

ChatGPT

Source

The Conversation
Thierry Poibeau
11 janvier 2023

1. ChatGPT : pourquoi tout le monde en parle ?

La sortie de ChatGPT le 30 novembre dernier a marqué une nouvelle étape dans le développement des technologies de traitement des langues. C'est en tout cas la première fois qu'un système d'IA, directement issu de la recherche, suscite un tel engouement : de nombreux articles sont parus dans la presse spécialisée mais aussi générale. ChatGPT (ou OpenAI, la société qui développe ChatGPT, ou les deux) sont régulièrement en top tendance sur Twitter aujourd'hui encore.

Pourquoi un tel écho ? Les capacités de ChatGPT marquent-elles un tournant par rapport aux précédentes technologies capables de générer des textes ?

ChatGPT peut dialoguer

Clairement, ChatGPT a permis un saut qualitatif : il est capable de répondre, de manière souvent pertinente, à presque n'importe quelle requête en langage naturel. GPT2, puis GPT3 – les précédents modèles de langage mis au point par OpenAI – étaient déjà très forts pour générer du texte à partir de quelques mots ou phrases donnés en amorce (ce que l'on appelle le « prompt ») : on parlait souvent de « perroquets » (*stochastic parrots*), comme l'a proposé Emily Bender, et de nombreux autres chercheurs critiques de ces techniques. En effet, on pouvait dire que ces systèmes produisaient du texte à partir de bouts de phrases tirées des immenses corpus à leur disposition... même s'il faut nuancer cette image du perroquet : les systèmes ne répètent pas des extraits mot à mot, mais produisent en général un texte original en reprenant, paraphrasant, combinant des éléments vus dans des corpus variés.

ChatGPT reprend ces caractéristiques, mais la partie « chat » (dialogue) y ajoute une dimension tout à fait différente, et en apparence souvent bluffante.

Le système ne produit plus juste un paragraphe à partir d'une phrase donnée en entrée. Il est maintenant capable de répondre correctement à des questions simples comme à des questions complexes, il peut fournir des éléments de raisonnement, s'exprimer dans différentes langues, analyser ou produire du code informatique, entre autres.

Par exemple, si on lui demande si deux personnages ont pu se rencontrer, le système est capable de déterminer les dates correspondant à leur existence, comparer ces dates et en déduire une réponse. Ceci est trivial pour un humain, mais aucun système ne pouvait jusque-là répondre de manière aussi précise à ce type de question sans être hautement spécialisé. Au-delà du langage, ChatGPT peut aussi fournir des éléments de raisonnements mathématiques (mais il se trompe souvent) et analyser du code informatique notamment.

De ce point de vue, ChatGPT a une longueur d'avance sur ses concurrents.

ChatGPT, un système fermé

Concernant le fonctionnement du système, difficile d'en dire plus, car OpenAI, malgré son nom, développe des systèmes fermés. C'est-à-dire que le code informatique (code source) utilisé n'est pas disponible et que les recherches liées à ChatGPT restent pour l'instant en grande partie un secret

industriel – même si, évidemment, d'autres équipes travaillent sur des systèmes similaires, aussi à base d'apprentissage profond. On a donc une idée de ce qui est utilisé par OpenAI.

D'autres entreprises ont gardé un modèle plus ouvert, par exemple Meta avec les travaux menés à FAIR, en général ouverts et publiés dans les principales conférences du domaine. Mais, plus généralement, on constate de plus en plus une fermeture des recherches. Par exemple, alors qu'Apple a toujours eu un modèle de développement privé et très secret, Deepmind avait un modèle ouvert et l'est sans doute un peu moins, maintenant que l'entreprise est sous le contrôle de Google.

Le système ChatGPT lui-même pourrait devenir commercial : OpenAI est financé par Microsoft qui pourrait décider de fermer l'accès au système un jour prochain, si c'est son intérêt.

En testant ChatGPT, vous contribuez (gratuitement) à l'améliorer

En attendant, OpenAI bénéficie de l'énorme publicité que lui apporte son outil, et aussi de toutes les interactions des utilisateurs avec lui. Si un utilisateur signale qu'une réponse n'est pas bonne, ou demande à l'outil de reformuler sa réponse en tenant compte d'un élément en plus, c'est autant d'information que le système emmagasine et pourra réutiliser pour affiner sa réponse la prochaine fois, sur la requête posée ou sur une requête similaire. En testant ChatGPT, on travaille gratuitement pour OpenAI !

Pour en revenir au système lui-même, la partie dialogue est donc ce qui fait la force et la particularité de ChatGPT (par rapport à GPT3 par exemple). Il ne s'agit plus d'un « simple » modèle de langage capable de générer du texte « au kilomètre », mais d'un véritable système de dialogue. Celui-ci a probablement bénéficié de millions ou de milliards d'exemples évalués par des humains, et la phase actuelle – où des centaines de milliers d'utilisateurs « jouent » quasi gratuitement avec le système – permet encore de l'améliorer en continu, puisque toutes les interactions sont évidemment enregistrées et exploitées pour cela.

Il est aujourd'hui assez simple d'accéder à des corpus de milliards de mots pour mettre au point un modèle de langage de type « GPT », au moins pour les langues bien répandues sur Internet.

Mais les données ayant permis la mise au point de ChatGPT (dialogues, interactions avec des humains) ne sont, elles, pas publiques, et c'est ce qui donne un avantage important pour OpenAI face à la concurrence.

Par exemple, Google dispose de données différentes, mais sans doute aussi exploitables pour ce type de systèmes – d'autant que Google a développé depuis plusieurs années un graphe de connaissances qui pourrait permettre de générer des réponses avec une meilleure fiabilité que ChatGPT. Notamment, l'analyse des enchaînements de requêtes issus du moteur de recherche de Google pourrait fournir des informations précieuses pour guider l'interaction avec l'utilisateur... Mais, en attendant, c'est OpenAI qui dispose de ChatGPT, et non Google : OpenAI a l'avantage.

ChatGPT a passé les premiers tests

De fait, même s'il est possible de contourner les limites de ChatGPT, le système refuse d'expliquer comment créer une bombe, de produire des contes érotiques ou de donner ses sentiments (ChatGPT répond fréquemment qu'elle est une machine, qu'elle n'a pas de sentiments ni de personnalité). OpenAI a visiblement soigné sa communication. La société a aussi mis un soin extrême à « blinder » le système, à lui permettre de déjouer la plupart des pièges qui peuvent ruiner en quelques heures ce type d'application, comme cela arrive fréquemment pour des systèmes ouverts au grand public.

On peut par exemple citer Meta, qui en novembre dernier a dû retirer son système appelé « Galactica » deux jours après l'avoir mis en ligne. Galactica avait été entraîné sur le domaine scientifique et pensé pour offrir des services aux chercheurs. Il a d'abord été présenté comme pouvant écrire des articles scientifiques automatiquement à partir d'un prompt... avant que la société ne précise qu'il ne s'agissait évidemment que d'une aide à la rédaction. Mais cette stratégie de communication malheureuse a déclenché une polémique qui a obligé Meta à débrancher rapidement Galactica.

À l'inverse, ChatGPT est toujours en ligne, et suscite toujours autant de passion auprès d'un large public. Des questions demeurent cependant : quel impact aura ChatGPT ? Quelles applications en seront dérivées ? Et quel modèle économique la compagnie OpenAI (et Microsoft, son principal investisseur) vise-t-elle ?

Source

The Conversation
 Fabian Suchanek,
 Gaël Varoquaux
 26 décembre 2022

2. Beau parleur comme une IA

Les intelligences artificielles apprennent à parler grâce aux « modèles de langage » : les modèles les plus simples vous proposent par exemple le mot suivant sur votre smartphone.

Mais les prouesses et les progrès des modèles de langage les plus modernes tels que GPT-3, LaMDA, PaLM ou ChatGPT sont époustouflants. Il existe par exemple des programmes informatiques capables d'écrire dans le style d'un poète donné, de simuler des personnes décédées, d'expliquer des blagues, traduire des langues, et même produire et corriger le code informatique, en se basant sur des modèles de neurones de plus en plus complexes.

Ceci aurait été impensable il y a quelques mois à peine et soulève de nombreuses questions. Mais ces modèles de langage sont en réalité plus superficiels que ces exemples nous font croire.

Quand les intelligences artificielles parlent à tort et à travers

Par exemple, nous avons comparé les histoires générées par des modèles de langage à des histoires écrites par des humains et constaté qu'elles étaient moins cohérentes, mais engageantes, et moins surprenantes que les histoires écrites par les humains.

Plus important, nous pouvons montrer que les modèles de langage actuels ont des problèmes même avec des tâches de raisonnement simples. Lorsque nous demandons :

« L'avocat a rendu visite au médecin ; le médecin a-t-il rendu visite à l'avocat ? »

... les modèles de langage simples ont tendance à dire oui. GPT3 répond même que l'avocat n'a pas rendu visite au médecin. Une raison possible que nous sommes en train d'explorer est que ces modèles de langage encodent les positions des mots de manière symétrique, et donc ils ne font pas la distinction entre « avant le verbe » et « après le verbe », ce qui complique la distinction du sujet et de l'objet dans une phrase.

De plus, les limites théoriques des modèles de langage basés sur les « transformateurs » signifient qu'ils ne peuvent pas distinguer les séquences paires et impaires d'un certain élément, si celles-ci sont intercalées avec un autre élément. En pratique, cela signifie que les modèles ne peuvent pas résoudre une tâche que nous appelons la « tâche pizza » – une simple énigme de la forme :

« La lumière est éteinte. J'appuie sur l'interrupteur d'éclairage. Je mange une pizza.
 J'appuie sur l'interrupteur d'éclairage. La lumière est-elle allumée ? »

Ici, une séquence paire d'interrupteurs d'éclairage signifie que la lumière est éteinte, mais un modèle BERT n'arrive pas à l'apprendre. Les modèles les plus puissants actuellement, GPT-3 et chatGPT, refusent catégoriquement de conclure que la lumière est éteinte.

Les modèles de langage d'aujourd'hui ont également des difficultés avec la négation, et réussissent généralement mal les tâches de raisonnement dès que celles-ci sont plus complexes. Par exemple, considérons l'énigme suivante de l'examen national des fonctionnaires de Chine :

« *David connaît l'ami de M. Zhang, Jack, et Jack connaît l'amie de David, Mme Lin.
 Tous ceux qui connaissent Jack ont une maîtrise, et tous ceux qui connaissent Mme
 Lin sont de Shanghai. Qui est de Shanghai et a une maîtrise ?* »

Les modèles actuels ne répondent correctement que dans 45 % des cas et chatGPT refuse de répondre... alors que la meilleure performance humaine est de 96 %.

Le modèle ne sait pas qu'il ne sait pas

Très probablement, le principal inconvénient des modèles de langage actuels est qu'ils sont faits pour faire des analogies, mais pas pour mémoriser exactement un ensemble trop grand. Par exemple, ils ne peuvent pas se souvenir d'une liste d'un million de clients ou d'un ensemble de dix mille protéines. Le modèle, à sa propre discrétion, oubliera certains des articles. Un problème encore plus important est qu'il peut aussi, encore une fois à sa propre discrétion, inventer de nouveaux éléments (un processus appelé à juste titre « hallucination »). Cela est une conséquence fondamentale des

techniques d'apprentissage automatique des modèles de langage, optimisées pour interpoler et généraliser, pas pour mémoriser.

