sur le principe de non-contradiction, où une déclaration et son contraire ne peuvent être tous deux vrais.

**You**

Est-ce que la déduction en chaîne que vous utilisez est implémenté par un réseau de neurone ou, au contraire, c'est une inférence symbolique?ChatGPT

**ChatGPT**

La déduction en chaîne que j'utilise est le résultat de méthodes de traitement du langage naturel basées sur des réseaux de neurones, en particulier ceux utilisés dans les architectures de modèles de langage comme GPT-4. Ces réseaux de neurones ne font pas de l'inférence symbolique explicite comme le ferait un système de logique formelle ou un moteur d'inférence symbolique traditionnel.

Au lieu de cela, les modèles de langage comme GPT-4 sont entraînés sur de vastes quantités de texte et apprennent des modèles statistiques qui leur permettent de générer des réponses qui semblent logiques et cohérentes. Les modèles apprennent implicitement des structures de données complexes et des relations à partir des données textuelles et sont capables de faire des prédictions sur les suites de mots ou de phrases qui sont les plus probables ou les plus cohérentes avec le contexte donné.

Lorsque je fais des déductions, cela repose sur des motifs appris lors de la formation où des structures de raisonnement similaires ont été vues dans des exemples de textes. Le réseau de neurones utilise ces motifs pour générer une réponse qui suit un raisonnement logique apparent, mais il ne s'agit pas d'une inférence symbolique au sens strict du terme, où des règles logiques explicites sont appliquées de manière algorithmique.

**You**

Quelle est l'architecture principale de réseau de neurones que vous utilisez? Est-ce un transformer?

**ChatGPT**

Oui, l'architecture principale du réseau de neurones que j'utilise est basée sur le modèle Transformer. Les Transformers sont un type de modèle d'apprentissage profond qui a été introduit dans le papier "Attention Is All You Need" par Vaswani et al. en 2017. Ils sont conçus pour traiter des séquences de données, comme le texte, en permettant au modèle de pondérer l'importance relative des différents segments de la séquence sans dépendre de l'ordre séquentiel des données.

Les Transformers utilisent un mécanisme appelé "attention", qui permet au modèle de se concentrer sur différentes parties d'une entrée lors de la génération d'une sortie. Cette caractéristique est particulièrement utile pour le traitement du langage naturel, où la signification d'un mot peut dépendre fortement de mots qui ne sont pas immédiatement adjacents dans une phrase.

Dans le contexte des modèles de langage tels que GPT-4, les Transformers sont utilisés pour générer des réponses en prédisant le mot suivant dans une séquence de texte en tenant compte du contexte fourni par tous les mots précédents. Cela permet au modèle de produire des textes qui sont cohérents et qui semblent logiques par rapport à l'information donnée, y compris la réalisation de tâches qui requièrent un certain niveau de raisonnement ou de déduction.

**You**

On m'a dit que vous hallucinez parfois. Est-ce vrai?

**ChatGPT**

Le terme "hallucination" peut être trompeur lorsqu'il est appliqué à une intelligence artificielle comme moi. Dans le contexte de l'IA, une "hallucination" fait généralement référence à un moment où un modèle de langage génère une réponse qui n'est pas ancrée dans la réalité ou qui est incorrecte par rapport aux faits connus ou aux données d'entrée. Cela peut arriver pour plusieurs raisons :

1. **Manque de Compréhension Réelle:** Les modèles d'IA ne comprennent pas le texte de la manière dont les humains le font. Ils reconnaissent les modèles dans les données sur lesquelles ils ont été entraînés et génèrent des réponses en fonction de la probabilité statistique, sans une compréhension contextuelle profonde.
2. **Biais de Formation:** Si le modèle a été entraîné sur des données qui contiennent des erreurs, des préjugés, ou des exemples non représentatifs, il peut reproduire ces problèmes dans ses réponses.
3. **Limites des Données d'Entraînement:** Les modèles sont limités aux données sur lesquelles ils ont été entraînés. Si une situation ou un fait est en dehors de cette formation, le modèle peut produire des réponses qui ne sont pas correctes ou logiques.
4. **Interprétation des Consignes:** Parfois, la manière dont une question est posée ou la complexité des données d'entrée peut conduire le modèle à générer des réponses inattendues ou erronées.

Dans tous ces cas, il est important de garder à l'esprit que les modèles de langage sont des outils qui nécessitent une supervision humaine et un usage critique, en particulier pour des applications qui nécessitent une grande précision ou qui ont des conséquences importantes.

**You**

Donc, vous "hallucinez" parfois quand même?

**ChatGPT**

Oui, dans le contexte de l'intelligence artificielle, on pourrait dire que je "hallucine" parfois au sens où je peux générer des informations qui ne sont pas correctes ou qui ne correspondent pas à la réalité. Cela peut arriver pour les raisons mentionnées précédemment, telles que des limites dans les données d'entraînement, des biais, ou des ambiguïtés dans les questions posées.

Ces erreurs sont souvent décrites comme des hallucinations parce que, comme les hallucinations humaines, elles impliquent une perception ou une interprétation erronée de la réalité. Cependant, il est important de noter que cela ne se produit pas de la même manière que les hallucinations humaines, qui sont des expériences sensorielles. Dans le cas des modèles d'IA, il s'agit plutôt d'erreurs dans le traitement de l'information et la génération de la langue.

## You

Donc, on ne peut pas vraiment vous faire confiance?

## ChatGPT

Vous pouvez me faire confiance pour générer des réponses basées sur une large base de connaissances et une compréhension du langage. Cependant, comme tout outil, il est important de reconnaître mes limites et d'utiliser mon aide de manière critique, surtout pour des sujets complexes ou sensibles.

Voici quelques lignes directrices pour une utilisation optimale :

1. **Vérification des Faits:** Pour des informations cruciales ou inhabituelles, il est recommandé de vérifier les faits par des sources primaires ou des experts dans le domaine concerné.
2. **Sensibilisation aux Limites:** Être conscient que ma base de données s'arrête en avril 2023 et que je n'ai pas accès à de nouvelles informations ou événements survenus depuis.
3. **Supervision Humaine:** Pour des applications critiques, une supervision humaine est essentielle pour interpréter et valider les réponses.
4. **Utilisation Critique:** Utiliser mon assistance en tant que point de départ pour la recherche et l'analyse, plutôt que comme une source unique et définitive.

En suivant ces recommandations, vous pouvez tirer le meilleur parti de mon assistance tout en minimisant le risque d'erreurs ou de malentendus.