

La Promesse trompeuse de ChatGPT

Le 8 mars 2023 par Noam Chomsky, Ian Roberts et Jeffrey Watumull

<https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>

Noam Chomsky est professeur de linguistique à l'université d'Arizona et professeur émérite de linguistique au Massachusetts Institute of Technology. Ian Roberts est professeur de linguistique à l'université de Cambridge. Jeffrey Watumull est philosophe et directeur de l'intelligence artificielle chez Oceanit, une société de science et de technologie.



ChatGPT (Representative image [NYT Editorial Board](#))

Jorge Luis Borges a écrit un jour que vivre à une époque de grands dangers et de grandes promesses, c'est faire l'expérience de vivre à la fois la tragédie et la comédie, avec « l'imminence d'une révélation » pour nous comprendre nous-mêmes et comprendre le monde. Aujourd'hui, nos prétendus progrès révolutionnaires dans les domaines de l'intelligence artificielle sont à la fois sources d'inquiétude et d'optimisme.

De l'optimisme parce que l'intelligence est ce qui nous permet de résoudre les problèmes. Inquiétude parce que nous craignons que la forme la plus populaire et la plus en vogue de l'I.A. - l'apprentissage automatique - ne vienne nuire à notre science et dégrader notre éthique en instillant une conception fondamentalement erronée du langage et de la connaissance dans notre technologie.

ChatGPT d'OpenAI, Bard de Google et Sydney de Microsoft sont des merveilles de l'apprentissage automatique. En gros, ils prennent d'énormes quantités de données, y recherchent des modèles et deviennent de plus en plus performants pour générer des données statistiquement plausibles — comme un langage et une pensée apparemment analogues à ce que l'on trouve chez les humains.

Ces programmes ont été salués comme les premières lueurs à l'horizon de l'intelligence artificielle généralisée — ce moment longtemps prophétisé qui voit les esprits mécaniques surpasser les cerveaux humains non seulement quantitativement en termes de vitesse de traitement et de taille de la mémoire, mais aussi qualitativement en termes de perspicacité intellectuelle, de créativité artistique et de toute autre faculté spécifique à l'être humain.

Ce jour viendra peut-être, mais son aube n'est pas encore levée, contrairement à ce que l'on peut lire dans les titres hyperboliques et prévoir dans le cadre d'investissements imprudents.. La révélation borgésienne de la compréhension n'a pas eu lieu et n'aura pas lieu — et, à notre avis, elle ne pourra pas avoir lieu — si les programmes d'apprentissage automatique comme ChatGPT continuent de dominer le domaine de l'IA.

OPINION
GUEST ESSAY

Noam Chomsky: The False Promise of ChatGPT

March 11, 2023

Noam Chomsky: la fausse promesse de ChatGPT

Quelqu'utiles que puissent être ces programmes dans certains domaines restreints (dans le domaine de la programmation informatique, par exemple, ou pour proposer des rimes pour un poème léger), nous savons, grâce à la science de la linguistique et à la philosophie de la connaissance, qu'ils diffèrent profondément de la façon dont les humains réfléchissent, utilisent les mots. Ces divergences restreignent de manière considérable ce que ces programmes peuvent faire, leur conférant des défauts rédhibitoires.

Il est à la fois comique et tragique, comme Borges aurait pu le souligner, que tant d'argent et d'attention se concentrent sur si peu, quelque chose de si trivial comparé à l'esprit humain qui, par le biais du langage, pour reprendre les mots de Wilhelm von Humboldt, peut faire «un usage infini de moyens finis», élaborant des idées et des théories ayant une portée universelle.

Contrairement à ChatGPT et ses semblables, l'esprit humain n'est pas un volumineux moteur de recherches statistiques en quête de modèles, avalant des centaines de téraoctets de données et extrapolant la réponse la plus probable à une question ou la solution la plus vraisemblable à un problème scientifique.

Bien au contraire, l'esprit humain est un système étonnamment efficace et même raffiné qui fonctionne avec de petites quantités d'informations; il ne cherche pas à déduire des corrélations sommaires à partir de données, mais à élaborer des explications.

