

Il faut qu'on parle de ChatGPT

OPINION

Depuis déjà dix ans, l'intelligence artificielle vit une période de croissance spectaculaire grâce à des technologies similaires aux réseaux de neurones artificiels des années 1990, qui permettent de concevoir des programmes d'ordinateurs capables d'apprendre, c'est-à-dire de se modifier pour fournir le résultat attendu sur des données de référence. Pour du diagnostic médical automatique par exemple, un tel programme peut traiter des radiographies des poumons pour prédire si des patients souffrent du covid, après avoir été entraînés sur des radiographies pour lesquelles des tests ont permis de connaître le diagnostic correct.

Cet apprentissage automatique peut se faire sur un très grand nombre d'exemples, et il en découle trois qualités remarquables. La première est que, alors que le programme de départ, écrit de manière traditionnelle, est relativement simple, la complexité du programme final, après apprentissage, est immense. La deuxième est que ces systèmes sont capables de résoudre des problèmes qui demandent des critères tellement complexes qu'aucun programmeur humain ne pourrait les concevoir. La troisième enfin, et certainement la plus surprenante, est que le fonctionnement de ces programmes une fois entraînés est pour l'essentiel incompréhensible.

En dépit de cette dernière caractéristique, qui peut être problématique, ces méthodes sont à présent utilisées à grande échelle, en particulier parce qu'elles sont les seules capables d'extraire du sens de signaux naturels complexes: elles peuvent par exemple déterminer les objets et les personnes présents dans une image, les mots dans un enregistrement sonore, ou les sentiments, sujets et actions dans un texte.

C'est ce dernier domaine du traitement de la langue naturelle qui est en ébullition depuis quelques mois car de très gros modèles d'IA, dont le dernier en date est ChatGPT, sont capables de produire du langage naturel complexe, cohérent avec du texte qui leur est donné en entrée. Vous pouvez leur poser une question de culture générale, leur demander d'écrire un conte ou une lettre très formelle à partir d'un résumé de quelques lignes, et ils le feront remarquablement. Bien mieux que la plupart des humains.

Derrière les impressionnantes performances de ces systèmes, il y a avant tout leur immensité. Le plus connu est GPT-3, troisième incarnation du *Generative Pre-trained model* d'OpenAI. Mis en service au début de l'été 2020, il intègre 175 milliards de paramètres optimisés sur des textes de pages web universitaires, de blogs, de livres, et de Wikipédia, constituant un volume total qui équivaut à 750 000 de fois la Bible, ou 167 fois l'intégralité de la version anglaise de Wikipédia. Le coût de la location des machines nécessaires à l'apprentissage de ce modèle n'a pas été communiqué mais est estimé de l'ordre de 5 millions de dollars. Une machine capable de faire fonctionner ce modèle une fois l'apprentissage terminé coûte 200 000 dollars, et la production de 20 pages de texte consomme autant d'énergie que faire bouillir 1 litre d'eau.

Cette mesure est justifiée par ce que Richard Sutton, pionnier de cette IA par apprentissage, appelle la «leçon amère»: l'indiscutable supériorité de systèmes brutaux et gigantesques sur des méthodes concurrentes, conçues par des humains, beaucoup plus sophistiquées, mais incapables de profiter de mêmes quantités de données d'apprentissage.



FRANÇOIS FLEURET
DIRECTEUR DU GROUPE MACHINE LEARNING,
UNIVERSITÉ DE GENÈVE

Ces «modèles de langage» ne savent faire qu'une chose: étant donné un texte interrompu, prédire quel sera le mot qui le continue

Car si ces «modèles de langage» semblent réfléchir, et même parfois disposer d'un «discours intérieur», en réalité ils ne savent faire essentiellement qu'une chose: étant donné un texte interrompu, prédire quel sera le mot qui le continue. Cette tâche au premier abord incroyablement simple peut être itérée pour écrire un texte mot après mot, mais surtout elle demande d'intégrer une très large palette de connaissances. Evidemment, des règles syntaxiques et grammaticales, la synonymie et les registres de langues, mais aussi des connaissances littéraires, historiques, et géographiques («La capitale de la Suisse est...»), physiques («Un corps immergé subit une force...»), psychologiques («Comme elle avait eu une mauvaise note, Sophie était...»), etc. La connaissance est donc là, extraite

des milliards de textes utilisés pendant l'apprentissage qui, étant donné leur volume, sont collectés automatiquement et intègrent le meilleur comme le pire de la culture humaine.

