



IA générative

mais qui a fait ça ?

Priscilla Haring-Kuipers (Pays-Bas)

DALL-E et ChatGPT – tous deux créés par OpenAI – fascinent et pénètrent tous les secteurs. L'IA générative devient un débat de société.

IA comme Inspiration Artificielle ?

Une des critiques adressées aux IA génératives d'images telles que DALL-E [1] est qu'elles ont été entraînées avec des œuvres sans que leurs auteurs aient donné leur autorisation. Les artistes ont toujours été inspirés par d'autres artistes. Copier une œuvre est même considéré comme une pratique formatrice, à l'artiste ensuite de trouver sa propre voie. Alors pourquoi une IA ne pourrait-elle pas faire la même chose ? Une des raisons est que ces IA n'existaient pas au moment où l'artiste a mis ses œuvres en ligne, qu'il ne l'avait pas fait pour entraîner une IA, et que personne ne lui a demandé son accord pour ça. Les règles ont été modifiées du jour au lendemain à son insu, et nous devons maintenant déterminer ensemble quelles règles sont acceptables et comment les appliquer.

Une autre raison est que l'IA ne s'inspire pas de l'art. Elle collecte un jeu de données à partir duquel elle construit un modèle. Nous appréhendons l'intelligence à travers le prisme de nos propres expériences, et nous nous attendons à ce que l'IA fonctionne plus ou moins comme nous. Or, ce n'est pas le cas. Nous

Source : P. Haring-Kuipers et DALL-E, avec l'invite AI working on a painting.

sommes des êtres multimodaux traitant l'information de façon directe et indirecte, notre apprentissage est associatif, et notre pensée est influencée par des facteurs tels que les filtres émotionnels et la neurochimie. L'IA peut traiter plus de données que nous et plus vite, mais elle le fait différemment. Il faut des années à l'humain avant que ses acquis intellectuels et émotionnels ne lui permettent de créer ce que nous jugeons être de l'art. L'IA ne se développera pas comme nous si nous ne faisons que lui fournir plus de temps ou de puissance de calcul. Nous façonnons l'IA.

Une IA à notre image

Les techniques d'exploration de données (de *data mining*) servent à construire des modèles capables de mimer nos comportements. Dans le cas du modèle linguistique sous-jacent à ChatGPT [2], les données sont le gigantesque corpus textuel laissés par vous et moi sur l'internet. Si ces modèles vous intéressent, Hannah Fry les explique dans l'épisode *Speaking of Intelligence* de sa série de podcasts DeepMind [3].

Un des problèmes de ChatGPT est d'être (involontairement) raciste, agressif et odieux. Il n'a aucun système de valeurs intégré pour juger du contenu servant à son entraînement. Alors d'où viennent ses « valeurs morales » ? D'humains. De travailleurs (sous-)payés à visionner des images et des textes immoraux, et à les étiqueter comme tels afin que le modèle apprenne à les reconnaître et cesse de les produire [4]. Le contenu des médias sociaux est modéré de la même façon. L'amélioration d'un modèle passe donc par un incessant travail d'étiquetage avec des mots comme « viol », « contenu raciste », etc. L'intelligence de l'IA est en partie la nôtre.

Autre problème, les faits énoncés par ChatGPT sont souvent incorrects. Le site *Stack Overflow* interdit de poster des réponses fournies par GPT, car « même si celles-ci ont un taux élevé d'erreurs, elles donnent souvent l'impression d'être correctes et sont très faciles à produire par quiconque. » [5] Nous interagissons avec l'IA sur le mode de la conversation, et avons donc tendance à traiter ses réponses de façon relâchée et à les juger plus vraisemblables – si tant est que nous les évaluons vraiment. En utilisant ChatGPT, gardez à l'esprit que vous parlez à la médiane de l'internet, et ne lui faites confiance qu'en conséquence.

Un revenu universel comme solution


Nous pourrions atténuer les effets d'une mainmise de l'IA sur le travail humain en versant à chacun un revenu universel. Assez d'argent pour vivre. Sans conditions, sans questions.

Is it safe to use ChatGPT for your task?

