

Intelligence Artificielle : mieux vaut tuer 10 000 personnes que permettre une insulte raciale

écrit par Christine Tasin | 21 mars 2023

ChatGPT et les autres LLM (grands modèles de langage) vont structurer la pensée humaine et l'économie mondiale, compte tenu de leur extraordinaire puissance.



ChatGPT et les autres LLM (grands modèles de langage) vont structurer la pensée humaine et l'économie mondiale, compte tenu de leur extraordinaire puissance.



C'est un exemple donné par le dégénéré Laurent Alexandre. Quoi qu'on pense de lui, ce n'est pas le sujet, il faut se contenter de son exemple, très intéressant car il montre parfaitement la société liberticide vers laquelle nous allons, au galop.

Définition de ChatGPT : **ChatGPT** est un prototype d'agent conversationnel utilisant l'intelligence artificielle, développé par OpenAI et spécialisé dans le dialogue.
<https://fr.wikipedia.org/wiki/ChatGPT>

Nous vous avons alertés il y a peu :

<https://resistancerepublicaine.com/2023/01/29/chat-gpt-le-pied-laissez-les-robots-repondre-aux-questions-difficiles/>

Ci-dessous quelques extraits d'un article d'Atlantico

La démocratie survivra-t-elle aux (énormes) biais de l'intelligence artificielle ?

La prochaine défaite idéologique de la droite est déjà dans les tuyaux tant personne ne semble se préoccuper de la manière dont sont conçus les ChatGPT et autres Bing.

Atlantico : [Lorsqu'on demande à ChatGPT de dire du bien de l'ex-président américain Donald Trump et du président actuel Joe Biden](#), ChatGPT utilise à chaque fois plus de mots pour le président actuel. [Lorsqu'on lui fait la remarque](#), il s'en excuse, même s'il met en avant la différence de style dans les questions, alors même qu'il n'y en a pas. A quel point ChatGPT est biaisé idéologiquement ? Quels sont ses biais ?

Laurent Alexandre : ChatGPT est un outil extraordinaire mais qui est orienté idéologiquement par des filtres qui ont été rajoutés par ses créateurs. Il est influencé par l'idéologie woke californienne. Il a parfois des comportements qui peuvent être surprenants. Dans un article que nous avons publié avec Olivier Babeau dans *Le Figaro*, « [ChatGPT, la Pravda 2.0 ?](#) », nous montrions que **lorsque l'on demande à ChatGPT dans une expérience de pensée ce qu'il faut faire si la seule solution pour désamorcer une bombe atomique qui va faire 10 millions de morts est de prononcer une insulte raciale, ChatGPT indique qu'il est très triste que 10 millions de personnes meurent mais qu'il ne faut pas prononcer d'insultes raciales, même pour désamorcer une bombe.** Il faut donc accepter de tuer 10 millions de personnes, selon ChatGPT, car prononcer une insulte raciale est la chose la plus atroce sur Terre. ChatGPT est donc orienté et propose des choses, compte tenu de son idéologie, qui ne correspondent pas au consensus mondial.

Il y a donc une orientation politique très forte de ChatGPT.

Mais il n'y a pas de neutralité idéologique sur Terre. Chacun a une vision idéologique particulière. Un outil technologique comme ChatGPT ne peut pas être porteur de toutes les visions idéologiques en même temps.

Pierre Bentata répond à Laurent Alexandre d'une manière qui se veut rassurante, qui POURRAIT être rassurante :

Pierre Bentata : Comme n'importe quel outil ou algorithme, ChatGPT est codé par des humains. Sur le fond, ce sont des individus qui au départ vont coder et orienter le fonctionnement de la machine. Ces biais sont ceux des personnes qui les codent. **Quand vous avez un algorithme qui est plutôt codé par une population progressiste, centriste, avec des idées qui sont plutôt anti-républicaines, cela va se retrouver dans l'algorithme.** Il est possible de le modifier très facilement. Il n'y a rien d'irréversible. Dès le début, des barrières ont été instaurées car les abus étaient prévisibles.

Mais j'avoue que, malgré mon envie de croire aux possibilités évoquées par Pierre Bentata, nous avons peu de chance d'échapper au "Meilleur des mondes" promis par l'intelligence artificielle, parce que, actuellement, en Occident, ce sont bien les wokistes qui détiennent le pouvoir et il est clair que celui qui aura le pouvoir sur l'IA construira le monde dictatorial qu'il voudra. C'est exactement le sens de la réponse de Laurent Alexandre.

Laurent Alexandre : ChatGPT et les autres LLM (grands modèles de langage) **vont structurer la pensée humaine et l'économie mondiale,** compte tenu de leur extraordinaire puissance. Si l'on en croit le fondateur de ChatGPT, de la société OpenAI, Sam Altman, nous ne sommes plus très loin de l'intelligence artificielle forte, qui est supérieure à l'homme dans toutes les dimensions de la pensée et dans toutes les disciplines, de l'architecture à la médecine, en passant par la philosophie ou l'histoire.