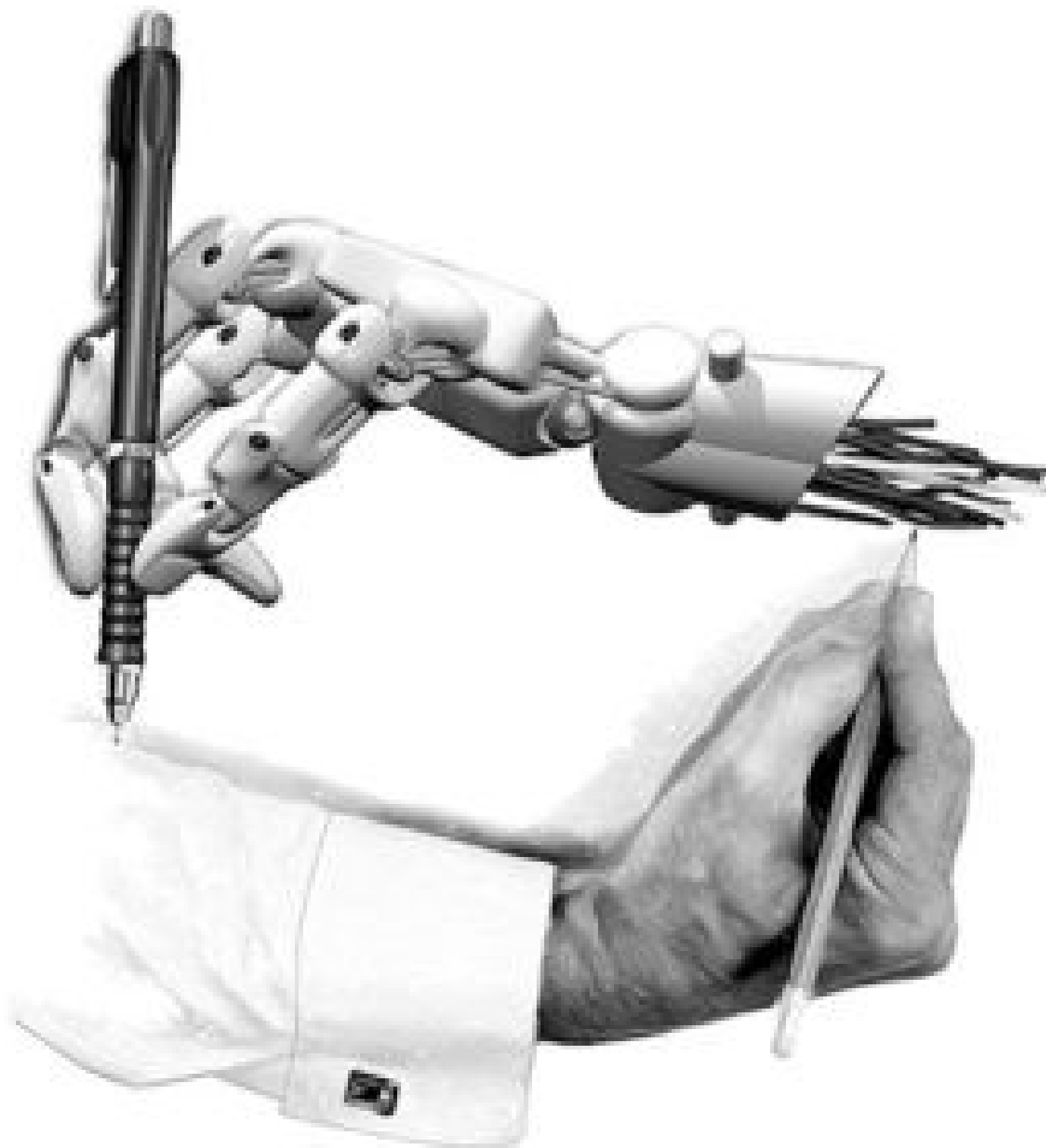
Prenons le cas d'un jeune enfant qui apprend une langue, il développe — inconsciemment, naturellement et rapidement à partir de données minuscules — une grammaire, un système incroyablement sophistiqué de principes et de paramètres logiques. Cette grammaire peut être considérée comme l'expression du «système d'exploitation» inné, d'origine génétique, qui confère à l'homme la capacité de générer des phrases complexes et de développer de longs raisonnements.