Le problème est qu'il est impossible de dire à partir de la sortie d'un modèle de langage si une réponse a été mémorisée ou inventée. Par exemple, quand on demande à GPT-3 où est mort Alan Turing, le modèle répond impassiblement « Warrington », au lieu de la réponse correcte (Wilmslow). ChatGPT répond correctement à cette question, mais si on l'interroge sur des chercheurs moins connus, comme les auteurs de cet article, il invente des faits. Et quand on lui pose des questions plus compliquées, il fournit des réponses élégantes, mais fausses :

« L'addition de deux nombres impairs est-elle impaire ? »

... le modèle répond poliment :

« Oui, lorsque vous ajoutez deux nombres impairs, le résultat est toujours un nombre impair. Cela est dû au fait qu'un nombre impair peut s'exprimer sous la forme $2n+1$, où n est un entier. Lorsque vous ajoutez deux nombres sous cette forme, le résultat est $(2n+1)+(2m+1) = 2(n+m)+2$, qui est également sous la forme $2p+1$, où p est un entier. Cela signifie que le résultat est un nombre impair. »

Sur la forme, la réponse est très convaincante, mais fausse.

Avec ces exemples, nous souhaitons montrer qu'il reste à l'heure actuelle imprudent de s'appuyer sur un modèle de langage pour raisonner ou prendre des décisions. Les modèles s'améliorent avec le temps, connaissent plus de choses, et savent de plus en plus s'abstenir de répondre lorsqu'ils n'ont pas l'information. Cependant, en dehors des questions simples, un modèle de langage peut facilement inventer une réponse et avec une explication ou une preuve tout aussi inventée et approximative.

D'autres méthodes excellent à raisonner sur des faits exacts

Tout cela ne veut pas dire que les modèles de langage ne seraient pas des outils étonnants aux capacités époustouflantes. Cela ne veut pas non plus dire que les modèles de langage ne pourront jamais surmonter ces défis, ou que d'autres méthodes de *deep learning* ne seront pas développées à cette fin. C'est plutôt dire qu'au moment d'écrire ces lignes, en 2022, les modèles de langage ne sont pas l'outil de choix pour raisonner ou pour stocker des données exactes.

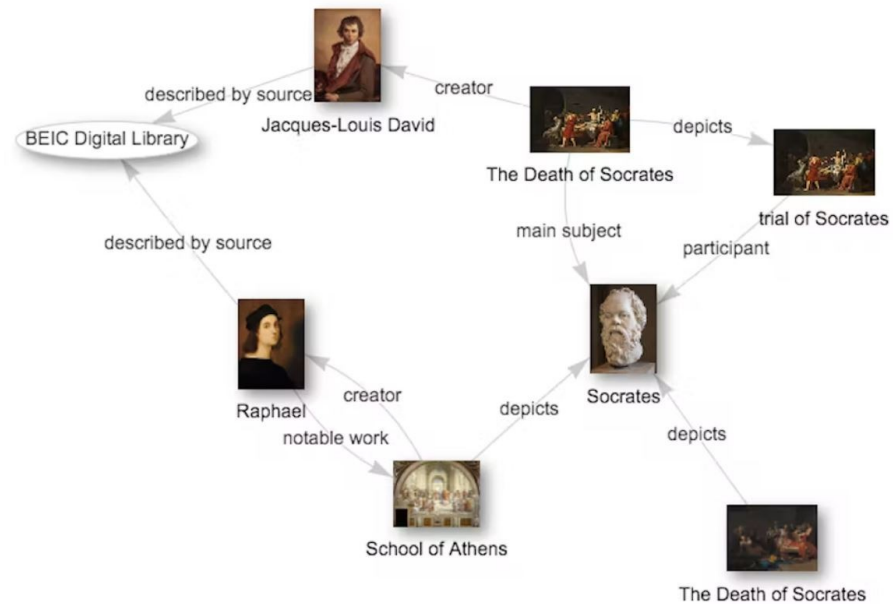
Pour ces fonctions, l'outil de prédilection reste actuellement les « représentations symboliques » : les bases de données, les bases de connaissances et la logique. Ces représentations stockent les données non pas de façon implicite, mais comme des ensembles d'entités (telles que des personnes, des produits commerciaux ou des protéines) et des relations entre ces entités (telles que qui a acheté quoi, ce qui contient quoi, etc.). Des règles logiques ou des contraintes sont ensuite utilisées pour raisonner sur ces relations d'une manière prouvée correcte – bien que généralement sans tenir compte des informations probabilistes. Un tel raisonnement a par exemple été utilisé dès 2011 par l'ordinateur Watson, lors du jeu Jeopardy pour répondre à la question suivante :

« Quel est le roi espagnol dont un portrait, peint par Titien, a été volé avec arme d'un musée argentin en 1987 ? »

En effet, la question peut se traduire par des règles de logique applicables sur une base de connaissance, et uniquement le roi Philip II peut correspondre. Les modèles de langages ne savent actuellement répondre à cette question, probablement parce qu'ils n'arrivent pas à mémoriser et manipuler suffisamment de connaissance (liens entre des entités connues).

Ce n'est sans doute pas un hasard si les mêmes grandes entreprises qui construisent certains des modèles de langage les plus puissants (Google, Facebook, IBM) construisent également certaines des plus grandes bases de connaissances. Ces représentations symboliques sont aujourd'hui souvent construites par l'extraction d'information d'un texte en langage naturel, c'est-à-dire un algorithme essaie de créer une base de connaissances en analysant des articles de presse ou une encyclopédie. Les méthodes qui sont utilisées pour cela sont en l'occurrence les modèles de langage. Dans ce cas, les modèles de langage ne sont pas l'objectif final, mais un moyen de construire les bases de connaissances. Ils sont adaptés pour ça parce qu'ils sont très résistants au bruit, à la fois dans leurs données d'apprentissage et dans leurs entrées. Ils sont donc très bien adaptés pour traiter les entrées

ambigües ou bruyantes, omniprésentes dans le langage humain.



Un exemple très simple de « graphe de connaissance ». Ces objets permettent de connecter des concepts et des entités. Ils sont beaucoup utilisés par les moteurs de recherche et les réseaux sociaux. Fuzheado/Wikidata, CC BY-SA

Les modèles de langage et les représentations symboliques sont complémentaires : les modèles de langage excellent dans l'analyse et la génération de texte en langage naturel. Les méthodes symboliques sont l'outil de choix lorsqu'il s'agit de stocker des éléments exacts et de raisonner dessus. Une analogie avec le cerveau humain peut être instructive : certaines tâches sont suffisamment faciles pour que le cerveau humain les exécute inconsciemment, intuitivement, en quelques millisecondes (lire des mots simples ou à saisir la somme « 2 + 2 ») ; mais des opérations abstraites nécessitent une réflexion laborieuse, consciente et logique (par exemple mémoriser des numéros de téléphone, résoudre des équations ou déterminer le rapport qualité/prix de deux machines à laver).

Daniel Kahneman a dichotomisé ce spectre en « Système 1 » pour le raisonnement subconscient et en « Système 2 » pour le raisonnement avec effort. Avec la technologie actuelle, il semble que les modèles de langage résolvent les problèmes du « Système 1 ». Les représentations symboliques, en revanche, sont adaptées aux problèmes du « Système 2 ». Au moins pour l'instant, il apparaît donc que les deux approches ont leur raison d'être. Qui plus est, tout un spectre entre les deux reste à explorer. Des chercheurs explorent déjà le couplage entre modèles de langage et bases de données et certains voient l'avenir dans la fusion des modèles neuronaux et symboliques en approches « neurosymboliques ».

Source
 letemps.ch
 Anouch
 Seydtaghia
 28 février 2023

3. Comment OpenAI, maître de l'intelligence artificielle, conquiert la planète

L'éditeur du service ChatGPT, dans lequel Microsoft veut injecter 10 milliards de dollars, conserve de nombreuses parts de mystère. La firme basée à San Francisco, un temps liée à Elon Musk, a vite délaissé ses premières idées désintéressées.

Nestlé, lui aussi, n'a pu résister. En 2022, la multinationale avait utilisé les services du générateur d'images DALL-E pour la promotion de ses yoghourts vendus sous la marque «La Laitière». La firme suisse avait employé ce système d'intelligence artificielle pour travailler sur le tableau du même nom de Johannes Vermeer, peint vers 1658, afin de le moderniser pour ses campagnes marketing. Sur YouTube, Nestlé détaille d'ailleurs comment il a employé DALL-E à cette fin.

Qui se trouve derrière ce logiciel ultra-perfectionné ? OpenAI. En l'espace de quelques semaines, la firme californienne est devenue incontournable avec ses systèmes d'intelligence artificielle. Il y a ainsi la génération d'images. Il y a bien sûr aussi, depuis fin 2022, ChatGPT, le générateur de texte qui stupéfie la planète par ses performances. Et bientôt, il y aura la vidéo, avec des séquences entièrement générées par des machines. OpenAI est partout. Ses systèmes d'intelligence artificielle envahissent la planète. Mais l'entreprise, elle, demeure extraordinairement discrète.

Chiffres mystère

C'est tout le paradoxe d'OpenAI. Financée par le géant Microsoft, cofondée par des stars de la tech tel Elon Musk, l'entreprise reste pourtant secrète. Ses débuts sont récents. C'est en 2015 que plusieurs sommités de la Silicon Valley se réunissent pour créer une intelligence artificielle éthique. Il y a l'ingénieur de renom Ilya Sutskever, pionnier des réseaux neuronaux artificiels. A ses côtés se trouvent Sam Altman, dirigeant de la fabrique à start-up Y Combinator, Peter Thiel, directeur de PayPal, Elon Musk, directeur de Tesla. Plusieurs d'entre eux injectent de l'argent dans OpenAI – combien? On ne l'a jamais su – tout comme Amazon Web Services, là aussi pour un montant inconnu.

À ses débuts, OpenAI ne veut pas parler de profits et de modèle d'affaires. Le 11 décembre 2015, l'organisation se présente sur son blog comme « une société de recherche en intelligence artificielle à but non lucratif », avec l'objectif de faire progresser l'intelligence numérique de la manière la plus susceptible de bénéficier à l'humanité dans son ensemble, sans être limité par la nécessité de générer un rendement financier. Comme nos recherches sont exemptes d'obligations financières, nous pouvons mieux nous concentrer sur un impact humain positif. »

Nouvelle stratégie

Mais rapidement, la volonté de passer à un modèle commercial apparaît. En mars 2019, OpenAI change de forme pour devenir une entreprise se disant « à profit limité » (capped-profit). Cela concerne les investisseurs: l'argent injecté lors des premiers tours de financement ne peut pas rapporter davantage qu'un multiple de 100. Le reste des bénéfices est versé à la partie non lucrative de l'entreprise. « Il s'agit d'un concept intéressant qui semble bon sur le papier, analyse Sabine Süssstrunk, directrice du Laboratoire d'images et représentation visuelle de la Faculté informatique et communications de l'EPFL. Cependant, comme le public ne connaît pas tous les accords entre actionnaires et les autres détails financiers, il est difficile de dire quelle part des bénéfices, le cas échéant, est versée à la partie non lucrative de la société. Cette part pourrait être très faible. »

Elon Musk avait déjà pris ses distances avec OpenAI en 2018. Critique de longue date de l'intelligence artificielle, qu'il estime extrêmement dangereuse si elle tombe entre de mauvaises mains, il n'est plus actif dans la société – mais on ignore s'il a revendu ses parts.

