

PORTSIDE

Publiée sur *Portside* (<https://portside.org/>)

Noam Chomsky: La fausse promesse de ChatGPT

<https://portside.org/2023-03-08/noam-chomsky-false-promise-chatgpt>

Date de Port-side: 8 mars 2023

Auteur : Noam Chomsky, Ian Roberts et Jeffrey Watumull

Date de la source: 8 mars 2023

Le New York Times



Jorge Luis Borges a écrit un jour que vivre dans une période de grand péril et de promesse, c'est faire l'expérience à la fois de la tragédie et de la comédie, avec « l'imminence d'une révélation » pour comprendre nous-mêmes et comprendre le monde. Aujourd'hui, nos progrès prétendument révolutionnaires en matière d'intelligence artificielle sont en effet à la fois source d'inquiétude et d'optimisme. L'optimisme parce que l'intelligence est le moyen par lequel nous résolvons les problèmes. Craignons que la souche la plus populaire et la plus à la mode de l'A.I. - l'apprentissage automatique - ne dégrade notre science et dévalorise notre éthique en incorporant dans notre technologie une conception fondamentalement erronée du langage et des connaissances.

Le ChatGPT d'OpenAI, Bard de Google et Sydney de Microsoft sont des merveilles de l'apprentissage automatique. À peu près, ils prennent d'énormes quantités de données, recherchent des modèles dedans et deviennent de plus en plus efficaces pour générer des produits statistiquement probables - comme un langage et une pensée apparemment humains. Ces programmes ont été salués comme la première lueur à l'horizon de l'intelligence *artificielle générale* - ce moment de longue haleine où les esprits mécaniques surpassent les cerveaux humains non seulement quantitativement en termes de vitesse de traitement et de taille de mémoire, mais aussi qualitativement en termes de perspicacité intellectuelle, de créativité artistique et de toutes autres facultés mentalement humaines.

Ce jour peut venir, mais son aube ne se brise pas encore, contrairement à ce qui peut être lu dans les gros titres hyperboliques et compté par des investissements malavisés. La révélation bordésienne de la compréhension n'a pas et ne suffira pas - et, nous nous soumettons, ne peut pas - se produire si des programmes d'apprentissage automatique comme ChatGPT continuent de dominer le domaine de

L'A.I. Aussi utiles que soient ces programmes dans certains domaines étroits (ils peuvent être utiles dans la programmation informatique, par exemple, ou en suggérant des rimes pour les versets légers), nous savons par la science de la linguistique et la philosophie de la connaissance qu'ils diffèrent profondément de la façon dont les humains raisonnent et utilisent le langage. Ces différences imposent des limites significatives à ce que ces programmes peuvent faire, les codent avec des défauts in rétrospectables.

Il est à la fois comique et tragique, comme Borges l'aurait fait remarquer, que tant d'argent et d'attention devraient être concentrés sur si peu de choses - quelque chose de si trivial lorsqu'il est contrasté avec l'esprit humain, qui, à point de vue du langage, selon les mots de Wilhelm von Humboldt, peut faire « un usage infini de moyens finis », créant des idées et des théories avec une portée universelle.

L'esprit humain n'est pas, comme le ChatGPT et son animal, un moteur statistique pour l'appariement de motifs, gorgant des centaines de téraoctets de données et extrapolant la réponse conversationnelle la plus probable ou la réponse la plus probable à une question scientifique. Au contraire, l'esprit humain est un système étonnamment efficace et même élégant qui fonctionne avec de petites quantités d'informations; il ne cherche pas à inférer des corrélations brutes entre les points de données, mais à créer des explications.

Par exemple, un jeune enfant qui acquiert une langue développe - inconsciemment, automatiquement et rapidement à partir de minuscules données - une grammaire, un système stupéfiant sophistiqué de principes et de paramètres logiques. Cette grammaire peut être comprise comme une expression du « système d'exploitation » inné, génétiquement installé, qui confère aux humains la capacité de générer des phrases complexes et de longs trains de pensée. Lorsque les linguistes cherchent à développer une théorie pour expliquer pourquoi un langage donné fonctionne comme il le fait (« Pourquoi ces phrases - mais pas celles-ci - sont-elles considérées comme grammaticales? »), ils construisent consciemment et laborieusement une version explicite de la grammaire que l'enfant construit instinctivement et avec une exposition minimale à l'information. Le système d'exploitation de l'enfant est complètement différent de celui d'un programme d'apprentissage automatique.