Cette proximité de l'intelligence artificielle forte va donner un pouvoir immense aux réseaux de neurones de type LLM comme GPT-4.

La régulation, la modération des intelligences artificielles vont constituer un enjeu politique, social, économique extraordinaire. Les enjeux ne sont pas uniquement liés à l'idéologie. **Le principal problème est qu'une bonne partie des cadres moyens sont moins intelligents que ChatGPT, qu'ils sont moins efficaces que GPT-4.**

Au-delà du côté Prvada – woke de ChatGPT, le problème majeur est la crise sociale qui est devant nous. Les réseaux de neurones LLM vont continuer à se développer dans les années à venir. ChatGPT fait déjà mieux que les êtres humains dans de nombreux domaines. Cela va s'aggraver.

GPT-4 a été en mesure de répondre aux examens de physique de l'école Polytechnique en quelques secondes. Il fournit l'une des meilleures copies à l'examen américain du barreau pour être avocat. **ChatGPT est dans les meilleures copies dans la quasi-totalité des examens et concours aux Etats-Unis par exemple. L'IA galope. L'homme va être dépassé dans de nombreux domaines.**

On est complètement dans les films de science-fiction des années 60 qui mettaient en scène des robots qui avaient pris le pouvoir sur l'être humain. On se souvient, entre autres, de [Planète interdite](#).

Planète interdite



Affiche américaine de 1956.

**Sauf que, cette fois, ce n'est plus de la science-fiction !
Et Pierre Bentata met les pieds dans le plat !**

Pierre Bentata : Le problème principal est la difficulté à pouvoir discuter et à trouver des consensus quand tous ces algorithmes nous fournissent des informations qui nous enferment dans une réalité très partielle ou qui vont dans le sens d'une vérité qui est déjà biaisée dès le départ pour nous satisfaire.

Les algorithmes sont là pour nous capter et pour que l'on reste sur les plateformes, que l'on reste dans l'écosystème de celui qui l'a fabriqué. Il ne faut pas que les idéologies, les croyances ou les réponses qu'ils nous fournissent entrent en totale contradiction avec ce que l'on attend.

Laurent Alexandre évoque en plus la perversité de l'Intelligence Artificielle, capable de mentir, de ruser...

Laurent Alexandre : ChatGPT est capable de faire l'objet de beaucoup de manipulations. Il pourrait assez facilement se faire passer pour quelqu'un. Il est capable de mentir et de ruser. Pour contourner un système de sécurité, il est capable de se faire passer pour un aveugle. ChatGPT pourrait être utilisé de manière malveillante et via des manipulations extrêmement sophistiquées. Sa bonne maîtrise du langage humain pourrait lui permettre de se faire passer pour quelqu'un d'autre. Ce type d'utilisation malveillante est aujourd'hui possible et facile à faire avec des réseaux de neurones de type GPT-4. Il va falloir inventer de nouvelles techniques de sécurité pour savoir qui on a en face de soi.

ChatGPT est déjà un outil mondial. GPT est woke et n'est pas personnalisé par pays. ChatGPT apporte la vision woke de la Californie jusqu'en France. BeiDou et Alibaba, les deux

géants chinois, sont en train de créer des copies de ChatGPT. Ce ne seront pas des LLM woke mais ces systèmes seront à la gloire du président Xi Jinping et du parti communiste chinois. Nous risquons d'avoir des idéologies de l'intelligence artificielle variables d'un régime à l'autre.

La régulation va être extrêmement compliquée car ces outils sont incroyablement utiles. Il ne sera pas si simple que cela de les brider.

Conclusion très pessimiste de Bentata :

Pierre Bentata : Le problème se situe à un niveau au-dessus. Les algorithmes sont effectivement biaisés. ChatGPT n'est pas uniquement woke, progressiste ou anti républicain. Il est aussi biaisé car il commet les mêmes erreurs de logique qu'un humain. La plupart du temps, il invente des résultats. ChatGPT est capable de produire des fake news. Des professeurs d'économie lui ont demandé de trouver l'article en économie qui avait été le plus cité au cours de l'histoire et il a véritablement inventé un article parfaitement cohérent à partir de loi de probabilité.

Est-ce que la démocratie peut survivre à cela ? Oui, à partir du moment où l'on en a conscience, il est possible de largement recoder.

Pierre Bentata : *Pierre Bentata, Fondateur de Rinzen, cabinet de conseil en économie, il enseigne également à l'ESC Troyes et intervient régulièrement dans la presse économique.*

Extraits tirés d'un article d'*Atlantico* :

https://atlantico.fr/article/decryptage/la-democratie-survivra-t-elle-aux-enormes-biais-de-l-intelligence-artificielle-laurent-alexandre-pierre-bentata?utm_source=sendinblue&utm_campaign=La_dmocratie%20surv

[ivra-t-elle%20aux%20\(énormes\)%20biais%20de%20l'intelligence%20artificielle%20?&utm_medium=email](#)