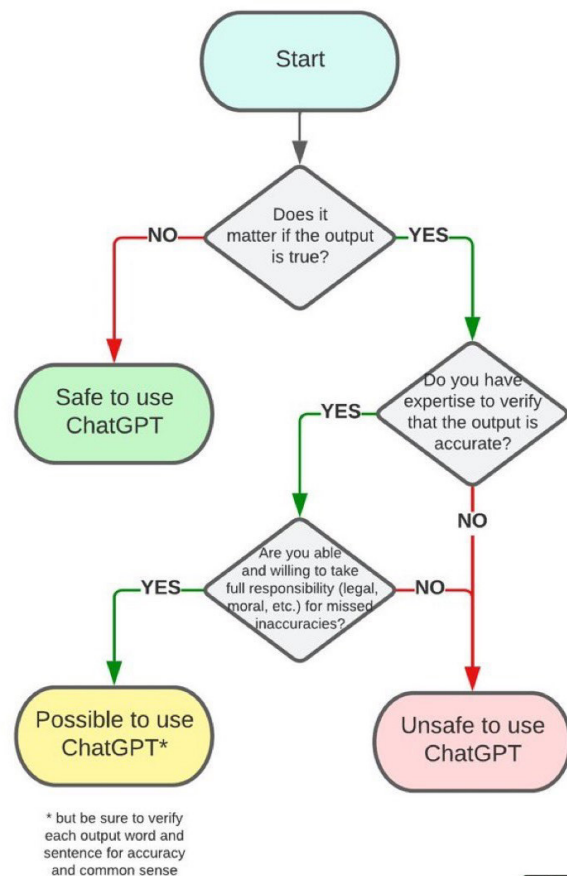
Aleksandr Tiulkanov | January 19, 2023

Les IA génératives ne menacent pas la créativité de l'artiste, elles menacent ses revenus. Je pourrais utiliser l'IA pour créer des visuels accompagnant mes articles – et non pas me les procurer auprès d'une banque d'images – mais je ne veux pas qu'une IA les écrive. Je veux les écrire moi-même. Par souci d'économie, mon éditeur pourrait toutefois décider l'inverse. Ce genre de décision peut affecter une vie professionnelle et personnelle à des degrés divers.

Ne vous y trompez pas : aucune profession n'est à l'abri de l'IA. Graphistes, auteurs, avocats, codeurs, ingénieurs, tous verront bientôt une partie de leur travail remplacée par une IA générative. Nous pourrions utiliser l'IA comme outil, et nous concentrer avec profit sur ce qui nous motive réellement en garantissant nos moyens de subsistance par un revenu universel. Nous aurions alors tout loisir de réfléchir à une division du travail entre ce qu'humains et IA font de mieux chacun de leur côté.

Face à l'IA comme outil, il est clair que ce qui nous revient à nous, humains, c'est la prise de décision. Décision d'appliquer (ou non) ce que l'IA produit, ou de relancer le programme après avoir modifié l'invite ou certains paramètres. Nous sommes également responsables de l'exploitation des résultats d'une IA, des informations qu'elle livre et de leur partage, p. ex. dans un cadre médical. 

VF : Hervé Moreau — 230048-04



Logigramme de ChatGPT selon Aleksandr Tiulkanov, chercheur en IA.

L'auteure

Priscilla Haring-Kuipers est titulaire d'un master en Psychologie des médias et directrice générale de *This Is Not Rocket Science*. Convaincue des bienfaits de la recherche et de la technologie sur le devenir de l'humanité, elle aborde ses articles sous l'angle des sciences sociales.

Forum mondial de l'électronique éthique

Le *Forum mondial de l'électronique éthique* (WEEF) se veut un lieu d'inspiration, de débats et de publications sur l'éthique et les objectifs de développement durable. Pour en savoir plus, visitez worldethicalelectronicsforum.com.

**WORLD
ETHICAL
ELECTRONICS
FORUM**



LIENS

- [1] OpenAI, DALL-E 2 : <https://openai.com/dall-e-2/>
- [2] OpenAI, ChatGPT: Optimizing Language Models for Dialogue, 30 novembre 2022 : <https://openai.com/blog/chatgpt/>
- [3] H. Fry, Speaking of intelligence, podcast DeepMind (S2, Ep2), 25 janvier 2022 : <https://youtu.be/21JSKHR7KWw>
- [4] B. Perrigo, OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic, 18 janvier 2023 : <https://time.com/6247678/openai-chatgpt-kenya-workers/>
- [5] Stack Overflow, Temporary Policy: ChatGPT Is Banned, 5 décembre 2022 : <https://meta.stackoverflow.com/questions/421831/temporary-policy-chatgpt-is-banned>