

ChatGPT : 5 minutes pour comprendre cette intelligence artificielle qui analyse et écrit (presque) comme nous

Développé par la start-up OpenAI, ce robot conversationnel est capable de simuler des conversations, d'écrire des poèmes ou des chansons. Sa marge de progression est spectaculaire.

Quelques lignes pour un robot mais un grand pas pour l'Humanité. Le projet ChatGPT, de la start-up OpenAI, marque une nouvelle étape dans le développement de l'intelligence artificielle, cette capacité des algorithmes à apprendre. Des travaux scientifiques sont menés sur les robots conversationnels depuis 2017 mais la version 3.5 de cet outil impressionne tant le langage est naturel dans plusieurs langues, dont le Français. Plus d'un million d'utilisateurs ont déjà mis à l'épreuve ce « bot » et l'ont donc entraîné au passage.

Comment fonctionne-t-il ?

ChatGPT (chat with pre-trained language model, en anglais) est un projet de pointe lancée par OpenAI - un institut de recherche en intelligence artificielle créé en 2015 par Elon Musk et Sam Altman - qui a franchi plusieurs étapes. Il repose sur une intelligence artificielle, baptisée GPT3, qui utilise des « transformers », des algorithmes de traitement automatique du langage naturel apparus en 2017. Doté de 175 milliards de paramètres, ce système informatique a appris en étant alimenté pendant des mois avec tous les textes disponibles sur Internet. « Il utilise l'apprentissage automatique, ou machine Learning en anglais, avec le modèle de l'attention où une phrase est comprise dans son ensemble et dans le contexte avec une capacité de traitement de 3 500 mots », analyse Stéphane Dadian, co-fondateur de la jeune pousse Juice et spécialiste de l'IA du traitement du langage naturel.

ChatGPT se définit, lui-même, comme « un type de programme informatique conçu pour simuler une conversation humaine en utilisant un modèle de langage pré-entraîné. Ces programmes utilisent des algorithmes d'apprentissage automatique pour analyser des textes et générer des réponses en utilisant le contexte et les informations précédemment apprises. » « Mais GPT 3 ne savait pas distinguer le vrai du faux, les chercheurs d'OpenAI ont ajouté dans la version 3.5 des couches supplémentaires d'apprentissage par renforcement avec des évaluateurs humains », explique Alexei Grinbaum, directeur de recherche au CEA. « Cela permet de réduire les réponses bizarres, inhumaines voire toxiques mais aussi les insultes ou les contre-vérités », détaille cet expert. « Son temps d'exécution rapide donne l'impression de dialoguer avec la machine alors que tout se passe sur des serveurs surpuissants aux États-Unis », note Thomas Janssoone, spécialiste des agents conversationnels chez la start-up de robotique Enchanted Tools.

Que peut-il créer ?

Depuis sa mise en ligne mercredi dernier, les internautes n'ont cessé de le défier. Chanson dans un style particulier, poème et même code informatique, ChatGPT a tout ingéré et reproduit à la perfection. « Il y a un gros pas en avant qui a été fait sur la qualité des réponses et des tâches accomplies, le tout en multilingues, c'est impressionnant car le modèle d'avant se contentait d'énumérer des mots », juge Stéphane Dadian, co-fondateur de Juice. « C'est le système le plus puissant à ce jour et il est même capable de réaliser des exercices impossibles pour les humains, comme composer un texte complexe ou debugger un programme informatique en quelques secondes », relève Alexei Grinbaum, du CEA. « Cela va poser des problèmes de plagiat pour les devoirs scolaires et dépasser les capacités des créatifs à qui cela prendra des heures », souligne ce physicien. Et d'ajouter que « OpenAI travaille sur un filigrane (« watermark ») qui permettra de détecter les créations de cette IA ».

Quelles sont ses marges de progression ?

La version 3.5 du GPT se révèle bluffante mais comporte encore des limites. « Il est important de noter que je suis un programme informatique et que ma capacité à apprendre et à progresser est limitée par mes algorithmes et mes paramètres d'entraînement », répond d'ailleurs le robot quand on lui pose la question. « Il faut garder à l'esprit qu'il n'est pas connecté à Internet et que ses connaissances s'arrêtent donc à 2021, et que les sources ne sont pas toujours fiables et traçables », tempère Stéphane Dadian.

L'outil ChatGPT inclut d'ailleurs un système de notation avec des pouces, remarque Thomas Janssoone, d'Enchanted Tools, qui développe un robot interactif, « car il dispose encore de biais sur des sujets clivants, comme les migrants, ou sur des sujets plus complexes qui sont liés à son entraînement, mais des équipes de OpenAI surveillent et corrigent cela ».

Interrogé lundi par nos soins sur l'écriture d'un tweet complotiste, ChatGPT avait par exemple répondu : « le gouvernement français cache la vérité sur les attentats du 13 novembre ». A contrario, il se disait ce mercredi incapable de formuler des idées complotistes.

Est-ce la première étape vers un robot proche des humains ?

« La machine reste une machine, même si elle va au-delà du test de Turing, en étant capable de passer pour un humain. Elle conserve un problème de stabilité dans ses réponses », pointe Alexeï Grinbaum du CEA. « Cette IA reste facile à tromper en faisant quelques fautes d'orthographe qu'un humain pourrait repérer », complète Thomas Janssoone. « Elle n'est pas encore proche de la singularité technologique car elle manque de cohérence sur la longueur et dans la compréhension du contexte d'une conversation entre deux humains, c'est le meilleur algorithme du moment mais il ne va pas casser le plafond de verre qu'a atteint l'IA ».

Lien	https://www.leparisien.fr/high-tech/chatgpt-5-minutes-pour-comprendre-cette-intelligence-artificielle-qui-analyse-et-ecrit-presque-comme-nous-06-12-2022-KWH52WTGUJCVTO4NVF7V4GGDEE.php
Auteur	Damien Licata Caruso (Journaliste au journal Le Parisien)
Date	06/12/2022