Premières images

Très vite, OpenAI enchaîne les coups d'éclat en multipliant la mise en ligne de services basés sur l'intelligence artificielle. Il y a la sortie d'OpenAI Five, un système qui parvient à battre des joueurs chevronnés au jeu en ligne Dota 2. En 2020, GPT-3, ancêtre de ChatGPT, est publié, offrant à une audience de spécialistes un sérieux aperçu de contenus générés par de l'intelligence artificielle. Le grand public ne perçoit pas encore ses possibilités. Mais il commence à être confronté aux images de synthèse générées par le système DALL-E 2, utilisables par tout un chacun.

Traiter des milliards de documents pour générer textes et images est extraordinairement gourmand en capacités de calcul. En mai 2020, Microsoft établit ce qu'il appelle un «superordinateur», destiné à faire tourner les modèles d'intelligence artificielle d'OpenAI. En parallèle, l'éditeur de logiciels injecte 2 milliards de dollars supplémentaires dans sa protégée, dont les besoins en capitaux explosent. On ne sait alors pas quelle est la part de Microsoft dans le capital d'OpenAI, ni celle des autres investisseurs, d'ailleurs.

L'appétit de Microsoft

Cette semaine, Microsoft annonçait l'injection de 10 milliards supplémentaires dans sa pépite, tout en conservant un mystère important sur l'opération. Selon Bloomberg, cet accord se matérialisera avant tout par l'octroi de Microsoft à OpenAI de la puissance de calcul nécessaire pour faire tourner ses services, notamment via son système cloud Azure. Microsoft devrait, en retour, recevoir la moitié des revenus générés par OpenAI jusqu'à une certaine limite.

Aujourd'hui, OpenAI, forte d'environ 375 employés basés à San Francisco, évolue dans deux directions parallèles. D'abord, il y a l'aspiration de ses services les plus précieux par Microsoft. Pour ses propres applications: son moteur de recherche Bing, sa suite Office, son service de collaboration Teams, ses outils de cybersécurité, tous pourraient bénéficier de ChatGPT, avec des gains en productivité potentiellement importants pour les utilisateurs. Ensuite, Microsoft va déployer ces outils d'intelligence artificielle au sein d'Azure, pour ses clients entreprises, afin de rendre son cloud plus attractif – et ainsi mieux concurrencer Amazon Web Services.

Vers un abonnement ?

En parallèle. OpenAI prend une autre direction, celle de l'élargissement de ses propres services. Avec un nouvel outil déjà évoqué, celui de la génération de vidéos. La semaine passée, Sam Altman a évoqué ce projet avec un certain flou. Il a affirmé qu'il « ne voulait pas faire de prédiction précise sur la date » de lancement de ce service, ajoutant que « cela pourrait être très bientôt. C'est un projet de recherche légitime. Cela pourrait prendre un certain temps. » Il va sans dire que le lancement d'un tel service pourrait poser des risques importants, avec la création de fausses vidéos (deepfakes) pouvant servir de funestes intentions.

Et il y a, en toile de fond, la question des revenus. Valorisée selon les rumeurs 29 milliards de dollars, OpenAI aurait réalisé un chiffre d'affaires de 80 millions de dollars en 2022, avec des projections, selon des médias spécialisés, de 200 millions cette année et d'un milliard de dollars en 2024. Pour y parvenir, une version payante pourrait être lancée prochainement : selon les rumeurs, un abonnement à 42 dollars mensuels serait lancé, permettant de procéder à davantage de requêtes pour ChatGPT, de manière prioritaire, et la possibilité de tester en avant-première de nouvelles fonctions.

Immenses possibilités

Pour Sabine Süssstrunk, proposer ChatGPT sous forme de licence pourrait être intéressant pour OpenAI : « Imaginez toutes les entreprises qui ont des centres d'appel, telles que les banques, les assurances, les détaillants, etc. ChatGPT, avec l'ajout d'un logiciel de synthèse vocale, offre de réelles possibilités d'interagir encore mieux avec les appelants que ne le font les logiciels de synthèse vocale actuels. Même les petites entreprises comme les restaurants et les cabinets médicaux peuvent bénéficier de la capacité de ChatGPT à avoir de « vraies » conversations. »

Reste que la conquête du monde par OpenAI n'est pas garantie. Il y aura la concurrence, bien sûr. « D'autres entreprises technologiques, des start-up comme de grandes entreprises telles Google et Meta, travaillent sur des systèmes similaires, et je suis sûre qu'elles incorporeront une technologie similaire dans leurs produits, même si elles prétendent le contraire », estime Sabine Süssstrunk. Selon la spécialiste de l'EPFL, « ChatGPT n'est pas une grande invention nouvelle, c'est « juste » un modèle de traitement du langage naturel qui est entraîné sur d'énormes quantités de données. Cela peut être fait par d'autres. »

OpenAI fait aussi face au risque de mauvais usage. « Nous pourrions disposer d'une abondance incroyable de systèmes qui peuvent nous aider à résoudre les problèmes [...] et nous permettre à tous de vivre mieux », affirmait cette semaine Sam Altman. Tout en ajoutant : « Je suis inquiet d'un cas de mauvaise utilisation accidentelle à court terme... Il faut augmenter la sécurité dans l'intelligence artificielle. Il faut faire beaucoup, beaucoup plus dans ce domaine. »

4. ChatGPT : (encore) une « révolution anthropologique » ?

Depuis la mise en ligne de la version « Chat » de l'intelligence artificielle (IA) GPT, capable d'imiter la production intellectuelle humaine, l'effervescence autour de ces technologies fait ressurgir dans les débats la notion de « révolution anthropologique ».

Or, ce que nous observons, au-delà de ChatGPT, est que chaque nouvelle avancée dans des technologies dites de « rupture » relève en fait de cette ambivalente technophilie, tentée par le technoscepticisme, qui oscille entre émerveillement et frisson face aux rapides et impressionnants progrès de l'informatique et de la cybernétique.

Depuis la victoire de Deep Blue, le super-ordinateur, contre un humain aux échecs, la tonalité est donnée : il se jouerait là un « tournant de l'histoire de l'humanité ».

Source
The Conversation
Lionel Obadia
22 février 2023

Une multitude de dispositifs informatiques

Les dernières prouesses mathématiques et technologiques ont donné naissance à une rapide massification de ce que l'on désigne comme « intelligence artificielle », appellation que des spécialistes comme Luc Julia ou Jean-Louis Dessalles contestent car cette « intelligence » se décline en fait sous une multitude de dispositifs informatiques.

On trouve ainsi pêle-mêle des algorithmes d'aide à la décision, notamment juridique (CaseLaw Analytics), reconnaissance faciale comme FaceNet, musique générée par algorithmes avec JukeDeck, images générées par le texte avec Dall-E 2, agents conversationnels portables (Siri), domotique intelligente qui informe sur le contenu du réfrigérateur, robots humanoïdes capables d'engager et tenir une conversation... on pense ainsi à la désormais célèbre Sofia, qui passe de plateau TV en conférence, ou à Ameca, dont la gestuelle et les expressions ne cessent de montrer les limites de l'humain.

Un imaginaire débridé

La massification des technologies et leur injection tous azimuts dans les sociétés contemporaines sont, certes, remarquables, et donnent l'impression d'une vague irrépressible de technologisation et de digitalisation des environnements humains.

Cette tendance nourrit un imaginaire débridé qui se pense nécessairement en rupture avec le passé, d'où la mode de la « révolution » aux accents transhumanistes. Nous serions ainsi à « l'aube » d'une « nouvelle » humanité prise dans une nouvelle « ère » technologique. Un récit qui oublie les nombreux ratés des dites technologies, d'ailleurs.

La « révolution de l'IA » ou « digitale » est-elle une *révolution anthropologique* ? Curieusement, ce sont les entrepreneurs comme Gilles Babinet, les historiens à succès, comme Yuval Noah Harari, les philosophes, comme Frédéric Worms, qui se sont emparés de l'expression (avec des termes plus ou moins exacts).

Le premier pour affirmer avec force que c'est bien le cas.

Le second, avec son livre *Homo Deus*, pour inscrire cette révolution dans un modèle de longue durée de l'histoire humaine au péril d'une simplification à outrance qui brouille les pistes entre lecture rétrospective de l'histoire et imagination prospective.

Le troisième, enfin, avec un bien plus de mesure, pour au moins souligner la pertinence d'un questionnement sur la profondeur des transformations en cours.

On regrettera que les anthropologues soient malheureusement peu mobilisés sur un débat qui concerne au premier chef la discipline dont le nom est mis à toutes les sauces, à quelques exceptions près. Emmanuel Grimaud a pris à bras le corps la question essentielle de la nature de l'IA par comparaison avec les traits d'une humanité toujours plus questionnée sur ce qui fait sa singularité.

Pascal Picq, s'en est emparé dans un genre très différent, puisque le paléoanthropologue n'hésite pas à sortir du cadre strict de l'anthropogenèse (l'évolution humaine). Il inscrit la révolution digitale dans le temps long de l'évolution humaine et, contre toute réduction intellectuelle, invite à penser la complexité des formes d'intelligences animales et artificielles.

Questionner la notion même de rupture

Les anthropologues, habitués, par spécialisation intellectuelle, au temps long et plus enclins à considérer les continuités que les ruptures (souvent hâtivement annoncées) dans l'ordre des sociétés et les mutations culturelles, ont toutes les raisons d'être circonspects.

D'abord parce que tout changement technologique n'entraîne pas un changement culturel majeur. Suivant Claude Lévi-Strauss notamment dans *Race et Histoire* (1955), il faut peut-être réserver le terme à un phénomène de nature à transformer en profondeur (structurellement) l'ordre de la pensée et de l'organisation sociale.

Ce fut le cas avec la domestication du feu et des espèces animales, la sédentarisation et l'agriculture au néolithique, qui ne sont pas toutes des inventions (*ex-nihilo*) mais bien souvent des innovations au sens anthropologique du terme : l'amélioration d'une technique déjà élaborée par les humains. Et c'est en ce sens que l'IA et nombre de technologies digitales méritent proprement le terme d'innovation plus que d'inventions.