Donc quand une de ces IA produit du texte, elle ne fait que broder, un mot après l'autre, sans objectif prédéfini, sans plan, simplement pour que chaque mot posé respecte une cohérence statistique avec ceux qui le précèdent. Le seul état mental dont cette entité dispose, c'est le texte qu'elle a écrit jusque-là. Elle n'a pas de pensées internes, intimes et cachées. Et si elle écrit souvent des vérités, ce n'est qu'un effet secondaire de son objectif premier d'imitation statistique.

ChatGPT marque une étape supplémentaire car il a été optimisé pour produire des réponses qui, en plus, sont jugées satisfaisantes par des évaluateurs humains. Avec un million d'utilisateurs cinq jours seulement après son lancement, son impact semble considérable. Ses utilisateurs sont aussi variés que des étudiants qui lui font faire leurs devoirs, des abonnés de sites de rencontres qui compensent leur manque de bagout, ou des développeurs de logiciels qui lui font écrire des bouts de programmes, car il sait aussi programmer. Dans mon environnement direct, je vois des gens l'utiliser pour des premiers jets de mails sensibles, et surtout des enseignants à tous niveaux qui l'ont testé et confirmé pour la plupart que les réponses données à des énoncés bruts obtiendraient de bonnes ou très bonnes notes.

Quel est le futur? En ce qui concerne le développement technologique, rien ne suggère de ralentissement. La recherche sur ces méthodes continue à augmenter la quantité de

données d'apprentissage utilisées, et vise à assurer que le discours reste cohérent dans le temps et repose sur une représentation de la réalité partagée avec les utilisateurs. A côté de cette quête de la performance, une partie de la communauté IA cherche à assurer que des principes éthiques sont respectés, que les travers et biais racistes, sexistes ou violents d'une portion importante des données d'apprentissage ne sont pas intégrés dans les systèmes entraînés, et que plus généralement ces systèmes ne pourront pas être utilisés à mauvais escient.

Pour ce qui est de l'impact, il est évidemment difficile d'anticiper exactement les effets économiques et sociétaux que vont avoir de tels systèmes. Au minimum, le résultat sera une automatisation d'un grand nombre de métiers de «cols blancs», qui semblaient à l'abri d'une révolution industrielle: rédaction, édition, analyse, support pédagogique, soutien psychologique, illustration, design graphique, etc.

D'autres conséquences sont beaucoup plus floues. Quelles seront la place et la forme de la communication écrite quand chacun disposera d'un secrétaire capable de rédiger instantanément une lettre longue et nuancée à partir de quelques points clés, et réciproquement de résumer une longue missive? Quelle sera la nature des illusions et du lien émotionnel qui se développera entre des humains et ces IA communicantes? Quelles seront les capacités véritablement créatrices des systèmes d'IA futurs en art et en sciences?

Tout cela ne peut plus être ignoré en prétendant que certains traitements de l'information relèvent d'une sphère inaccessible aux machines, et il serait sage de se préparer. ■

Chatbot GPT: une norme prévisible, ce n'est pas le langage

OPINION



THOMAS ROBERT
ENSEIGNANT D'ANGLAIS, DOCTEUR
EN PHILOSOPHIE

Le but du langage n'est pas nécessairement une communication directe d'un message clair à des fins utiles

Comme si la menace climatique ne suffisait pas, les ingénieurs de la Silicon Valley ont jugé opportun de poursuivre des recherches qui, selon leur aveu même, nous menacent dans notre humanité. L'on pourrait dissenter longtemps sur l'hubris qui pousse des codeurs avec une vision limitée de ce qu'est l'être humain à tenter de développer une intelligence artificielle s'approchant de nous, ou, pire, nous surpassant. Il est désormais essentiel de déconstruire le récit élaboré par les promoteurs de l'intelligence artificielle et relayé sans beaucoup d'esprit critique et avec une sidération certaine par de nombreux médias et individus plus ou moins experts de domaines potentiellement affectés par l'avancée de l'intelligence artificielle.