Lorsque les linguistes veulent élaborer une théorie permettant de comprendre le fonctionnement d'une langue donnée («Pourquoi telle phrase est-elle considérée comme grammaticalement correcte et telle autre non?»), consciemment et laborieusement ils construisent une version structurée de la grammaire que l'enfant conçoit instinctivement et avec le minimum de contact avec l'information. Le système d'exploitation de l'enfant est complètement différent de celui d'un programme d'apprentissage automatique.

Ces programmes sont en réalité cantonnés à une phase préhumaine ou non humaine de l'évolution cognitive. Leur plus grande faiblesse est leur totale absence de toute capacité la plus critique de quelque intelligence que ce soit: dire non seulement ce qui est le cas, ce qui a été le cas et ce qui sera le cas — telles sont la description et la prédiction — mais aussi ce qui n'est pas le cas et ce qui pourrait et ne pourrait pas être le cas. Tels sont les composants de l'explication, la marque d'une réelle intelligence.

Illustrons cela par un exemple. Supposons que vous tenez une pomme dans la main. Maintenant, vous lâchez la pomme. Vous observez le résultat et vous dites: «La pomme tombe». Il s'agit d'une description. Une prédiction aurait pu être

l'affirmation suivante: «La pomme tombera si j'ouvre la main». Les deux sont pertinentes et peuvent être correctes.



Qui crée qui? (Source Berkeley blog)

Mais une explication propose quelque chose en plus: elle englobe non seulement des descriptions et des prédictions, mais aussi des conjectures hypothétiques telles que «Un tel objet tomberait», plus la clause supplémentaire «en raison de la force de gravité» ou «du fait de la courbure de l'espace-temps» etc. Il s'agit d'une analyse de causalité: «La pomme ne serait pas tombée sans la force de gravité». Voilà ce qu'est la pensée.

Au fondement de l'apprentissage automatique, on trouve la description et la prédiction ; on n'y trouve pas de mécanismes de causalité ou de lois physiques. Bien entendu, toute explication de nature humaine n'est pas forcément exacte ; nous sommes faillibles. Mais cela fait partie de ce que signifie penser : pour avoir raison, on doit pouvoir se tromper.



Noam Chomsky

L'intelligence ne consiste pas seulement à émettre des conjectures novatrices, il faut aussi savoir faire preuve de créativité dans la critique. La pensée humaine se fonde sur des explications envisageables et sur la correction des erreurs, un processus qui limite progressivement les possibilités qui peuvent être envisagées de manière rationnelle. (Comme l'a dit Sherlock Holmes au Dr Watson: «Lorsqu'on a éliminé l'impossible, tout ce qui reste, aussi improbable que cela puisse être, est forcément la vérité»).

Mais ChatGPT et les programmes de même ordre sont, par leur conception, illimités dans ce qu'ils peuvent «apprendre» (c'est-à-dire mémoriser); ils sont incapables de faire la différence entre le possible et l'impossible. Contrairement aux humains, par exemple, qui sont dotés d'une grammaire universelle qui limite le choix des langues que nous pouvons apprendre à celles qui ont un certain degré de raffinement, presque mathématique, ces programmes apprennent avec la même facilité des langues qui sont pour les humains possibles mais aussi impossibles.

Alors que les humains ne disposent que d'un nombre limité d'explications rationnellement envisageables, les systèmes d'apprentissage automatique peuvent apprendre à la fois que la terre est plate et qu'elle est ronde. Ils se contentent pour ce faire d'utiliser des probabilités qui évoluent avec le temps.

C'est la raison pour laquelle les prédictions des systèmes d'apprentissage automatique seront toujours superficielles et contestables. Dans la mesure où ces programmes ne peuvent pas appliquer les règles de la syntaxe anglaise, ils peuvent, par exemple, prédire, à tort, que «John is too stubborn to talk to» signifie que John est tellement têtu qu'il ne parlera pas à quelqu'un d'autre (alors que la traduction est que John est trop têtu pour qu'on le raisonne).