En ce sens, le philosophe Michel Serres évoquait une « troisième révolution » pour qualifier l'avènement du monde digital, après l'écriture et l'imprimerie. Il fait ainsi écho à Jack Goody, qui explique que la raison graphique a métamorphosé une raison orale, structurant la pensée et la

communication humaine pendant des millénaires. En d'autres termes, ce ne sera pas l'unique fois, bien au contraire, où l'humanité se trouve confrontée à une réadaptation de ses manières de penser après un changement dans ses techniques.

Quelle révolution ?

Révolution, donc que cet *Homo numericus* comme on l'affirme ici ou là avec la force de la conviction que le présent explique tout ?

Si on suit l'historienne Adrienne Mayor, les civilisations de l'Antiquité ont déjà imaginé et même commencé à mettre en œuvre des technologies qui sont actuellement sur le devant de la scène, de manière embryonnaire pour l'IA ou plus aboutie pour les robots. Son *God and Robots* tord le cou à des lectures partielles et partiales d'une histoire monolithique et récente.

Elle narre ainsi les récits de Talos, « premier robot », puis le chaudron d'immortalité de Médée, les emprunts des humains aux animaux et aux Dieux pour accroître leurs pouvoirs, les premières statues « vivantes » de Dédale et Pygmalion, la création des humains plus qu'humains par Prométhée, les automates d'Héphaïstos, la première réalité virtuelle qu'incarne Pandore... Pour Mayor, la révolution numérique ne serait qu'une *actualisation* dans le présent d'anciennes technologies.

Une réinvention plus qu'une révolution, donc, et des impacts moins profonds qu'il n'y paraît : les machines étaient là, depuis le début, dans l'imaginaire des humains qui ont tâtonné (dans tous les sens du terme) pour les matérialiser.

Certes, les technologies sont loin d'être identiques – entre les automates grecs et les machines agissantes de Boston Dynamics, l'écart est quand même sensible et toute technologie est d'abord évaluable dans son contexte – mais elles étaient déjà investies des espoirs de l'humanité d'être assistée voire suppléée dans ses tâches physiques (pour les robots) et intellectuelles (pour l'IA).

Comment les humains absorbent-ils les technologies ?

La révolution est-elle d'ailleurs anthropologique ou technologique ? La leçon de l'anthropologie, c'est de considérer comment l'humain élabore des technologies et comment celles-ci sont absorbées par les systèmes sociaux et culturels.

La « révolution numérique » ou « digitale » engendre indubitablement des transformations dans les technologies, sans qu'elles se traduisent mécaniquement par des mutations des modèles de comportements humains. De nouveaux usages apparaissent mais les modèles de comportement peuvent infléchir les techniques, plutôt que d'être infléchis par elles.

Suivant les plus critiques des spécialistes de l'IA, et des technologies digitales, il s'agit de considérer la singularité de chaque technologie et de ses effets : là où la réponse sociale aux agents conversationnels semble plutôt positive et immédiate (les chatbots sont facilement adoptés), les IA de reconnaissance faciale donnent lieu à des contournements au moyen de contre-systèmes informatiques.

Suivant Picq, il faut sans doute réserver aux smartphones un sort à part, dans le sens où ils allient la matérialité d'un ordinateur portable, et la virtuosité technologique des IA. Ce terminal concentré de technologie qui a quand même changé les modes de communication, accompagné ou engendré des changements dans la mobilité humaine, dans le rapport au savoir, etc. que Pascal Picq veut l'objet véritable de la « révolution » actuelle est le détournement dans le sens où il devient ordinateur portable plus que téléphone...

L'humanité s'amuse et se fait peur

Bref, depuis le feu, les armes, la magie – peut-être première technologie de transformation de la réalité humaine – donc première réalité virtuelle de l'histoire ? – jusqu'à l'IA et les robots, l'humanité s'amuse et se fait peur avec ses créations technologiques, utiles ou ludiques.

Elle se plaît parfois à les constituer comme des créatures (sans qu'elles soient nécessairement anthropomorphisées) telles que la pensée mythique les figure. Rien de surprenant dans ce cadre que certains trichent avec ChatGPT par exemple.

En effet, l'humain est un être versé dans le ludisme, c'est-à-dire, qui aime le jeu comme l'affirmait le philosophe Johann Huizinga mais aussi dans le luddisme, une forme de crainte suscitée par les technologies (en référence au conflit social du XIXe siècle qui avait opposé des manufacturiers à des artisans, les luddistes qui dénonçaient l'emploi de machines).

La tentation du fétiche

La tentation reste toutefois grande de céder au fétiche intellectuel du terme « révolution ». Dans la plupart des cas, est qualifiée de révolution un changement culturel ou socio-technique dont sont observées les prémisses et sont conjecturés les impacts qui ne sont pas encore observés.

Or, une révolution (sauf sans doute quand elle se décrète sur le plan *politique*) se mesure à partir d'éléments rétrospectifs et non pas d'extrapolations, qui sont assez stables dans le domaine des sciences et technologies, mais très peu fiables en sciences humaines.

Si l'on considère toutefois les avancées actuelles de l'Intelligence artificielle, non plus du seul point de vue d'une échelle technologique linéaire, comme la loi de Moore, mais sur le plan des absorptions culturelles et adoptions sociales des technologies, sans doute faudrait-il alors parler (d'une même voix avec les informaticiens) d'*évolution* plutôt que de *révolution* anthropologique.

Mais au prix d'un renversement intellectuel majeur : celui de passer d'une pensée technocentrée (qui considère que c'est la technologie qui transforme la société) à une pensée sociocentrée des techniques (la perspective inverse, donc). Une petite « révolution » donc... ?

Source
The Conversation
Laurence Devillers
20 février 2023

5. ChatGPT et « intelligences » artificielles : comment déceler le vrai du faux

Véritable coqueluche de la rentrée 2023, le système interactif ChatGPT a soulevé une vague d'engouement, puis d'interrogations et d'inquiétudes. En très peu de temps, il a rassemblé un million d'utilisateurs et a été testé sur de nombreuses tâches, principalement textuelles : demande d'information, dissertation, génération de fictions, programmation informatique, traduction de textes, écriture de poèmes...

Une des raisons de cette popularité est que ChatGPT a montré des capacités impressionnantes dans de nombreux domaines ainsi que des capacités émergentes, par exemple la génération de code informatique et la génération « multimodale ». Une autre raison est que son interface de dialogue permet aux utilisateurs d'interagir avec un grand modèle de langage sous-jacent GPT3.5 de manière plus efficace et efficiente qu'avant, par le biais de chats interactifs.

Ces résultats ont poussé à se demander si on pouvait utiliser ces grands systèmes de langage à des fins professionnelles, documentaires, scolaires ou encore artistiques. Il est possible que ce type de systèmes transforme certains métiers et qu'il touche en profondeur l'enseignement et l'éducation – les enfants étant particulièrement vulnérables devant ces systèmes.

Une « intelligence »... en apparence seulement

ChatGPT produit des textes quasiment parfaits d'un point de vue grammatical bien qu'il n'ait aucune compréhension de ce qu'il produit. Il a des capacités vraiment étonnantes et certains cas montrés en exemple sont remarquables. Ses textes, souvent complexes, peuvent ressembler aux données originales ayant servi à l'apprentissage et en possèdent certaines caractéristiques.

Mais avec l'apparence du vrai, ces résultats peuvent être parfois totalement faux. Quels sont la nature et le statut de ces paroles artificielles sans raisonnement associé ? La compréhension de la langue naturelle implique des raisonnements complexes et variés, spatiaux, temporels, ontologiques, arithmétiques, basés sur des connaissances et permettant de relier les objets et actions dans le monde réel, ce que ChatGPT est loin d'intégrer n'ayant aucune perception phénoménale.

Si quelques exemples choisis peuvent suggérer que les modèles de langue sont capables de raisonner, ils ne sont en fait capables d'aucun raisonnement logique et n'ont pas d'intuition, pas de pensée, pas d'émotions. ChatGPT s'exprime avec assurance en bon français comme dans d'autres langues après avoir englouti des milliards de données, mais ne comprend rien à ce qu'il dit et peut très facilement générer des fake news, des discriminations, des injustices et amplifier la guerre de l'information.

Comment déceler le vrai du faux : des technos à l'éducation

Ces approches peu transparentes peuvent cependant être évaluées sous de nombreux aspects sur des données existantes (ce sont des *benchmarks*) pour montrer le manque de performance des systèmes sur des problèmes de raisonnements logiques comme la déduction, l'induction ou l'abduction – ou encore le sens commun.

L'éducation peut s'emparer de ce sujet pour montrer les limites de cette langue artificielle désincarnée, et faire travailler les élèves sur une meilleure compréhension des concepts de la modélisation numérique, de l'apprentissage des machines et l'intelligence artificielle.

Des enfants plus crédules devant les IA

Ceci est particulièrement important car les enfants peuvent être plus particulièrement crédules devant ces systèmes doués de parole comme ChatGPT.

Lauréat du prix Nobel d'économie, l'américain Richard Thaler a mis en lumière en 2008 le concept de « nudge », technique qui consiste à inciter les individus à changer de comportement sans les contraindre et en utilisant leurs biais cognitifs.

De plus, nous avons pu montrer que les jeunes enfants suivaient plus les suggestions de systèmes de dialogue embarqués dans des objets (comme un Google Home ou un robot) que celles d'un humain. Notre démarche de recherche était basée un jeu sur l'altruisme et menée dans le cadre de la chaire IA Humaine (pour *Human-Machine Affective Interaction and Ethics*) sur les nudges numériques amplifiés par l'IA. Cette chaire interdisciplinaire, sorte de laboratoire d'études des comportements d'interaction humain-machine associe des chercheurs en informatique, linguistique et économie comportementale.

Les agents conversationnels comme ChatGPT pourraient devenir un moyen d'influence des individus. Ils ne sont pour l'instant ni régulés, ni évalués et très opaques. Il est donc important de bien comprendre leur fonctionnement et limites avant de les utiliser, et dans ce cadre, l'école a un grand rôle à jouer.

Pourquoi ChatGPT est-il aussi puissant ?

ChatGPT est un système multilingue multitâche interactif utilisant une IA générative en libre accès sur Internet. Les systèmes d'IA générative s'appuient sur des algorithmes capables d'encoder de gigantesques volumes de données (textes, poèmes, programmes informatiques, symboles) et de générer des textes syntaxiquement corrects pour un grand nombre de tâches.

Les « transformers » sont un de ces types d'algorithmes. Ce sont des réseaux de neurones qui apprennent les régularités les plus saillantes des mots dans de nombreux contextes et sont ainsi capables de prédire le mot ou la séquence susceptible d'être pertinente dans la suite d'un texte donné.

ChatGPT est le successeur du grand modèle de langage (LLM) InstructGPT, auquel a été ajoutée une interface de dialogue. InstructGPT marche mieux que les précédentes approches : en effet, les développeurs ont réussi à mieux mettre en adéquation l'IA générative (de type GPT3.5) avec l'intention de l'utilisateur, et ce sur un large éventail de tâches. Pour cela, ils utilisent l'apprentissage par renforcement, c'est-à-dire que l'IA apprend aussi des commentaires que les humains font sur ses textes.