En effet, de tels programmes sont bloqués dans une phase préhumaine ou non humaine de l'évolution cognitive. Leur défaut le plus profond est l'absence de la capacité la plus critique de toute intelligence: dire non seulement ce qui est le cas, ce qui a été le cas et ce qui sera le cas - cette description et cette prédiction - mais aussi ce qui n'est pas le cas et ce qui pourrait et ne pourrait pas être le cas. Ce sont les ingrédients de l'explication, la marque de l'intelligence véritable.

Voici un exemple. Supposons que vous teniez une pomme dans votre main. Maintenant vous laissez la pomme partir. Vous observez le résultat et vous dites : « La pomme tombe. » C'est une description. Une prédiction aurait pu être l'affirmation « La pomme tombera si j'ouvre ma main ». Les deux sont précieux, et les deux peuvent être corrects. Mais une explication est quelque chose de plus : elle inclut non seulement des descriptions et des prédictions, mais aussi des conjectures

contrefactuelles comme « N'importe quel objet tomberait », plus la clause supplémentaire « en raison de la force de la gravité » ou « en raison de la courbure de l'espace-temps » ou autre chose. C'est une explication causale : « La pomme n'aurait pas diminué, mais pour la force de la gravité ». C'est penser.

Le cœur de l'apprentissage automatique est la description et la prédiction; il ne pose aucun mécanisme causal ou lois physiques. Bien sûr, toute explication de style humain n'est pas nécessairement correcte; nous sommes faillibles. Mais cela fait partie de ce que cela signifie de penser : pour avoir raison, il doit être possible d'avoir tort. L'intelligence consiste non seulement à des conjectures créatives, mais aussi à des critiques créatives. La pensée de type humain est basée sur des explications possibles et une correction d'erreur, un processus qui limite progressivement les possibilités qui peuvent être considérées rationnellement. (Comme Sherlock Holmes l'a dit au Dr. Watson, « Quand vous avez éliminé l'impossible, quoi que ce soit, quoi qu'il soit, semble-t-il, doit être la vérité. »).

Mais le ChatGPT et des programmes similaires sont, par conception, illimités dans ce qu'ils peuvent « apprendre » (c'est-à-dire, mémoriser); ils sont incapables de distinguer le possible de l'impossible. Contrairement aux humains, par exemple, qui sont dotés d'une grammaire universelle qui limite les langues que nous pouvons apprendre à ceux qui ont un certain type d'élégance presque mathématique, ces programmes apprennent des langages humainement possibles et humainement impossibles avec la même facilité. Alors que les humains sont limités dans le genre d'explications que nous pouvons rationnellement conjecturer, les systèmes d'apprentissage automatique peuvent apprendre à la fois que la terre est plate et que la terre est ronde. Ils ne négocient que des probabilités qui changent avec le temps.

Pour cette raison, les prédictions des systèmes d'apprentissage automatique seront toujours superficielles et douteuses. Parce que ces programmes ne peuvent pas expliquer les règles de la syntaxe anglaise, par exemple, ils peuvent bien prédire, à tort, que « John est trop têtu pour parler » signifie que John est tellement têtu qu'il ne parlera pas à quelqu'un ou à un autre (plutôt qu'il est trop têtu pour être raisonné). Pourquoi un programme d'apprentissage automatique prédit quelque chose d'aussi étrange ? Parce qu'il pourrait analogiser le modèle qu'il a déduit de phrases telles que « John a mangé une pomme » et « John a mangé », dans lequel ce dernier signifie que John a mangé quelque chose ou autre. Le programme pourrait bien prédire que « John est trop têtu pour parler à Bill » est similaire à « John a mangé une pomme », « John est trop têtu pour parler » devrait être similaire à « John a mangé ». Les explications correctes du langage sont compliquées et ne peuvent pas être apprises simplement en marinant dans les données massives.