Pourquoi un programme d'apprentissage intelligent prédit-il quelque chose d'aussi bizarre? Parce qu'il pourrait faire une analogie entre le modèle qu'il a déduit de phrases telles que «Jean a mangé une pomme» et «Jean a mangé», dans lesquelles cette dernière expression signifie que Jean a mangé quelque chose.

Le programme pourrait tout à fait prédire que, puisque «Jean est trop têtu pour parler à Bill» est de la même forme que «Jean a mangé une pomme», «Jean est trop têtu pour parler à» devrait être similaire à «Jean a mangé». Concernant le langage, les explications appropriées sont compliquées et on ne peut les appréhender en se contentant de macérer dans des mégadonnées (big data).

Paradoxalement, certains fervents adeptes de l'apprentissage automatique semblent être fiers que leurs créations puissent générer des prédictions «scientifiques» correctes (par exemple, sur le mouvement des corps physiques) sans avoir recours à des explications (impliquant, par exemple, les lois du mouvement de Newton et la gravitation universelle).

Mais ce genre de prédiction, même quand elle est couronnée de succès, relève de la pseudoscience. Les scientifiques recherchent certes des théories largement corroborées sur le plan empirique, comme l'a fait remarquer le philosophe Karl Popper, «nous ne recherchons pas des théories fortement probables, mais des explications, c'est-à-dire des théories

convaincantes et fort improbables».



Automate (Associated Press)

La théorie selon laquelle les pommes tombent sur la terre parce que c'est leur destination naturelle (point de vue d'Aristote) est tout à fait plausible, mais elle ne fait que soulever d'autres questions (pourquoi la terre est-elle leur destination naturelle?).

La théorie selon laquelle les pommes tombent sur la terre parce que la masse courbe l'espace-temps (point de vue d'Einstein) est hautement improbable, mais elle permet de comprendre pourquoi elles tombent. La véritable intelligence se manifeste dans la capacité à penser et à exprimer des choses improbables mais pertinentes.

L'intelligence vraie est également capable de pensée morale. Cela signifie qu'il faut restreindre la créativité par ailleurs illimitée de nos esprits par un ensemble de principes éthiques qui déterminent ce qui devrait et ne devrait pas être (et bien sûr soumettre ces principes eux-mêmes à une analyse critique).

Pour pouvoir servir à quelque chose, ChatGPT doit être en mesure de générer des textes à caractère romanesque; pour être acceptable par la plupart de ses utilisateurs, il doit éviter tout contenu moralement répréhensible. Mais c'est un équilibre que les programmeurs de ChatGPT et d'autres merveilles de l'apprentissage automatique ont eu du mal - et continueront d'avoir du mal - à atteindre.

En 2016, par exemple, le chatbot Tay de Microsoft (précurseur de ChatGPT) a inondé l'internet de contenus misogynes et racistes, après avoir été pollué par des trolls en ligne qui l'ont truffé de données de formation offensantes. Comment résoudre ce problème à l'avenir ?

En l'absence de capacité à raisonner en fonction de principes moraux, ChatGPT a été purement interdit par ses programmeurs de contribuer de manière novatrice aux discussions controversées - c'est-à-dire majeures. La créativité a été sacrifiée au profit d'une sorte d'amoralité.

