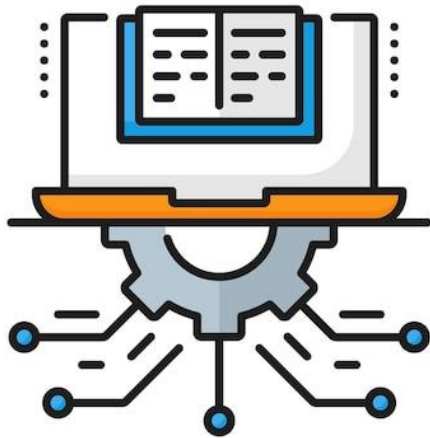
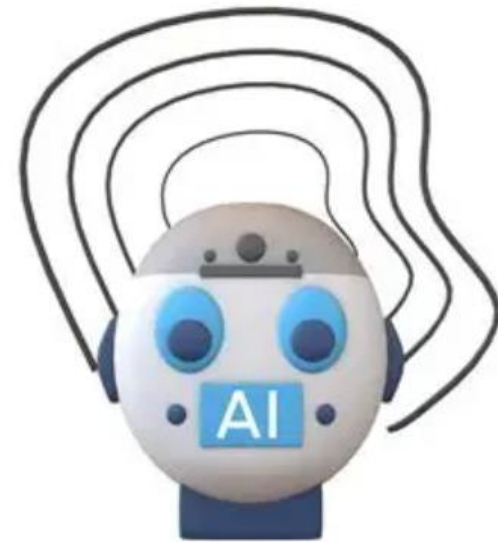


Les données d'entraînement



Hallucination



Hallucination

Désigne une situation dans laquelle une IA générative produit une information fausse, inventée ou non vérifiable, tout en la présentant de manière convaincante.

Ce phénomène est particulièrement fréquent avec les modèles de langage. Ils ne disposent pas nécessairement d'un mécanisme permettant de distinguer systématiquement une information vraie d'une information plausible.

Une hallucination peut prendre différentes formes :

- Inventer une information ou un fait ;
- Attribuer une citation à une personne qui ne l'a jamais prononcée ;
- Créer une référence bibliographique inexistante ;
- Produire une réponse incorrecte à partir d'une information ambiguë.

La fluidité et l'assurance du texte ne garantissent donc pas son exactitude.

Il est essentiel de vérifier les informations produites par l'IA, particulièrement lorsqu'elles sont utilisées dans un contexte académique ou professionnel.

Les données d'entraînement

Sont les données utilisées pour apprendre à un modèle d'IA à réaliser une tâche. Elles peuvent être constituées de textes, d'images, de sons, de vidéos, de données scientifiques ou d'autres types d'informations.

Pendant son entraînement, le modèle analyse ces données pour identifier des régularités, des relations et des structures, qu'il pourra ensuite mobiliser lorsqu'il génère une réponse ou produit un résultat.

La qualité, la diversité et la représentativité de ces données influencent donc directement les capacités et les limites du modèle.

Si certaines populations, langues, situations ou connaissances sont sous-représentées, le modèle peut être moins performant dans ces contextes.

Les données peuvent également contenir des erreurs, des stéréotypes ou des biais, susceptibles de se retrouver dans les résultats produits par l'IA.

Il est donc important de considérer les données d'entraînement comme un élément essentiel pour comprendre les performances, les biais et les limites d'un système d'IA.