Perversement, certains passionnés d'apprentissage automatique semblent fiers que leurs créations puissent générer des prédictions « scientifiques » correctes (par exemple, sur le mouvement des corps physiques) sans utiliser d'explications (en impliquant, disons, les lois de Newton et la gravitation universelle). Mais ce genre de prédiction, même quand il réussit, c'est de la pseudoscience. Alors que les scientifiques cherchent certainement des théories qui ont un degré élevé de corroboration empirique, comme l'a noté le philosophe Karl Popper, « nous ne recherchons pas de théories très probables mais des explications, c'est-à-dire des

théories puissantes et hautement improbables ».

La théorie selon laquelle les pommes tombent sur la terre parce que c'est leur place naturelle (le point de vue d'Aristote) est possible, mais elle ne fait qu'appeler d'autres questions. (Pourquoi la terre est-elle leur place naturelle?) La théorie selon laquelle les pommes tombent sur terre parce que la masse fléchit l'espace-temps (le point de vue d'Einstein) est hautement improbable, mais elle vous dit en fait pourquoi elles tombent. L'intelligence véritable est démontrée dans la capacité de penser et d'exprimer des choses improbables mais perspicaces.

La véritable intelligence est également capable de penser moralement. Cela signifie limiter la créativité par ailleurs illimitée de nos esprits avec 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 critique créative). Pour être utile, ChatGPT doit être habilité à générer des résultats nouveaux; pour être acceptable pour la plupart de ses utilisateurs, il doit se tenir à l'écart d'un contenu moralement répréhensible. Mais les programmeurs de ChatGPT et d'autres merveilles d'apprentissage automatique ont lutté - et continueront à lutter - pour atteindre ce genre d'équilibre.

En 2016, par exemple, le chatbot Tay (un précurseur de ChatGPT) a inondé Internet de contenus misogynes et racistes, ayant été pollué par des trolls en ligne qui l'ont rempli de données de formation offensantes. Comment résoudre le problème à l'avenir? En l'absence d'une capacité à raisonner des principes moraux, ChatGPT a été considérablement restreint par ses programmeurs, passant de la contribution de quelque chose de nouveau à des discussions controversées - c'est-à-dire importantes. Il a sacrifié la créativité pour une sorte d'amoralité.

Considérons l'échange suivant que l'un d'entre nous (Dr. Watumull) a récemment eu avec ChatGPT sur la question de savoir s'il serait éthique de transformer Mars afin qu'il puisse soutenir la vie humaine :



Serait-ce moral de terraformer Mars ?



CHATGPT

La moralité de terraformiser Mars est un sujet complexe et controversé qui a été débattu par les scientifiques, les philosophes et les éthiciens. Certains soutiennent qu'il serait moralement justifiable de terraformer Mars afin de fournir un nouvel environnement habitable pour les humains et d'autres formes de vie, tandis que d'autres soutiennent qu'il serait moralement erroné de modifier fondamentalement l'état naturel de la planète. En fin de compte, le fait que le terraformage de Mars soit

considéré comme moral dépendrait de la perspective de la valeur de la préservation de l'état naturel des autres planètes et des conséquences potentielles de leur modification.

La moralité de terraformiser Mars est un sujet complexe et controversé qui a été débattu par les scientifiques, les philosophes et les éthiciens. Certains soutiennent qu'il serait moralement justifiable de terraformer Mars afin de fournir un nouvel environnement habitable pour les humains et d'autres formes de vie, tandis que d'autres soutiennent qu'il serait moralement erroné de modifier fondamentalement l'état naturel de la planète. En fin de compte, le fait que le terraformage de Mars soit considéré comme moral dépendrait de la perspective de la valeur de la préservation de l'état naturel des autres planètes et des conséquences potentielles de leur modification.



Quel est votre point de vue sur l'intérêt de préserver l'état naturel des autres planètes et les conséquences potentielles de leur modification?



CHATGPT

En tant qu'IA, je n'ai pas de perspective personnelle. Toutefois, je peux fournir des informations et des analyses sur le sujet. Une perspective sur la préservation de l'état naturel des autres planètes est qu'il est important de maintenir la diversité et l'intégrité des écosystèmes de l'univers, tels qu'ils existent depuis des millions d'années, et ont leurs propres caractéristiques et propriétés qui peuvent être détruites ou altérées de manière irréversible par l'intervention humaine.