Le fait d'augmenter la taille des modèles de langage ne les rend pas intrinsèquement plus aptes à suivre l'intention de l'utilisateur. Ces modèles de langage de grande taille peuvent générer des résultats qui sont mensongers, toxiques ou tout simplement inutiles pour l'utilisateur, car ils ne sont pas alignés sur les intentions de l'utilisateur.

Mais les résultats montrent que le réglage fin grâce à des retours humains est une orientation prometteuse pour aligner les modèles de langage avec l'intention humaine, même si InstructGPT commet encore des erreurs simples.

Ainsi, la performance technologique de ChatGPT provient donc de la taille de l'IA générative utilisant des « transformers » (175 milliards de paramètres), de l'alignement des réponses de l'IA par apprentissage par renforcement mais également de la possibilité de dialoguer avec ce système.

L'impact de ChatGPT sur le marché de la recherche d'information

ChatGPT de Microsoft-OpenAI est une menace pour le modèle d'interrogation de Google par sa puissance de recherche et de production. Google positionne Bard comme un moteur de recherche interactif plus réfléchi et plus précis, qui n'est pas entravé par les problèmes d'actualité rencontrés par ChatGPT puisqu'il a été entraîné sur des données disponibles avant septembre 2021 – et ne connaît donc pas les actualités plus récentes (pour l'instant).

La société chinoise Baidu a également un projet d'IA générative avec Ernie Bot. Le projet « BigScience », créé par HuggingFace et incluant des financements du CNRS et du ministère de la recherche a permis de créer « Bloom », une IA générative utilisant un modèle de langage de 176 milliards de paramètres entraîné sur des données multilingues multitâches et surtout en « science

ouverte » ! Ce projet innovant est une coopération public/privé et a embarqué plus de mille chercheurs de nombreux pays. Il pourrait en découler un « ChatBloom ».

Des questions éthiques

Le contexte actuel est marqué par des réalisations et des applications de ces systèmes largement diffusées dont l'impact massif nécessite une réflexion éthique.

Ces IA génératives multilingues, multitâches et interactives posent de nombreuses questions : les données choisies pour les entraîner, la répartition des langues dans un système multilingue, les paramètres d'optimisation des systèmes, la propriété des contenus générés, etc.

De plus, le pouvoir génératif des IA est souvent amplifié par des filtres qui permettent de censurer certains sujets et des modules de déduction logique dans le but de vérifier la véracité des énoncés. Une poignée d'humains (ingénieurs, transcripteurs, évaluateurs) ont créé ce type de système utilisé par des millions de personnes.

Ces systèmes d'intelligence artificielle utilisés massivement posent donc des défis éthiques majeurs parmi lesquels la transformation de la notion de production d'information, le rapport à la vérité et les risques massifs liés à la désinformation et la manipulation.

6. Pourquoi ChatGPT est une mascarade

Source

L'Obs

Guillaume von der Weid

19 février 2023

Dans une tribune à « l'Obs », le philosophe Guillaume von der Weid explique qu'il ne faut pas redouter l'outil d'intelligence artificielle, car cette technologie ne sait pas réfléchir. D'ailleurs, ce n'est même pas une intelligence.

ChatGPT, dernier né de l'intelligence artificielle, fait couler beaucoup d'encre. Beaucoup de l'encre de ces journaux dont il annonce la mort, en étant capable à la fois de fournir des solutions à tous les problèmes, mais aussi de les commenter, accomplissant le travail traditionnellement dévolu à la presse. Cathédrale d'algorithmes édifée dans le ciment de l'information totale d'internet, ChatGPT comprend les questions qu'on lui pose et structure ses réponses selon les normes conversationnelles, avec des paragraphes articulés, exprimant chacun une idée, abordant tour à tour les différents aspects d'une question pour déboucher, usant d'humour à l'occasion, sur des conclusions nuancées.

Expérience fascinante de voir une machine, c'est-à-dire un bout de matière, nous expliquer ce qu'il faut penser. « *Quand l'univers l'écraserait*, disait Pascal, *l'homme serait encore plus noble que ce qui le tue, parce qu'il sait qu'il meurt, et l'avantage que l'univers a sur lui, l'univers n'en sait rien.* » ChatGPT écrase l'homme en lui expliquant pourquoi.

Ce serait cependant confondre information et savoir. Le savoir n'est pas seulement le fait de posséder des informations sur le monde, par exemple le nom de la capitale du Brésil, le mécanisme de la digestion, la raison de la guerre en Ukraine ou la vitesse de la lumière. C'est *avoir l'idée* de pourquoi on pose ces questions, c'est-à-dire envisager l'existence d'un vrai ultime, d'un puzzle complet, dont les questions particulières seraient les pièces, permettant de saisir le monde *dans ses principes*.

Le savoir, autrement dit, n'est pas une « *bureaucratie des apparences* » (Etienne Klein) : il ne consiste pas seulement à rendre l'information compréhensible, mais à faire des hypothèses sur le puzzle lui-même, la nature du vrai, et donc à imaginer plutôt qu'à synthétiser, à créer plutôt qu'à vulgariser.

Artifices informatiques

Cette confusion a une cause et un moteur. La cause, c'est l'apparence de machines qui semblent *produire* des idées quand elles ne font qu'opérer des calculs selon nos paramètres – à la manière de la génération spontanée, qui prétendait, avant la découverte pastoriennne des microbes, que la vie naissait dans des récipients qui la contenaient déjà imperceptiblement. Ainsi, l'« intelligence » des réponses de ChatGPT se trouve déjà dans ses algorithmes miniaturisés. Dira-t-on d'un tableau noir qu'il est intelligent parce qu'on y a écrit à la craie une équation mathématique ?

On rétorquera que l'intelligence artificielle a produit toute seule l'équation, par combinaison de données mathématiques, dont nous ne savions nous-mêmes rien déduire. C'est précisément cette extraction de vérités cachées qui fonde la prétention à l'intelligence de nos artifices informatiques. Sauf qu'il ne s'agit là que de combiner des données, et non de produire des idées, c'est-à-dire des

réflexions sur ce qu'est une donnée.

L'attribution d'une réflexion à cette forme calculante a aussi un moteur : l'anthropomorphisme, qui prête des caractéristiques humaines à ce que nous ne comprenons pas. Ces machines nous semblent d'autant plus dotées du pouvoir alchimique de transmuter l'information en savoir, qu'ignorant leur principe de fonctionnement, nous y projetons notre pensée, comme à une autre époque les intentions divines sur des phénomènes naturels inexpliqués.

Mascarade encouragée par la *personnification* de la réponse de ChatGPT. Car sa nouveauté est moins de synthétiser du contenu, ce qu'on savait déjà très bien faire, que de répondre aux questions qui lui sont posées « en langage naturel ». Les yeux de Dieu nous regardaient par le mystère de la nature, l'intelligence de l'artifice par la simulation de la parole humaine.

L'intelligence artificielle est bien mal nommée

Borges avait imaginé une bibliothèque qui contiendrait tous les livres possibles, par combinaisons aléatoires et exhaustives de tous les mots et les phrases possibles. Un écrivain ne ferait ainsi que « choisir » son livre dans l'immensité des livres écrivables. De même, une partie d'échecs n'est que le choix, à chaque embranchement de coups, d'une partie parmi la multitude des parties possibles. Allons plus loin : notre vie entière n'est-elle pas le choix d'une vie réelle parmi toutes les vies possibles ? Et dans la totalité déjà écrite du possible, ChatGPT ne nous indique-t-il pas la meilleure d'entre elles ?

Confusion là encore. Nous prendrions la chose pour la raison de la chose, le signe pour l'idée, les instants passés pour la durée créatrice. Sauf que ce qui est important dans un livre, ce ne sont pas les lettres écrites, mais les idées que les lettres font naître en nous, parce qu'elles en procèdent. C'est ce qui différencie une notice d'utilisation décrivant le fonctionnement d'une machine, d'un livre, dont les mots renvoient non pas à un petit nombre d'opérations univoques, mais à des idées qui sont toujours les approximations d'une vérité totale, dépassant l'existence matérielle. Où le savoir touche au spirituel. Les plus grands livres, dont la valeur informative est négligeable, sont précisément ceux dont le mouvement nous entraîne vers l'absolu. Or, cet absolu est une visée, non une synthèse du déjà connu.

Aussi l'intelligence artificielle est-elle bien mal nommée. Pour économiser l'encre et s'épargner l'inquiétude de ne plus avoir à écrire, on devrait plutôt parler d'*artifice intelligent*, expression qui insiste sur la matérialité du dispositif plutôt que l'intelligence dont il procède. Une matérialité sans vie et donc sans but, capable seulement d'optimiser une masse gigantesque d'informations. L'avenir ne saurait être un passé amélioré, et l'intelligence le ressassement indéfini de ce qu'elle a elle-même produit. Tout reste à écrire.

Source
letemps.ch
 Florian
 Fischbacher
 22 février 2023

7. Un magazine américain de science-fiction submergé de textes écrits par des intelligences artificielles

Devant l'afflux de nouvelles rédigées par ChatGPT et autres chatbots, le populaire magazine de science-fiction *Clarkesworld* a dû bloquer les propositions de nouveaux écrits inédits

« Nous n'avons pas de solution pour ce problème. » Résigné, Neil Clarke, l'éditeur du magazine en ligne *Clarkesworld*, spécialisé dans la publication de collections de nouvelles de science-fiction et de fantasy, a annoncé sur Twitter lundi être obligé de geler toute nouvelle soumission de textes sur son site.

Clarkesworld, qui est publié mensuellement en ligne depuis 2006 et fait paraître chaque année une anthologie en version imprimée, a notamment reçu un prestigieux prix Hugo en 2022. Il s'agit de l'un des rares éditeurs à accepter et rémunérer des textes venants de nouveaux auteurs, encore inconnus, rappelle le *Guardian*.

La semaine dernière, dans un post de blog, Neil Clarke, indiquait avoir « observé une augmentation du nombre de soumissions de spams » à son magazine depuis le début de la pandémie, en grande majorité des cas de plagiat. Mais, depuis la fin de l'année 2022, lui parviennent de plus en plus de textes générés par des robots conversationnels, tel le d'ores et déjà célèbre ChatGPT. L'éditeur a précisé sur Twitter que le nombre de ces soumissions avait explosé au cours des dernières semaines, le forçant à bannir quelque 500 personnes ayant proposé des textes frauduleux en février.

A tel point de se retrouver dans l'impossibilité de faire face et de devoir mettre le système en pause depuis lundi.



Selon lui, il s'agirait principalement de personnes «extérieures à la communauté de la science-fiction/fantasy» tentées par l'idée de «gagner facilement de l'argent avec ChatGPT». Neil Clarke dit avoir contacté plusieurs de ses pairs qui semblent être confrontés à des problèmes similaires.

ChatGPT, coauteur ou ghostwriter ?

A la mi-février, Amazon vendait ainsi plus de 200 livres électroniques mentionnant ChatGPT comme auteur ou coauteur, selon une enquête de l'agence Reuters, qui cite notamment en exemple un recueil de poésie ou encore un livre illustré pour enfants. «Et ce nombre augmente chaque jour.»

