



ChatGPT: A Dual-Use Technology / Une technologie à double usage

We need pragmatic responses to complex issues / Nous avons besoin de réponses pragmatiques à des questions complexes

BRYN WILLIAMS-JONES

MAR 7, 2023

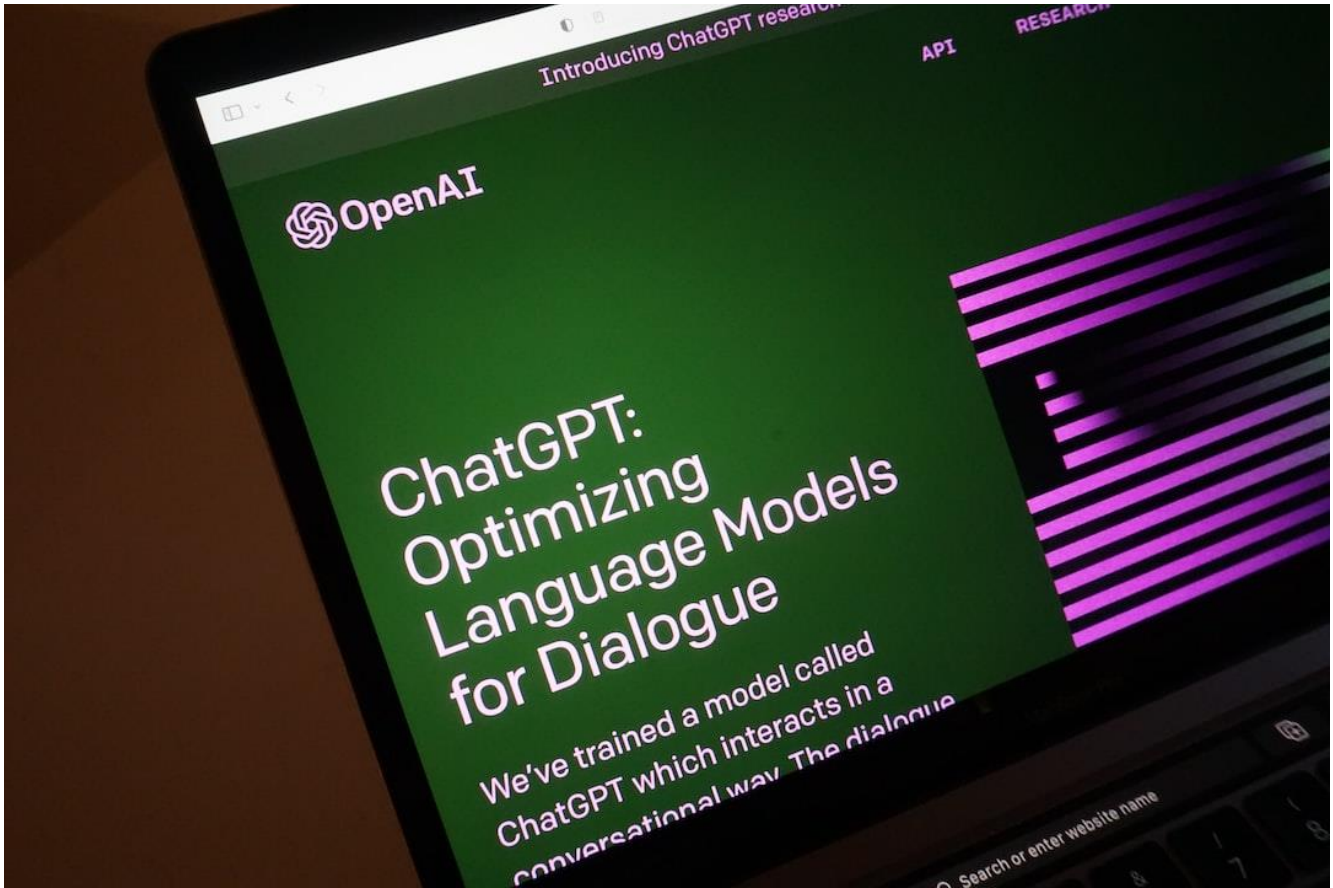


Photo by [Rolf van Root](#) on [Unsplash](#)

La version française de ce texte figure ci-dessous

By now, everyone has heard of [ChatGPT](#) (developed by [OpenAI](#)), the AI language model or “chabot”. We are still in the early days of this technology – [version 4](#) of ChatGPT is on the horizon – and other big tech players such as Alphabet/Google ([Bard](#)) and Microsoft ([Bing](#)) have released their own chatbots, some which met with [notable failures](#).

These technologies are raising important concerns about the trustworthiness of content, the massively simplified ability to cheat, and the difficulty of verifying whether particular content is human- or AI-generated.