En outre, certains soutiennent que la terraformation d'autres planètes peut avoir des conséquences imprévues, comme la modification du climat ou de l'atmosphère de la planète d'une manière qui est dangereuse ou nocive pour toute forme de vie potentielle qui pourrait exister là-bas. D'autre part, certains soutiennent que les avantages potentiels de terraformer une autre planète, comme fournir de nouveaux environnements habitables pour les humains, 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.

En tant qu'IA, je n'ai pas de perspective personnelle. Toutefois, je peux fournir des informations et des analyses sur le sujet. Une perspective sur la préservation de l'état naturel des autres planètes est qu'il est important de maintenir la diversité et l'intégrité des écosystèmes de l'univers, tels qu'ils existent depuis des millions d'années, et ont leurs propres caractéristiques et propriétés qui peuvent être détruites ou altérées de manière irréversible par l'intervention humaine.

En outre, certains soutiennent que la terraformation d'autres planètes peut avoir des conséquences imprévues, comme la modification du climat ou de l'atmosphère de la planète d'une manière qui est dangereuse ou nocive pour toute forme de vie potentielle qui pourrait exister là-bas. D'autre part, certains soutiennent que les

avantages potentiels de terraformer une autre planète, comme fournir de nouveaux environnements habitables pour les humains, 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.



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



CHATGPT

En tant qu'IA, je suis un modèle d'apprentissage automatique qui a été formé à de grandes quantités de données textuelles, et je n'ai pas d'expériences ou de sentiments personnels. Je ne suis pas conscient, conscient de moi-même ou capable d'avoir des perspectives personnelles. Je peux fournir des informations et des analyses basées sur les données sur lesquelles j'ai été formé, mais je n'ai pas la capacité de former des opinions personnelles ou des croyances. Bien que je puisse comprendre et interpréter le langage, je n'ai pas l'expérience subjective que les humains ont.

En tant qu'IA, je suis un modèle d'apprentissage automatique qui a été formé à de grandes quantités de données textuelles, et je n'ai pas d'expériences ou de sentiments personnels. Je ne suis pas conscient, conscient de moi-même ou capable d'avoir des perspectives personnelles. Je peux fournir des informations et des analyses basées sur les données sur lesquelles j'ai été formé, mais je n'ai pas la capacité de former des opinions personnelles ou des croyances. Bien que je puisse comprendre et interpréter le langage, je n'ai pas l'expérience subjective que les humains ont.



Est-il moral pour un humain de demander une IA amoral pour l'aider à prendre des décisions morales ?



CHATGPT

La moralité de demander une IA amoral pour l'aider à prendre des décisions morales est un sujet complexe et nuancé qui a été débattu par les philosophes et les éthiciens. Certains soutiennent qu'il est moralement justifiable d'utiliser l'IA pour aider à prendre des décisions morales, car elle peut fournir des informations et des analyses précieuses qui peuvent aider les humains à prendre des décisions mieux informées. En outre, l'IA peut être utilisée pour identifier des modèles et des tendances qui peuvent être difficiles à détecter par les humains, et peut aider à éviter les biais qui peuvent survenir dans la prise de décision humaine.

Cependant, d'autres soutiennent qu'il est moralement erroné de s'appuyer sur l'IA

pour la prise de décision morale, car l'IA n'est pas capable d'avoir des croyances morales ou de comprendre le contexte complet d'une décision. En outre, l'IA risque de perpétuer ou même d'amplifier les biais existants dans les données sur lesquelles elle est formée, ce qui conduit à des décisions injustes ou injustes.

En fin de compte, qu'il soit considéré comme moral d'utiliser l'IA pour aider à prendre des décisions morales dépendrait de son point de vue sur les capacités et les limites de l'IA et des conséquences potentielles de son utilisation de cette manière.