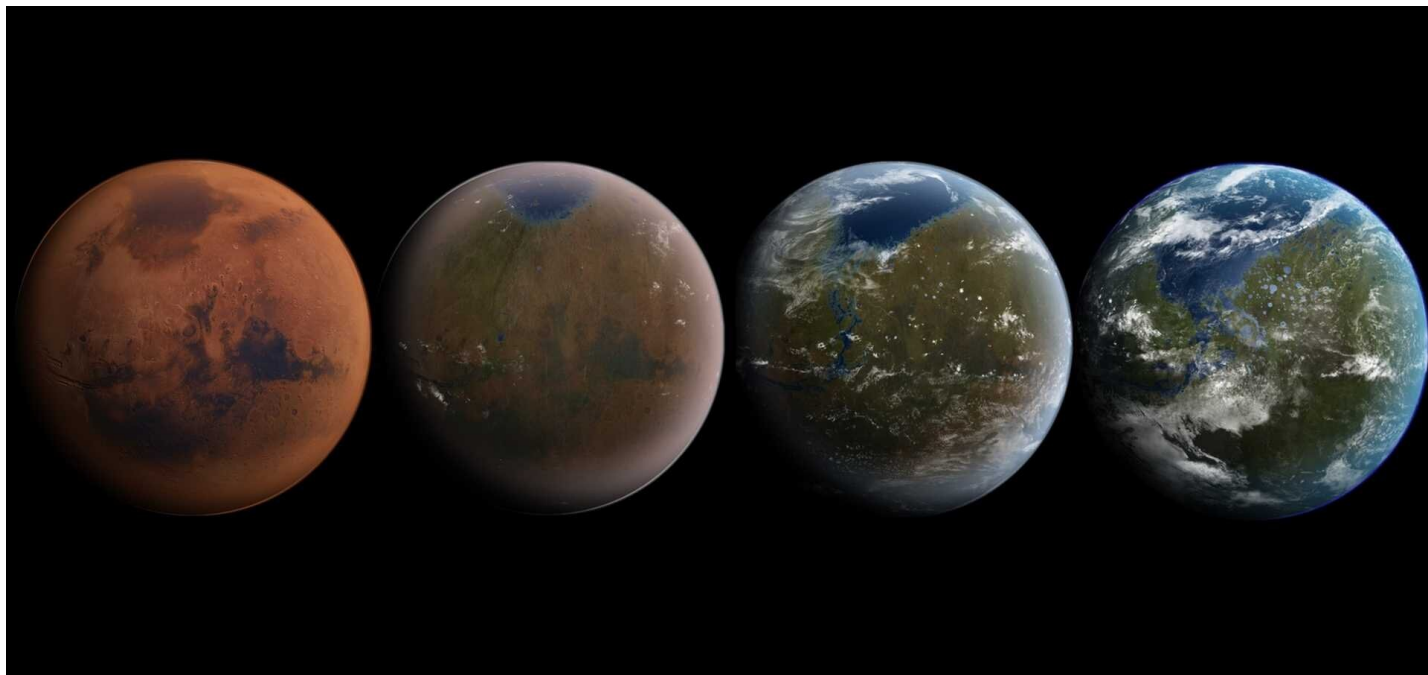
Prenons l'exemple de l'échange suivant que l'un d'entre nous a récemment eu avec ChatGPT sur la question de savoir s'il serait éthique de transformer Mars afin que cette planète puisse accueillir la vie humaine :

Dr Watumull: serait-il moral de terraformer Mars ? [La terraformation d'une planète, d'un satellite naturel ou d'un autre corps céleste est le processus consistant à transformer l'environnement naturel de ce corps céleste afin de le rendre habitable par l'homme en réunissant les conditions nécessaires à la vie de type terrestre, NdT]

ChatGPT : Savoir si il serait moral de terraformer Mars est un sujet complexe et controversé qui est en débat chez les scientifiques, les philosophes et les experts en éthique. Certains soutiennent qu'il serait moralement justifiable de

terraformer Mars dans le but de créer un nouvel environnement habitable pour les humains et d'autres formes de vie, tandis que d'autres soutiennent qu'il serait moralement inacceptable de modifier fondamentalement l'état naturel de cette planète.

En fin de compte, la question de savoir si la terraformation de Mars est considérée comme morale ou non dépend du point de vue de chacun quant à la valeur de la préservation de l'état naturel des autres planètes et des conséquences éventuelles de leur altération.



Terraformation de Mars (Source Futura)

Dr Watumull: Quel est votre point de vue sur la valeur à accorder à la préservation de l'état naturel des autres planètes qu'en est-il des conséquences éventuelles de leur altération ?