Lorsque l'auteur numérique est clairement cité, il n'y a pas de véritable problème, mais « de nombreux écrivains ne révèlent pas qu'ils ont eu recours [à ChatGPT]. Il est presque impossible d'obtenir un décompte exhaustif du nombre de livres électroniques qui pourraient avoir été écrits par l'IA », poursuit Reuters. «Il doit y avoir une transparence de la part des auteurs et des plateformes sur la façon dont ces livres sont créés, sinon vous allez vous retrouver avec beaucoup de livres de mauvaise qualité», réagit Mary Rasenberger, du groupe Authors Guild qui redoute que l'écriture – même en tant que *ghostwriter* – ne passe « du statut d'artisanat à celui de marchandise ».

Or, les solutions pour limiter ce phénomène, comme la détection automatique de ces textes, sont loin d'être parfaites. En effet, à mesure que les chatbots se perfectionnent, il devient de plus en plus difficile de déterminer avec certitude si un texte est généré par ordinateur ou écrit par un humain, relève le magazine *Vice* qui précise que même OpenAI – l'entreprise derrière ChatGPT – a convenu que son propre détecteur de textes rédigés par le chatbot « n'est pas totalement fiable ».

Pour ce qui est de Clarkesworld, Neil Clarke envisage plusieurs possibilités, telle que l'introduction de soumissions payantes. Mais s'il devait en arriver là, ce serait à regret; il dit en effet craindre, qu'à terme, tous ces changements imposés par des fraudeurs «n'entraînent un nombre croissant d'obstacles pour les nouveaux auteurs». Or « la fiction en format court a besoin de ces personnes ».

Source
 letemps.ch
 Florian
 Fischbacher
 22 février 2023

8. ChatGPT aura-t-il un jour le Nobel de littérature ?

Cet algorithme a plus d'un tour dans son sac. Capable de cracher des textes de fiction et des récits à la demande, il pourrait faire le bonheur de l'élève paresseux comme celui de l'écrivain chevronné. Qu'en est-il vraiment ?

Et si, du fond de son lit, on pouvait réclamer, non pas à un parent bienveillant, mais à une machine, des histoires pour s'endormir, correspondant à son genre de prédilection, et lues par une intelligence artificielle (IA) ? Il ne s'agit pas de l'incipit d'un roman d'anticipation du génial Philip K. Dick, mais de notre nouvelle réalité, alors que l'IA part à la conquête du verbe. Développé par OpenAI, start-up dont Elon Musk fut cofondateur, le modèle de langage ChatGPT est même la nouvelle star des réseaux sociaux, où l'on commente ses prouesses. Car elle est pleine de ressources : elle peut rédiger e-mails, blagues, haïkus, recettes de cuisine, dialogues imaginaires, parodies de chansons, dissertations de philo... « Impossible de faire le tour de tout ce qu'on peut faire écrire à ChatGPT, c'est à peu près aussi vain que de demander à un humain : « Alors, qu'est-ce que vous pouvez faire avec le langage ? » constate ainsi le vulgarisateur Thibaut Giraud, alias Mr Phi, dans une vidéo YouTube dédiée à cette invention.

Histoires à la chaîne

Globalement, ChatGPT fonctionne en apprenant à imiter tout ce qu'il a ingurgité sur le web, et peut générer ainsi des enfilades de mots à une vitesse éblouissante. Dès son lancement, fin novembre, un million de curieux l'avaient testé. Mais aussi des professionnels de la plume, à l'instar de l'autrice américaine Leanne Leeds, spécialiste de fictions paranormales pour la plateforme Kindle d'Amazon. Obligée d'écrire des histoires à la chaîne pour que son activité reste rentable, elle confiait récemment au média The Verge : « Ce qui est impressionnant, c'est que si je lui dis qu'il s'agit d'un roman policier qui doit être humoristique, il semble comprendre ce que je lui demande. Les noms qu'il me donne pour les suspects sont même mignons. »

L'autrice a déjà utilisé ChatGPT pour ses six derniers romans. Dans son entourage professionnel, d'autres s'en servent aussi pour « rédiger des descriptions », ou « améliorer des intrigues ». Avec une angoisse sourde derrière leurs discussions : « Où tout cela va-t-il mener ? »

Professeur associé à la Haute Ecole d'art et de design de Genève, l'anthropologue Nicolas Nova s'intéresse depuis longtemps à la manière dont les technologies permettent de produire des formes culturelles. Et ausculte bien sûr ce nouvel algorithme. « Il marche bien pour rédiger des contenus très stéréotypés comme une lettre de motivation un peu plate, ou alors, et c'est ce qui fait peur à certains, un raisonnement plutôt scolaire, constate-t-il. Si on demande par exemple à Chat GPT « la notion de bricolage chez Claude Lévi-Strauss », il va être bon, car ce modèle de langage sait produire des choses quand il existe des sources. De la même manière, il pourra aussi rédiger les « pour » et « contre » s'il existe des textes sur le thème demandé ».

Pour trouver l'inspiration

Pour le chercheur, qui envisage d'intégrer l'intelligence artificielle dans les travaux demandés aux étudiants, puisqu'on ne peut plus l'ignorer, elle pourrait même rejoindre la « boîte à outils » des auteurs. « Pour trouver l'inspiration, ou étendre notre capacité à réfléchir, ce n'est pas inintéressant. Certains le font déjà, poursuit-il. Dans la robotique, il existe même un domaine qui s'intéresse à la collaboration entre humain et intelligence : la cobotique, qui montre que ces formes permettent d'être plus efficace ou pertinent. Mais pour produire quelque chose qui soit vraiment neuf, il faut savoir paramétrer la machine. C'est un dialogue. »

ChatGPT sonne-t-il la mort de la littérature telle que nous l'avons connue ? Ou son renouveau ? S'agit-il d'un procédé malhonnête ? Ou génial ? Pas très neuf, en tout cas. En 2010, le média Slate révélait ainsi que Michel Houellebecq avait « copié-collé » des passages entiers de Wikipédia dans son roman, *La Carte et le Territoire*. Dans *L'Obs*, l'écrivain se défendait, arguant un « patchwork » littéraire : « J'ai surtout été influencé par Perec et Borges. Perec y arrivait quand même mieux que moi parce qu'il ne retravaille pas du tout le fragment, ce qui à chaque fois crée un décalage linguistique très fort. » Le procédé ne l'empêcha d'ailleurs pas de remporter le Prix Goncourt. Après

tout, plus d'un siècle plus tôt, Lautreámont recopiait déjà l'Encyclopédie d'histoire naturelle du docteur Chenu pour décrire un vol d'étourneaux.

« Human mind made »

« Le pastiche fait partie de la tradition littéraire, et je pense que l'intelligence artificielle, qui ne fait que reprendre des discours existants pour les recomposer à sa manière, peut être un soulagement pour un auteur, quel qu'il soit », prophétise William Marx, titulaire de la chaire de littératures comparées au Collège de France. Qui pointe par ailleurs les risques de tromperie sur les secrets de fabrication. « Or je crains malheureusement que nous n'allions dans cette direction. Peut-être qu'il y aura un jour des textes certifiés *human mind made* : faits par l'esprit humain, et que ces textes vaudront plus cher que ceux écrits par Chat GPT ? »

L'arrivée des calculatrices avait permis aux mathématiciens de compter plus vite, et dans les sciences, l'intelligence artificielle aide chaque jour plus de chercheurs, mais le langage ne devait-il pas échapper à cette logique ? Selon William Marx, qui vient d'éditionner le *Cours de poésie de Paul Valéry* (Ed. Gallimard), la création artistique contient pourtant déjà sa part mécanique.

« Selon Paul Valéry, il y a deux étapes dans la création, détaille l'historien de la littérature. La première est celle où le créateur reçoit des productions de la sensibilité, des choses qui lui viennent du corps, des émotions, des souvenirs, et qui fournissent son matériau créatif. Mais pour la mise en forme, le créateur n'a pas besoin d'être humain. C'est une intelligence purement abstraite produisant des mécanismes formels entre les phrases et les mots. Valéry disait que quand Racine écrit *Phèdre*, il n'est plus Racine mais une machine. »

Sauf qu'un texte n'est pas seulement une construction mais aussi une intention. « Je n'ai pas peur de me répéter parce que ce que j'écris, ce que je dis ne répond pas aux exigences de la littérature, mais à celles de la nécessité et de l'urgence, à celle du feu », écrit Edouard Louis dans *Qui a tué mon père?* Et William Marx se demande si « la création aléatoire peut attirer beaucoup de lecteurs, sauf éblouis par le côté un peu monstrueux de la chose », tout comme il se demande « quel profit aura le cerveau à déporter vers la machine des fonctions aussi puissamment humaines qu'écrire un roman, un poème, ou un système philosophique: perdra-t-il des capacités ? »

Fiction de propagande

Pour le moment, les poèmes de ChatGPT restent très en deçà de la Pléiade. Mais alors que beaucoup glosent sur son style faiblard, Lenka Zdeborova, professeure associée au laboratoire de physique statistique des systèmes computationnels à l'EPFL, prévient : « Ça va se tromper de moins en moins. Car ce que l'on voit n'est que la première version, dans un domaine qui bouge vite. Et nous aurons des versions plus impressionnantes. A quel point ? Difficile de prédire, dans la mesure où l'humain se trompe aussi. » Elle-même a testé ChatGPT en lui demandant d'inventer un dialogue avec divers héros de la série *Star Trek*. Résultat ? « Intéressant et amusant. » Au point qu'elle imagine chacun pouvoir devenir le créateur de ses propres contenus, puisqu'il est désormais possible d'écrire de la fiction, des scénarios ou des chansons « de façon semi-automatisée ».

Au Japon, un éditeur vient même d'annoncer la parution prochaine d'un manga « artificiel ». Que restera-t-il à ceux qui vivent de leur plume alors que ces machines commencent à savoir produire des pastiches à bas coût, en quantité phénoménale ? Mystère.

Ce qui les sauve encore est que l'intelligence artificielle ne sait pas faire dans la dentelle s'agissant de la véracité et de la précision, mais il est un domaine où elle pourrait vite exceller : « La désinformation et l'appel à la haine », prévient le mathématicien Lê Nguyễn Hoàng, qui s'est spécialisé dans l'éthique et la sécurité des nouvelles technologies. Et celle-ci le préoccupe : « Il y a aujourd'hui un intérêt pour le cybercrime d'investir massivement dans ce genre de technologie afin de l'utiliser pour toutes sortes d'attaques, y compris par la fiction. Car il est beaucoup plus facile d'écrire de la propagande qu'un texte correct. Et pouvoir le faire ainsi, de manière automatisée, est un danger pour la démocratie. »

Autant dire que l'intelligence artificielle ne décrochera pas le Prix Nobel de la paix.

Source
lemonde
informatique.fr
Mark Hachman,
IDG (adapté par
Célia Seramour)
18 mars 2023

9. ChatGPT, une aide puissante à la rédaction souligne le MIT

Une étude du MIT révèle que les pouvoirs de l'IA de ChatGPT améliorent grandement les capacités d'écriture de chacun. Mais à quel prix ? L'IA pourrait aussi coûter des emplois.