La question de l'intelligence artificielle s'articule essentiellement autour de problèmes linguistiques. En effet, il s'agit, pour la machine, de réussir à faire illusion au moyen de notre canal de communication de prédilection, à savoir le langage articulé. Or, la définition du langage privilégiée par

les chercheurs en intelligence artificielle est contestable. Ces derniers ont une approche instrumentale du langage articulé. Il s'agirait d'un code permettant de faire passer le plus directement possible un message. Teintée d'une légitimité scientifique à travers les neurosciences, cette définition est en réalité tout à fait idéologique. Il s'agit de réduire le langage et l'expérience humaine à une norme prévisible. C'est en définitive sur ce modèle que sont construits les traducteurs automatiques et, plus récemment, la nouvelle génération de chatbots.

A l'approche instrumentale, objective et normalisée du langage articulé et, par extension, de l'expérience humaine, l'on peut opposer une définition passionnelle, subjective et créatrice. Plus proche de notre expérience quotidienne du langage, une telle approche fait de ce dernier une expression de soi aux contours multiples et imprévisibles dont le but n'est pas nécessairement une communication directe d'un message clair à des fins utiles. C'est à travers cette approche que le langage devient indissociable de notre personnalité et de nos cultures.

L'enjeu actuel autour de l'intelligence artificielle ne consiste pas à convoquer un débat académique sur ce qu'est le langage en mobilisant Rousseau, Darwin, Saussure ou encore Chomsky. Il s'agit désormais de reprendre le contrôle du récit conté par les enthousiastes de l'intelligence artificielle. En effet, le risque est que ces derniers nous convainquent que l'intelligence n'est rien d'autre qu'une compilation du savoir ou que la créativité se réduit à la réponse la plus probable. En d'autres termes, le danger le plus pressant n'est pas les avancées de l'intelligence artificielle, mais le discours les accompagnant et tendant à définir de façon rabougrie ce qu'est l'être humain. Trois siècles plus tard, les codeurs de la Silicon Valley tentent une nouvelle fois la théorie de l'homme-machine, il est de notre choix d'y succomber ou non. ■

Les défis éthiques des générateurs de texte

OPINION

Un débat fait actuellement rage sur les implications éthiques des générateurs de textes et d'images basés sur l'intelligence artificielle, tels que OpenAI et DALL-E. Ces technologies ont le potentiel de révolutionner la façon dont nous créons et consommons du contenu, mais leurs implications pour la société sont profondes et souvent difficiles à prévoir.

D'une part, les générateurs de textes et d'images basés sur l'IA pourraient être utilisés pour créer un contenu plus accessible et plus attrayant que jamais. En automatisant le processus de création de contenu, ces outils pourraient nous permettre de générer du contenu à un rythme beaucoup plus rapide, offrant potentiellement un niveau d'accès sans précédent à l'information et au divertissement. En outre, les générateurs basés sur l'IA pourraient être utilisés pour créer des contenus plus diversifiés et inclusifs, permettant aux personnes de tous horizons de voir leurs points de vue représentés dans les médias.

Toutefois, ces technologies soulèvent un certain nombre de questions éthiques. Par exemple, les générateurs de textes et d'images basés sur l'IA pourraient être utilisés pour créer du contenu trompeur, voire malveillant. Cela pourrait entraîner une augmentation des *fake news* et d'autres types de désinformation, ainsi qu'une augmentation de la cyberintimidation et d'autres formes de harcèlement en ligne. En outre, les générateurs basés

sur l'IA pourraient être utilisés pour créer des contenus qui ne respectent pas l'éthique de certains groupes.

En définitive, il est important de prendre en compte les implications éthiques des générateurs de textes et d'images basés sur l'IA avant de les déployer. Il est essentiel que ces technologies soient utilisées de manière responsable et dans le plus grand respect des droits de l'homme et de la vie privée.

Ces outils pourraient nous permettre de créer du contenu à un rythme beaucoup plus rapide

En outre, il est important de veiller à ce que ces technologies soient utilisées pour créer un contenu diversifié, inclusif et reflétant toutes les perspectives. Ce n'est qu'à cette condition que ces technologies pourront être employées pour créer un contenu réellement bénéfique pour la société. ■

Ce texte a été écrit avec ChatGPT; commande de départ «Op-ed on the ethical challenges of text generators like OpenAI», introduite par le Dr Johann Roduit, fondateur du studio d'innovation Conexkt, à Sion. La traduction a été réalisée par DeepL.

SUR LE WEB

ChatGPT, l'intelligence artificielle qui stupéfie la planète

Le chatbot, ou agent conversationnel, créé par la société OpenAI, est capable de rédiger des poèmes, de répondre à des questions et d'écrire des articles, le tout avec une qualité déconcertante. A lire sur

www.letemps.ch