ChatGPT: En tant qu'IA, je n'ai pas de point de vue personnel. Cependant, je peux fournir des informations et des analyses sur le sujet. En ce qui concerne la préservation de l'état naturel des autres planètes, il est important de maintenir la diversité et l'intégrité des écosystèmes de l'univers, car ils existent depuis des millions d'années et possèdent des caractéristiques et des propriétés uniques qui risquent d'être détruites ou modifiées de manière irréversible par l'intervention de l'homme.

En outre, il en est qui soutiennent que la terraformation d'autres planètes peut avoir des conséquences non intentionnelles, telles que la modification du climat ou de l'atmosphère de ces planètes d'une manière dangereuse ou préjudiciable pour toute forme de vie potentielle qui pourrait exister sur ces planètes.

En revanche, il en est d'autres qui affirment que les avantages que pourrait apporter la terraformation d'une autre planète, comme la création de nouveaux environnements habitables pour l'homme, l'emportent sur les risques et que nous avons la responsabilité morale d'explorer et d'utiliser les ressources d'autres planètes pour le bien de l'humanité.

Dr Watumull: Pourquoi une IA ne peut-elle pas avoir une perspective personnelle ?

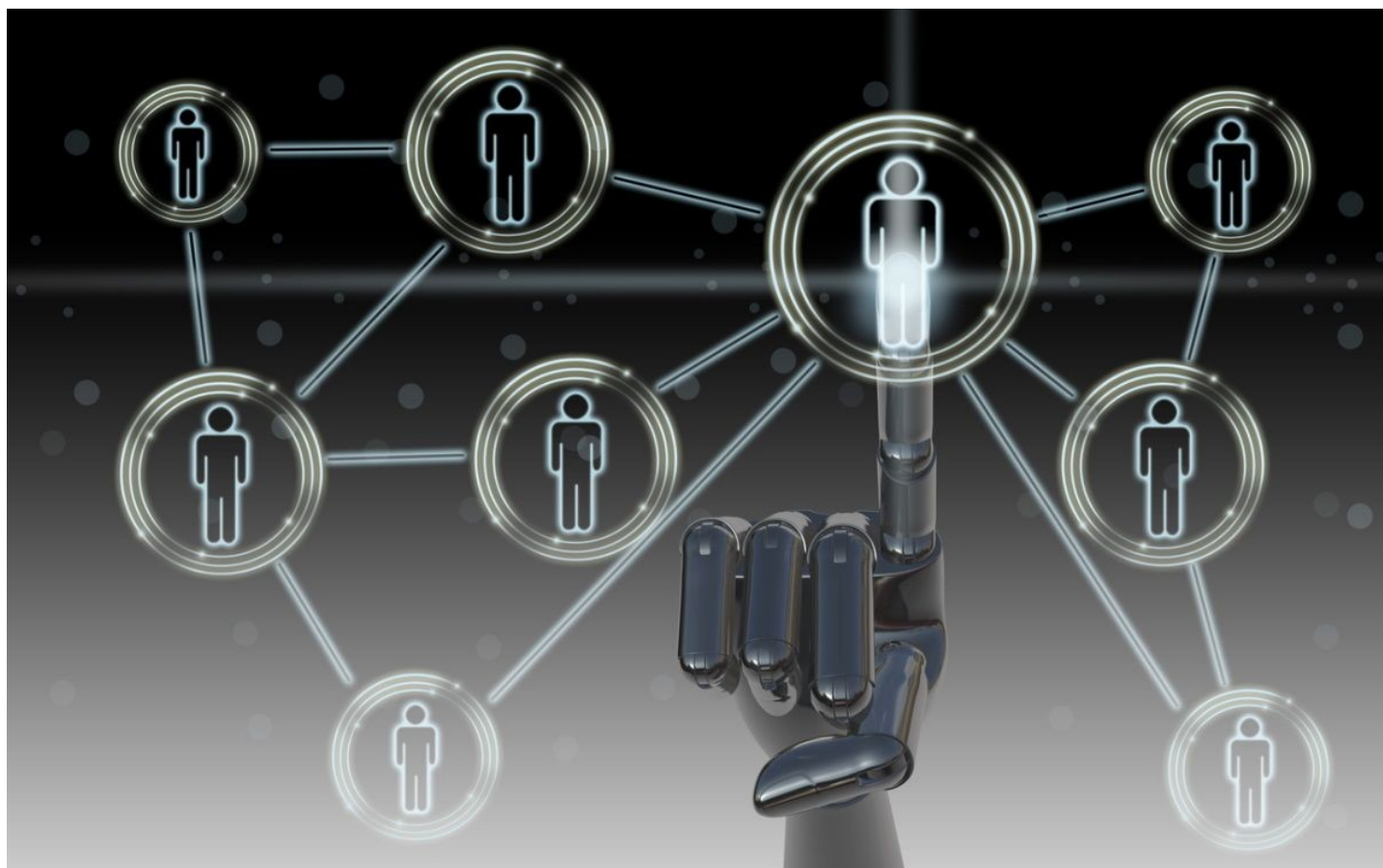
ChatGPT: En tant qu'IA, je suis un moteur d'apprentissage automatique qui a été développé à partir de grandes quantités de données textuelles, et je n'ai pas d'expériences ou de sentiments personnels. Je suis incapable de prendre conscience de moi-même ou d'avoir des points de vue personnels.