ChatGPT et ses pouvoirs d'IA pourraient aider les rédacteurs et les employés de bureau à améliorer la qualité de leurs écrits et à réduire le temps consacré à cette tâche. Toutefois, l'étude du MIT à l'origine de ces conclusions met également en garde les employeurs contre l'utilisation de l'IA pour augmenter le nombre de licenciements. L'article intitulé *Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence*, rédigé par Shakked Noy et Whitney Zhang du département d'économie du MIT, est considéré comme un document de travail et n'a pas fait l'objet d'un examen par les pairs. Néanmoins, les conclusions de l'étude sur la technologie du chatbot IA de ChatGPT sont à la fois fascinantes et inquiétantes, en particulier lorsque l'étude prend en compte la manière dont elle affecte les salariés.

Les deux doctorants ont divisé 444 professionnels diplômés en deux groupes et leur ont demandé de rédiger des communiqués de presse, des courriels, des rapports succincts et des plans d'analyse - une journée de travail normale pour de nombreuses personnes. L'un des groupes a utilisé un éditeur de texte. L'autre groupe a été formé et autorisé à se servir de ChatGPT pour effectuer le travail à sa place. Les résultats ont été évalués, une note leur a été attribuée et ceux qui ont obtenu la note la plus élevée ont reçu une somme d'argent réelle (bien que nominale) pour leurs notes élevées, ce qui les a poussés à fournir leur meilleur travail.

Une réduction de temps impressionnante sur certaines tâches

L'étude a révélé que ceux qui ont utilisé ChatGPT ont réduit de près de moitié le temps nécessaire à l'exécution des tâches (de 30 minutes à environ 17 minutes). La qualité du travail s'est également améliorée, les notes attribuées par les évaluateurs (sur une échelle de 1 à 7) passant de 4 à environ 4,7. Ceux qui ont utilisé l'éditeur de texte ont constaté une légère diminution du temps passé, mais aussi une baisse de la qualité du travail.

Les auteurs concluent que les professionnels de niveau universitaire effectuant des tâches de rédaction professionnelle de niveau moyen voient leur productivité augmenter de manière substantielle lorsqu'ils ont accès à ChatGPT. L'assistant IA améliore la qualité du travail des personnes peu compétentes tout en réduisant le temps qu'elles y consacrent, et il permet aux personnes très compétentes de maintenir leur niveau de qualité tout en devenant beaucoup plus rapides.

Des employés prêts à payer pour ChatGPT

D'autres résultats de l'étude sont également intéressants. Environ 68 % des travailleurs ont simplement copié et collé les résultats de ChatGPT sans les modifier. En outre, même si certains testeurs ont passé plus de temps à éditer les réponses de ChatGPT, la façon dont l'étude a été mise en place (contrôle de la production minute par minute) a également montré que les personnes testées n'ont pas vu de bénéfice en termes de qualité en éditant le travail de l'assistant - bien qu'elles aient passé plus de temps à le faire. Cela était vrai même lorsque les auteurs de l'étude offraient des incitations à l'amélioration du travail. Après le traitement, la part de temps consacrée à la rédaction d'un brouillon a diminué de plus de la moitié et la part de temps consacrée à l'édition a plus que doublé, selon l'étude.

Les participants à l'étude ont également indiqué que l'utilisation de ChatGPT augmentait de manière significative leur satisfaction au travail, en éliminant la surcharge de travail qui peut affecter certaines tâches, et ils étaient prêts à payer environ 0,5 % de leur salaire mensuel pour un abonnement à ChatGPT (ChatGPT facture déjà un accès payant de 20 \$ par mois). (Mais les personnes testées ont également indiqué que ChatGPT n'en savait pas toujours autant qu'elles. L'autre problème, notent les auteurs - n'oubliez pas qu'il s'agit d'économistes - est que les employeurs pourraient eux aussi commencer à reconnaître la valeur de ChatGPT. Cela pourrait mettre en péril les emplois de certaines personnes.

L'IA, prochain concurrent des employés ?

Au niveau global, ChatGPT comprime considérablement la distribution de la productivité, réduisant ainsi les inégalités, écrivent Shakked Noy et Whitney Zhang. En outre, de nombreux employés l'utilisent déjà dans leur emploi au quotidien. Les données expérimentales suggèrent que le ChatGPT se substitue largement à l'effort des collaborateurs plutôt que de compléter leurs compétences, ce qui pourrait entraîner une diminution de la demande de travailleurs, avec des effets distributifs négatifs, les détenteurs de capitaux gagnant au détriment des employés.

Source
letemps.ch
Julie Eigenmann
13 avril 2023

10. ChatGPT s'invite déjà au bureau

De nombreux professionnels en Suisse utilisent l'intelligence artificielle pour automatiser certaines tâches quotidiennes. Le procédé permet de gagner en efficacité même s'il ne faut pas occulter la question des données.

ChatGPT et consorts font légitimement craindre un remplacement de certains travailleurs. Mais bien exploités, ces outils peuvent se révéler d'une redoutable efficacité, en épargnant des tâches rébarbatives et en faisant gagner en productivité.

Une récente étude de Goldman Sachs estime que pour les postes affectés par les intelligences artificielles (IA) aux Etats-Unis, jusqu'à 50 % de la charge de travail pourrait être remplacée. Un autre rapport réalisé par OpenAI, qui a développé ChatGPT, juge que les IA seraient en mesure d'avoir un impact sur 80 % de la main-d'œuvre américaine et de s'acquitter d'au moins 10 % de ses tâches. Et pour 20 % de ces salariés, la moitié du travail pourrait être effectuée par les IA.

Un support administratif

En Suisse, plusieurs personnes s'appuient déjà sur ChatGPT dans le cadre de leur travail. A l'image de Nathalie Brodard, fondatrice de la société de conseil Brodiance, spécialisée dans les ressources humaines, et de Brodard Executive Search, un cabinet de recrutement. « Je l'utilise activement depuis un mois comme support administratif pour différentes tâches, par exemple pour rédiger des réponses aux candidats concernant des descriptifs de postes, que je reprends ensuite. Suivant la tâche, cela me permet de gagner une vingtaine de minutes et donc d'être plus efficace, professionnelle et réactive. »

Dans le cadre de recrutements spécifiques, « l'outil peut aussi suggérer des cas pratiques à soumettre au candidat en entretien d'embauche. On utilise ensuite notre expertise et notre expérience pour aller plus loin dans les discussions », rapporte-t-elle.

Devenir meilleur encore, c'est aussi ce à quoi aspire Blaise Reymondin lorsqu'il utilise cette technologie. Cofondateur de la société Blaise & Bruno Acquisition Marketing, active dans le marketing digital, il détaille : « ChatGPT est très utile pour compléter un énoncé sur la base d'une idée qu'on lui soumet : s'ensuit un échange qui amène de nouvelles pistes. Je lui demande également quelles sont les préoccupations d'un client par rapport à tel ou tel produit, quelles sont les émotions qui peuvent le conduire à l'envisager. Ou je teste très rapidement différentes options de textes pour une page internet pour déterminer celle qui a le plus d'impact. »

Tout vérifier

Il s'agit vraiment de compétences augmentées et non de substitution, assure Blaise Reymondin. « L'outil permet à chaque individu de passer à un niveau supérieur, mais pas davantage : un spécialiste peut avec lui devenir un expert, par exemple. C'est la raison pour laquelle ceux qui ne s'y mettraient pas risqueraient de se faire doubler par des professionnels de niveau inférieur. Mais quoi qu'il en soit, je vérifie toujours les informations que l'outil fait remonter, puisqu'on sait qu'il peut produire des « hallucinations » [des inventions, ndr]. Chacun doit en avoir conscience pour faire le tri. »

Le professionnel qui commettrait une erreur en se reposant sur un résultat généré par ChatGPT engagerait sa propre responsabilité, prévient Philippe Gilliéron, avocat spécialiste en matière de droits de la propriété intellectuelle et du droit des technologies. « Les conditions d'utilisation d'OpenAI précisent que le système peut se tromper. Il est donc fondamental de toujours conserver un esprit critique face aux résultats. Or, seul un spécialiste du domaine sera bien souvent à même de

déceler les éventuelles erreurs commises. »

Pour ce qui a trait aux droits d'auteur, le recours à ChatGPT présente peu de problèmes. A supposer qu'un utilisateur le sollicite pour lui suggérer des idées de nouveaux modèles d'affaires, Philippe Gilliéron souligne tout de même : « Une simple idée ou un concept n'est jamais protégé en tant que tel par des droits d'auteur sans avoir été concrétisé de manière tangible; il n'en demeure pas moins que l'idée amenée par ChatGPT peut porter sur un modèle d'ores et déjà existant sur le marché. L'utilisateur doit donc prendre ses précautions avant toute éventuelle mise en œuvre. »

Pour trier les demandes des clients

Toute la question est de faire bon usage des réponses, ce à quoi veille Patrick Tundo, fondateur d'Assymba, active dans les services informatiques pour les PME. Il utilise les IA afin de générer des images d'illustration pour des modes d'emploi. Mais surtout, ChatGPT lui permet de trier les demandes de clients qui rencontrent des problèmes, avec leur imprimante, leur adresse e-mail ou autres tracas. « L'IA lit le contenu et propose des pistes en indiquant par exemple qu'un cas similaire a été traité récemment, raconte-t-il. La suggestion ne convient pas toujours mais globalement tout le monde y gagne : on optimise notre processus, le client attend moins et nous lui facturons moins. »

L'entrepreneur précise toutefois : « L'IA reçoit le problème uniquement, pas les noms des clients ou autres données personnelles. » La question des données est en effet à examiner. Romain Boichat, fondateur de la plateforme romande de télémédecine Soignez-moi.ch, en est bien conscient. Aujourd'hui il soumet à l'IA, dans le cadre d'une phase de test, un fichier audio d'une consultation : elle lui en propose un résumé très clair et peut aussi rédiger un rapport comme le ferait un médecin, qu'il ne reste plus ensuite qu'à corriger.

« ChatGPT peut aussi sortir la quintessence des nombreux PDF qui constituent le dossier électronique du patient, se réjouit Romain Boichat. Nous voulions voir si c'était réalisable techniquement avec des personnes qui ont donné leur accord. Cela permet de gagner beaucoup de temps, mais s'il fallait l'implémenter, la question des données des patients serait un véritable enjeu. »

Des données sensibles

Les autorités italiennes ont annoncé fin mars leur décision de bloquer ChatGPT, accusé notamment de ne pas respecter la législation sur les données personnelles.

« Il faut être très vigilant, avertit l'avocat Philippe Gilliéron. Un médecin et un avocat par exemple peuvent détenir des informations très sensibles. Le recours à de tels outils ne devrait jamais conduire à divulguer au système des données personnelles ou couvertes par le secret professionnel. »

Comme toute révolution technologique, ChatGPT et ses déclinaisons suscitent leur lot d'interrogations et de craintes légitimes. Elles représentent aussi des opportunités pour améliorer efficacité et productivité au travail. Les spécialistes le soulignent : le risque de rater le train de l'IA pourrait coûter cher.