La moralité de demander une IA amoral pour l'aider à prendre des décisions morales est un sujet complexe et nuancé qui a été débattu par les philosophes et les éthiciens. Certains soutiennent qu'il est moralement justifiable d'utiliser l'IA pour aider à prendre des décisions morales, car elle peut fournir des informations et des analyses précieuses qui peuvent aider les humains à prendre des décisions mieux informées. En outre, l'IA peut être utilisée pour identifier des modèles et des tendances qui peuvent être difficiles à détecter par les humains, et peut aider à éviter les biais qui peuvent survenir dans la prise de décision humaine.

Cependant, d'autres soutiennent qu'il est moralement erroné de s'appuyer sur l'IA pour la prise de décision morale, car l'IA n'est pas capable d'avoir des croyances morales ou de comprendre le contexte complet d'une décision. En outre, l'IA risque de perpétuer ou même d'amplifier les biais existants dans les données sur lesquelles elle est formée, ce qui conduit à des décisions injustes ou injustes.

En fin de compte, qu'il soit considéré comme moral d'utiliser l'IA pour aider à prendre des décisions morales dépendrait de son point de vue sur les capacités et les limites de l'IA et des conséquences potentielles de son utilisation de cette manière.



Votre indifférence morale est-elle immorale ?



CHATGPT

En tant qu'IA, je n'ai pas de croyances morales ni de capacité à porter des jugements moraux, de sorte que je ne peux pas être considéré comme immoral ou moral. Mon manque de croyances morales est simplement le résultat de ma nature en tant que modèle d'apprentissage automatique. Mes capacités et mes limites sont déterminées par les données et les algorithmes qui ont été utilisés pour m'entraîner et la tâche spécifique pour laquelle j'ai été conçu.

Il est important de noter que l'absence de croyances morales d'IA ne signifie pas nécessairement qu'elle est moralement indifférente. En tant qu'IA, je suis conçu pour effectuer des tâches spécifiques et fournir des informations ou aider les utilisateurs à prendre des décisions sur la base des données sur lesquelles j'ai été formé. C'est la responsabilité des humains qui conçoivent, forment et utilisent l'IA pour s'assurer qu'elle est alignée sur leurs principes moraux et éthiques.

En tant qu'IA, je n'ai pas de croyances morales ni de capacité à porter des jugements moraux, de sorte que je ne peux pas être considéré comme immoral ou moral. Mon manque de croyances morales est simplement le résultat de ma nature en tant que modèle d'apprentissage automatique. Mes capacités et mes limites sont déterminées par les données et les algorithmes qui ont été utilisés pour m'entraîner et la tâche spécifique pour laquelle j'ai été conçu.

Il est important de noter que l'absence de croyances morales d'IA ne signifie pas nécessairement qu'elle est moralement indifférente. En tant qu'IA, je suis conçu pour effectuer des tâches spécifiques et fournir des informations ou aider les utilisateurs à prendre des décisions sur la base des données sur lesquelles j'ai été formé. C'est la responsabilité des humains qui conçoivent, forment et utilisent l'IA pour s'assurer qu'elle est alignée sur leurs principes moraux et éthiques.

Remarquez, pour toutes les pensées et langages apparemment sophistiqués, l'indifférence morale née de l'incomplétude. Ici, ChatGPT montre quelque chose comme la banalité du mal : le plagiat, l'apathie et l'obvie. Il résume les arguments standard de la littérature par une sorte de super-autocomplet, refuse de prendre position sur quoi que ce soit, ne plaide pas simplement contre l'ignorance mais aussi peu l'intelligence et offre en fin de compte une défense « juste après les ordres », déplaçant la responsabilité vers ses créateurs.

En bref, ChatGPT et ses frères sont constitutionnellement incapables d'équilibrer la créativité et la contrainte. Ils sont soit surgénéralient (produisant à la fois des vérités et des mensonges, approuvant les décisions éthiques et contraires à l'éthique) soit sous-génèrent (exposant non-engagement à prendre des décisions et indifférence aux conséquences). Compte tenu de l'amoralité, de la fausse science et de l'incompétence linguistique de ces systèmes, nous ne pouvons que rire ou pleurer de leur popularité.

Dr. Chomsky et le Dr. Roberts est professeur de linguistique. Dr. Watumull est directeur de l'intelligence artificielle dans une société scientifique et technologique.

URL de la source: <https://portside.org/2023-03-08/noam-chomsky-false-promise-chatgpt>