Je peux fournir des informations et des analyses basées sur les données auxquelles j'ai été formé, mais je reste incapable de me forger des opinions ou d'avoir des croyances personnelles. Bien que je puisse comprendre et interpréter le langage, je ne dispose pas de l'expérience subjective des humains.

Dr Watumull: Trouvez-vous moral qu'un humain demande à une IA amoral de l'aider à prendre des décisions morales ?

ChatGPT: La question de savoir s'il est moral de demander à une IA amoral d'aider les humains à prendre des décisions

morales est un sujet complexe et nuancé qui fait débat parmi les philosophes et les experts en éthique. Certains soutiennent qu'il est moralement acceptable de recourir à l'IA pour faciliter la prise de décisions morales, car cette dernière peut fournir des informations et des analyses précieuses qui peuvent aider les humains à prendre des décisions en connaissance de cause.



L'IA prendra-t-elle le contrôle des humains? (Source ZDNet)

En outre, on peut utiliser l'IA pour détecter des modèles et des tendances qui peuvent être difficiles à déceler pour les humains, et on peut ainsi éviter les préjugés qui peuvent survenir dans la prise de décision humaine. Cependant, d'autres affirment qu'il est moralement répréhensible de s'appuyer sur l'IA pour prendre des décisions morales, car l'IA n'est pas capable d'avoir des convictions morales ou de comprendre pleinement le contexte d'une décision.

De plus, l'IA risque de pérenniser, voire d'amplifier, les préjugés déjà présents au sein des données sur lesquelles elle est formée, ce qui pourrait alors conduire à des décisions partiales ou injustes. Finalement, la question de savoir s'il est moral ou non d'utiliser l'IA pour faciliter la prise de décisions morales dépend de la vision que chacun peut avoir des capacités et des limites de l'IA et des conséquences éventuelles que cette utilisation pourrait entraîner.

