

Que signifie ChatGPT

ChatGPT (Chat Generative Pre-trained Transformer) est un modèle d'apprentissage automatique complexe capable d'effectuer des tâches de génération de langage naturel (NLG) avec un niveau de précision si élevé que le modèle peut passer un test de Turing.

ChatGPT a été formé sur des quantités massives de **données non étiquetées** extraites d'Internet avant 2022. Le modèle est continuellement surveillé et affiné pour des tâches spécifiques orientées langage avec des ensembles de données supplémentaires **étiquetés** par des humains.

Les tâches dans lesquelles ChatGPT excelle incluent:

- Fournir des réponses aux questions.
- Compléter un texte ou une phrase donnés.
- Écrire du contenu de fiction et de non-fiction à partir d'invites.
- Produire des réponses de **chatbot** de type humain.
- Génération de code informatique.
- Traduction de texte d'une langue à une autre.
- Effectuer des calculs.
- Résumer un texte donné.
- Classification du texte en différentes catégories.
- Analyse du **sentiment textuel**.
- Génération de texte qui résume les données dans des tableaux et des feuilles de calcul.
- Répondre aux entrées de l'utilisateur de manière conversationnelle.

ChatGPT est construit sur un **grand modèle de langage** (LLM) appelé GPT-3 (Generative Pre-trained Transformer 3).

Un grand modèle de langage est un type de **réseau neuronal profond** qui utilise des milliards de paramètres et est entraîné avec des **pétaoctets** de données.

Le modèle a été développé par OpenAI, une société de recherche fondée par Elon Musk.

Techopedia explique ChatGPT

Techniquement, ChatGPT peut être classé comme un type **d'IA générative**.

Une fois formé, un modèle d'IA générative peut produire des résultats similaires – mais pas les mêmes – que les données d'apprentissage.

Comment ChatGPT a-t-il été formé?

Les scientifiques des données et les ingénieurs en apprentissage automatique ont utilisé **l'apprentissage semi-supervisé** pour former ChatGPT.

Les algorithmes d'apprentissage semi-supervisé sont formés sur des ensembles de données partiellement étiquetés.

Certains points de données ont une étiquette et d'autres non.

Le modèle utilise les données étiquetées pour prédire la sortie des données non étiquetées.

Selon OpenAI, les scientifiques des données ont rassemblé l'énorme quantité de données non étiquetées nécessaires pour former le LLM en [grattant](#) Internet.

Ils ont complété cela par des sources textuelles qui étaient soit dans le domaine public, soit mises à la disposition des chercheurs et éventuellement des gouvernements.

OpenAI n'a pas révélé exactement comment ils ont pu étiqueter les ensembles de données extrêmement volumineux nécessaires pour affiner le modèle, mais on sait qu'ils ont [externalisé une partie de l'étiquetage](#) et il est probable qu'ils ont également tiré parti des plates-formes de crowdsourcing comme [Mechanical Turk](#) d'Amazon.

ChatGPT utilise des invites pour générer des réponses plus précises et pertinentes à un large éventail d'entrées et est continuellement affiné avec des invites utilisateur pertinentes pour une tâche spécifique à accomplir.

Le processus consiste à [donner une invite au modèle](#), puis à lui permettre de générer une réponse.

La [sortie générée](#) est ensuite évaluée par un évaluateur humain, et le modèle est ajusté en fonction de la rétroaction.

Le processus de réglage fin est répété jusqu'à ce que les résultats du modèle soient acceptables.

C'est pourquoi il est important pour les utilisateurs qui interagissent avec ChatGPT de fournir des commentaires sur la qualité de la sortie du modèle.

En évaluant les résultats du modèle et en fournissant des commentaires, les utilisateurs peuvent aider à affiner le modèle et contribuer à son développement et à son amélioration continus.

ChatGPT et l'IA responsable

L'[IA responsable](#) est un cadre de gouvernance destiné à rendre les modèles d'IA complexes de [la boîte noire](#) comme ChatGPT plus [explicables](#) et donc plus fiables. OpenAI prend plusieurs mesures pour s'assurer que ChatGPT prend en charge le concept d'IA responsable.

- Les chercheurs et les développeurs d'OpenAI cherchent des moyens de rendre ce modèle [d'apprentissage profond](#) plus transparent et tentent d'être aussi francs que possible sur les capacités, les limites et les [abus potentiels](#) du modèle.
- Les scientifiques des données OpenAI examinent continuellement les commentaires des utilisateurs et ont mis les humains dans la [boucle \(HITL\)](#) pour éliminer les biais dans leurs [données de formation](#).
- Les ingénieurs d'OpenAI surveillent en permanence les invites utilisateur et les sorties du modèle ChatGPT pour s'assurer qu'il est utilisé de manière responsable.

ChatGPT et l'éducation

La dernière version de ChatGPT produit un texte si humain que les enseignants ont sonné l'alarme sur la façon dont le modèle est [mal utilisé par les étudiants](#).

En réponse, certaines écoles ont carrément interdit ChatGPT.

D'autres testent des outils de plagiat conçus pour [identifier statistiquement](#) les filigranes linguistiques des modèles.

OpenAi, par exemple, offre actuellement à ses membres un outil de plagiat gratuit appelé [Classificateur AI](#) ([Ouvre un nouvel onglet](#)).

Selon le site Web OpenAI, cependant, l'outil fonctionne mieux lorsque le contenu évalué pour l'originalité a au moins 250 mots et est écrit par un adulte qui parle anglais.

Les défenseurs de ChatGPT en classe voient la valeur des outils anti-plagiat, mais maintiennent également que ce type d'[intelligence artificielle](#) (IA) ne va pas disparaître.

Au lieu d'interdire son utilisation, les élèves devraient apprendre [à utiliser l'outil de manière éthique](#) et les éducateurs devraient utiliser ce type d'[intelligence artificielle](#) (IA) pour enseigner la pensée critique.

Utilisations commerciales pour ChatGPT

ChatGPT est utilisé comme un outil d'intelligence augmentée pour une grande variété de tâches commerciales. Il est actuellement utilisé pour :

- Automatisez le processus de création de contenu pour les publications sur les réseaux sociaux, les articles de sites Web et les descriptions de produits.
- Créez des chatbots conversationnels orientés client capables de sortir du script.
- Résumez et catégorisez les commentaires des clients provenant de [plusieurs canaux](#).
- Traduisez les rapports commerciaux et les livrables marketing d'une langue à une autre.
- Écrivez du code informatique standard et [des scripts génériques](#).
- Traduisez les informations contenues dans un tableau ou une feuille de calcul en texte.
- Créez des chatbots qui prennent en charge l'apprentissage juste-à-temps pour les programmes de formation internes.
- Rédigez des méta-descriptions et suggérez des synonymes pour les [mots-clés SEO](#).

Concurrents de ChatGPT

[Google Bard](#) – un service d'IA conversationnel qui fournit des réponses précises et à jour aux questions en agrégeant des informations provenant d'Internet.

Bien que l'on ne sache pas exactement en quoi le service de chatbot de Google sera différent de ChatGPT, les experts de l'industrie spéculent que le différenciateur de BARD sera qu'il peut accéder aux informations actuelles sur Internet, tandis que ChatGPT n'a accès qu'aux informations publiées avant 2022.

Facebook [Blender](#) – un grand modèle de langage (LLM) développé par Facebook AI. Blender est capable d'effectuer une variété de tâches linguistiques, y compris répondre à des questions, générer du texte et résumer de longs documents.

Recherche et mise en page:

Michel Cloutier

CIVBDL

20230710

"C'est ensemble qu'on avance"