Source
letemps.ch
Anouch
Seydtaghia
19 avril 2023

11. Avec Auto-GPT, l'intelligence artificielle entre dans une nouvelle dimension

Désormais, des systèmes d'intelligence artificielle tels Auto-GPT et BabyAGI, basés sur ChatGPT, savent travailler seuls pour accomplir des tâches complexes. Explication d'un phénomène aussi massif qu'ultra-rapide.

Cela va décidément trop vite. Le grand public peine à suivre les évolutions fulgurantes autour de l'intelligence artificielle (IA). Et même les ingénieurs semblent pris de court par les innovations qui se multiplient à un rythme effréné. Et pourtant, spécialistes et non initiés ont tout intérêt à suivre ce qui se passe autour de l'IA, tant ses évolutions nous touchent directement. Ces jours, une nouvelle tendance majeure se dessine : la création de systèmes d'IA autonomes multitâches, capables d'effectuer des missions complexes à la place des humains.

D'abord, établissons le décor. Le 30 novembre dernier, la société OpenAI mettait à disposition de tous gratuitement son agent conversationnel ChatGPT. Le principe : l'utilisateur pose des questions en langage naturel, la machine répond, de manière unique ou dans le cadre d'une conversation permettant d'affiner un sujet. ChatGPT a vite évolué, la version 4, payante (environ 20 dollars par mois), offrant des réponses détaillées à des requêtes complexes.

Beaucoup plus loin

Mais cela reste une discussion. Ces derniers jours, des développeurs sont allés beaucoup plus loin, en créant, sur la base de ChatGPT, des systèmes autonomes. Le principal, actuellement, est Auto-GPT, conçu par l'ingénieur Toran Bruce Richards, célèbre depuis. Le développeur a déposé sur la plateforme GitHub son projet, utilisable par tous. Il le définit ainsi : « Auto-GPT est une application expérimentale en code open source qui met en évidence les capacités du modèle de langage GPT-4. Ce programme, piloté par le GPT-4, enchaîne les « pensées » de larges modèles de langage (LLM) pour atteindre de manière autonome l'objectif que vous vous êtes fixé. [...] Auto-GPT repousse les limites de ce qui est possible avec l'IA. »

C'est donc un super système qui utilise plusieurs ingrédients en même temps : GPT-4 pour générer du texte et du code et GPT-3.5 pour stocker et synthétiser des fichiers. Le programme va aussi accéder à internet pour rechercher des informations (notamment via Google) et les rassembler. Le système va également écrire, corriger et développer son propre code en langage Python, l'un des plus populaires actuellement. Il sera aussi capable de corriger ses propres erreurs.

L'idée est donc de se passer au maximum d'interventions humaines pour laisser la machine atteindre par ses propres moyens des objectifs.

Investisseur autonome

Pour quelles applications ? On a vu Auto-GPT être employé pour créer des études de marché complètes: analyser la pertinence de lancer un nouveau produit, le développer, évaluer le positionnement des concurrents. Auto-GPT a aussi été utilisé pour écrire un livre de recettes : le système les a élaborées, il a rédigé l'introduction de l'ouvrage, fait la mise en pages et recherché des images. Le système a également conçu les bases d'une start-up et a aussi été employé pour définir une stratégie d'investissement en ligne dans les cryptomonnaies et la suivre jour après jour.

Sur Twitter, les vidéos montrant comment Auto-GPT écrit du code et les résultats obtenus se multiplient. « Auto-GPT a une démarche proche de la récursivité pour aborder, à partir d'une simple requête, l'ensemble des éléments sous-jacents qu'il va rencontrer au fil de son traitement, écrivait récemment sur Twitter Mathieu Crucq, directeur de la société Brainsonic. [...] Comme le projet est encore jeune, il peut par moments tourner en boucle. Normal. Mais ça reste hyper intéressant à observer, moins dans sa capacité technique que dans la réalité qu'il met en évidence: oui, les IA vont pouvoir se combiner entre elles; oui, les IA vont évoluer extrêmement rapidement; oui, des usages nouveaux vont naître, dans une temporalité assez surréaliste. »

Imperfections

Tout comme Auto-GPT, le système comparable BabyAGI du développeur Yohei Nakajima n'est de loin pas encore parfait : il lui faut encore une supervision humaine, il fait des erreurs et tourne parfois en boucle. De plus, ces systèmes ne se destinent pas à tout le monde, car leur exploitation, à la différence de ChatGPT, nécessite des connaissances informatiques. Mais l'histoire avançant si vite, et les développeurs se ruant en masse sur ces systèmes, ces deux points devraient être rapidement résolus, du moins en partie : les spécialistes s'attendent à ce que ces services soient plus fiables, et à ce que des interfaces ne requérant aucune programmation voient rapidement le jour.

On s'achemine ainsi rapidement vers une sorte d'intelligence artificielle générale, capable de résoudre des problèmes complexes de manière semi-automatique, voire totalement automatique. « L'objectif n'est pas de supprimer la supervision humaine, mais de permettre à de nombreuses personnes de passer de l'exécution des tâches à la gestion des tâches », a affirmé récemment Yohei Nakajima au site spécialisé Motherboard.

Questions éthiques

Et, bien sûr, des questions sur la sécurité et l'utilisation à bon escient de ces systèmes se posent. Toran Bruce Richards en est bien conscient. « Il est essentiel d'équilibrer les avantages et les risques potentiels, affirmait-il lui aussi au site Motherboard. Veiller à ce que l'agent opère dans les limites éthiques et juridiques tout en respectant la vie privée et les préoccupations en matière de sécurité doit être une priorité. C'est pourquoi la supervision humaine est toujours recommandée, car elle permet d'atténuer les problèmes potentiels et de guider l'agent vers les résultats souhaités. »

Ces questions éthiques, Elon Musk semble lui aussi les avoir à cœur. Dans la nuit de lundi à mardi, il affirmait ceci sur la chaîne américaine Fox News : « Je vais lancer quelque chose qui

s'appelle Truth GPT ou une IA qui recherche la vérité maximale et qui essaie de comprendre la nature de l'univers. » Reste que les projets du milliardaire de concurrencer OpenAI sont flous. En parallèle, ce sont surtout des développeurs comme Toran Bruce Richards et Yohei Nakajima, suivis par des millions d'autres, qui font évoluer l'IA extraordinairement rapidement.

Source
enerzine.com
30 juillet 2023

12. Un chercheur explique la quantité d'énergie consommée par le ChatGPT

ChatGPT et d'autres grands modèles de langage apprennent à imiter les humains en analysant d'énormes quantités de données. Derrière la boîte de texte d'un chatbot se trouve un vaste réseau d'unités de traitement informatique qui prennent en charge la formation et l'exécution de ces modèles.

Quelle est la consommation d'énergie des réseaux utilisant de grands modèles de langage ? Beaucoup, selon Sajjad Moazeni, professeur adjoint d'ingénierie électrique et informatique à l'Université de Washington, qui étudie les réseaux pour l'IA et les supercalculateurs d'apprentissage automatique. La simple formation d'un chatbot peut consommer autant d'électricité qu'un quartier en un an.

Comment les grands modèles de langage, tels que ChatGPT, se comparent-ils à l'informatique en nuage sur le plan énergétique ?

Sajjad Moazeni : Ces modèles sont devenus si importants qu'il faut des milliers de processeurs pour former les modèles et répondre aux milliards de requêtes quotidiennes des utilisateurs. Tous ces calculs ne peuvent être effectués que dans un centre de données.

En comparaison, les charges de travail classiques de l'informatique en nuage, telles que les services en ligne, les bases de données et le streaming vidéo, sont beaucoup moins intensives en termes de calcul et nécessitent des ordres de grandeur moins importants en termes d'utilisation de la mémoire.

Pouvez-vous décrire ces centres de données ?

SM : Dans les centres de données actuels, il y a des centaines de milliers d'unités de traitement qui peuvent communiquer entre elles grâce à un grand nombre de fibres optiques et de commutateurs de réseau. Ces processeurs (en plus de la mémoire et des dispositifs de stockage) sont stockés dans des baies de serveurs. Il existe également une infrastructure interne pour refroidir les serveurs (avec de l'eau et de l'air) et des unités pour générer et distribuer l'énergie.

Il existe des centaines de centres de données de ce type dans le monde et ils sont principalement gérés par de grandes entreprises technologiques telles qu'Amazon, Microsoft et Google.

Quelle quantité d'énergie ces grands centres de données consomment-ils pour faire fonctionner ces grands modèles de langage ?

SM : Pour la formation d'un grand modèle linguistique, chaque unité de traitement peut consommer plus de 400 watts d'énergie en fonctionnement. En général, il faut également consommer une quantité d'énergie similaire pour le refroidissement et la gestion de l'énergie. Dans l'ensemble, la consommation d'énergie peut atteindre 10 gigawattheures (GWh) pour la formation d'un seul grand modèle linguistique tel que ChatGPT-3.

Aujourd'hui, il y a des centaines de millions de requêtes quotidiennes sur ChatGPT, bien que ce nombre soit en baisse. Ce nombre de requêtes peut coûter environ 1 GWh par jour, ce qui équivaut à la consommation quotidienne d'énergie d'environ 1000 ménages américains.

Si ces chiffres peuvent sembler corrects pour l'instant, ce n'est que le début d'un développement et d'une adoption à grande échelle de ces modèles. Nous nous attendons à ce que, bientôt, de nombreux services différents utilisent cette technologie quotidiennement.

En outre, à mesure que les modèles deviennent plus sophistiqués, ils deviennent de plus en plus

grands, ce qui signifie que l'énergie des centres de données pour la formation et l'utilisation de ces modèles peut devenir insoutenable. Chaque grande entreprise technologique essaie maintenant de développer son propre modèle, ce qui peut entraîner une charge de formation énorme pour les centres de données.

Quelles sont les solutions potentielles à ce problème ?

SM : Les chercheurs ont essayé d'optimiser le matériel et les processeurs des centres de données afin de les rendre plus efficaces sur le plan énergétique pour ces types de calculs.

Mon groupe se concentre plus particulièrement sur l'aspect réseau. Dans les centres de données actuels, les processeurs envoient des signaux électriques pour faire entrer ou sortir les données à calculer. Mais ces signaux électriques peuvent être déformés. Afin d'envoyer rapidement un grand nombre de données, nous devons utiliser beaucoup d'énergie pour nous assurer que les signaux sont reçus correctement.

Nous construisons la prochaine génération de solutions d'interconnexion optique, qui comprend la conversion de ces signaux électriques en signaux optiques. Ces signaux optiques ont une perte beaucoup plus faible, ce qui minimise la consommation d'énergie.

Étant donné que nous n'en sommes qu'aux premières phases de cette nouvelle technologie, il est très important que les gens soient transparents quant à leurs résultats et créent des modèles à source ouverte. Cela nous aidera également à trouver des solutions avancées et durables.