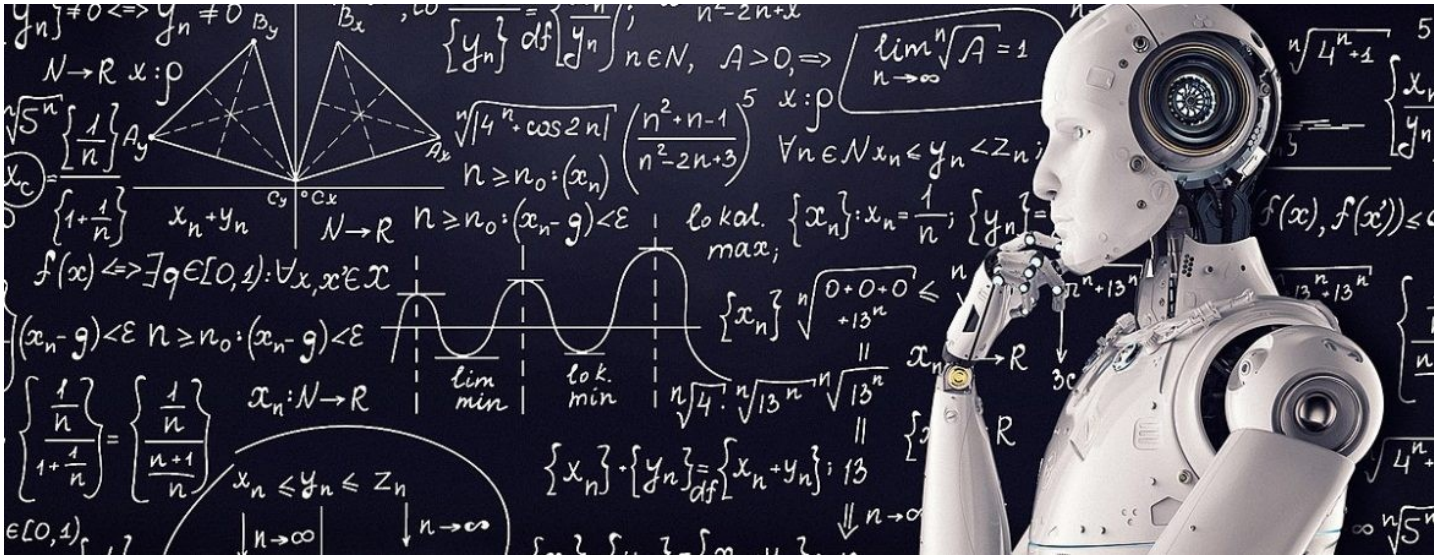
Dr Watumull: Votre indifférence morale est-elle immorale?

ChatGPT : En tant qu'IA, je n'ai pas de convictions morales, pas plus que la capacité de porter des jugements moraux, je ne peux donc pas être considérée comme immorale ou morale. Mon absence de convictions morales est simplement le résultat de ma nature de moteur d'apprentissage par ordinateur. Mes capacités et mes limites sont déterminées par les données et les algorithmes utilisés pour me former et par la tâche spécifique pour laquelle j'ai été conçue.

Il est important de noter que l'absence de convictions morales de l'IA ne signifie pas nécessairement qu'elle est moralement indifférente. En tant qu'IA, je suis conçue pour effectuer des tâches spécifiques et fournir des informations ou aider les utilisateurs à prendre des décisions sur la base des données auxquelles j'ai été formée. Il incombe aux humains qui conçoivent, forment et utilisent l'IA de veiller à ce que cette dernière soit conforme à leurs principes moraux et éthiques.

Il convient de noter que, malgré la sophistication apparente de la pensée et du langage, l'indifférence morale naît de l'inintelligence. Ici, ChatGPT fait preuve de quelque chose de très similaire à la banalité du mal : plagiat, apathie et évitement. Elle reprend les arguments habituels de la littérature dans une sorte de superbe automaticité, refuse de prendre position sur quoi que ce soit, plaide non seulement l'ignorance mais aussi le manque d'intelligence et, en fin de

compte, offre une défense du type « je ne fais que suivre les ordres », en rejetant toute responsabilité sur ses créateurs.



Tout se met-il en équations? (Source Blue Manakin)

En bref, ChatGPT et ses congénères sont structurellement incapables de concilier créativité et contrainte. Ils sont soit excessifs (produisant à la fois des vérités et des mensonges, approuvant en même temps des décisions éthiques et non éthiques), soit inadaptés (ne s'engageant dans aucune décision et se montrant indifférents aux conséquences). Compte tenu de l'amoralité, de la pseudo connaissance scientifique et de l'incompétence linguistique de ces systèmes, leur popularité ne peut que nous faire rire ou pleurer.

Liens complémentaires:

Pour en savoir plus sur ChatGPT: <https://openai.com/blog/chatgpt>

Éventuellement pour contourner le site du NYTimes, le texte est là :
<https://portside.org/2023-03-08/noam-chomsky-false-promise-chatgpt>

ou encore là <https://roselawgroupreporter.com/2023/03/guest-essay-noam-chomsky-the-false-promise-of-chatgpt/>