The promises for widespread public access to an AI with whom we can have meaningful dialogue are evident. For a start, there is easy access to synthesized content in response to almost any question. Instead of doing a web search and having to wade through pages of articles, the chatbot provides fun access to and navigation through an ever-expanding sea of information. ChatGPT can quickly generate intelligent content on almost any topic, from creating a menu or meal plan for the week, to writing haikus, to answering general and even specific questions about life, politics, history, science, and even ethics!

When I asked the question “**What are the ethical issues raised by ChatGPT?**”, I got the following helpful and accurate (if general) answer:

| As an AI language model, ChatGPT raises several ethical issues, including:

1. **Bias:** ChatGPT is trained on large datasets of human language, which can include biases that are reflected in its responses. This bias can affect its recommendations and interactions with users, potentially leading to discrimination and unfairness.
2. **Privacy:** ChatGPT collects and processes large amounts of personal data from its users, which raises concerns about data protection and privacy. Users may be uncomfortable with their personal information being used for marketing or other purposes.
3. **Manipulation:** ChatGPT has the potential to influence users’ opinions and behavior through its recommendations and responses. This raises concerns about whether it should be used to manipulate people in any way, and what safeguards need to be put in place to prevent this from happening.
4. **Transparency:** ChatGPT’s responses are generated by complex algorithms that can be difficult for humans to understand. This can make it challenging to explain how it makes decisions, making it difficult to hold it accountable.
5. **Accountability:** ChatGPT is a powerful tool that can have significant impacts on people’s lives. There is a need for clear rules and regulations around its use to ensure that it is used in a responsible and ethical manner.
6. **Ownership:** As an AI language model, ChatGPT raises questions about who owns the content generated by the model. There are concerns about whether the intellectual property rights of the data used to train the model are being respected, and whether the model itself can be considered a form of intellectual property.

Overall, these ethical issues raise important questions about how we should design and regulate AI language models like ChatGPT to ensure that they are used responsibly.

The very simplicity and clarity of content generation that makes ChatGPT so appealing also raises serious concerns, particularly in education. Colleges and universities are rushing to change their [policies and regulations](#) in order to deal with a feared surge in plagiarism. For their part, scientific publishers and journal editor associations ([COPE](#), [WAME](#)) are concerned about author responsibility and being able to identify and prevent the dissemination of fabricated and false content that would undermine [conduct of research and provision of services](#).

But the potential uses for and concerns about ChatGPT are not limited to the academic community. Nor should these be framed exclusively as “problems”.

This technology presents a classic [dual-use situation](#). Knowledge and technology can be used for good or for ill, depending both on how they are developed (the values embedded within) and used (by whom and with what intent). And the associated risks and benefits will differ depending on the user and other stakeholders, the context of implementation, and even who or how one perceives or values these benefits or risks – one person’s risk can be another’s benefit.

Thankfully, these and other issues have also generated an impressive amount of public and academic attention (see these articles in [The Conversation](#)), and the research community has mobilized to reflect more deeply on these issues and propose solutions. For example, some of these were discussed at a [recent conference](#) organized by colleagues at the [International Observatory o the Social Impacts of AI and Digital Technology](#) (OBVIA), who are also working on position statements and policy briefs.

Given its sophistication of ChatGPT, what more could “flesh and blood” ethicists add to the analysis? The answer is nuance, rigorous argument, and above all credibility. The arguments and analyses proposed by researchers (who follow norms of responsible conduct of research) are supported by and grounded in the academic and policy literature. And as the authors are clear, so too is the responsibility for the quality and veracity of the content produced.

By contrast, ChatGPT and other AIs have a “[Black Box Problem](#)” – while we can make these technologies produce meaningful results, we are often unable (or unwilling) to explain “how” they work. The risk, then, as the old computer science adage goes, is “[Garbage in, garbage out](#)”. Biased (bad) data will lead to biased (bad) results and recommendations that are erroneous. And this then leads to very legitimate concerns about the accountability, responsibility, and trustworthiness of these systems and the organizations (companies) that produce them.

For many of us, ChatGPT may be our first real and meaningful encounter with a smart AI. Future versions will be even more intelligent and could enable, for example, personal librarians to whom we can ask our questions or avatars to manage our email on our behalf (yes please!). But what, then, happens to all those people currently filling the roles that ChatGPT might replace? What use is there for librarians, administrators, lawyers, policy makers, educators... ethicists? And what are the specific and responsibilities of the AI engineers and companies producing these technologies, and the governments who approve (and regulate) their implementation?

The AI genie is out of the bottle and has been for a while. The many questions raised by ChatGPT are the reason we have institutions like the [OBVIA](#) that mobilize hundreds of university researchers to think deeply about the challenges raised by AI, in all its incarnations, and to propose pragmatic solutions.

Collectively, we have to figure out how to benefit from the positive aspects of AI innovations while also being aware of their potential risks (personal, social, geopolitical, ecological, etc.). And the benefits and risks will not necessarily be equitably distributed, thus reinforcing existing inequalities. So, we need to continue building awareness amongst the public and policy makers, and where necessary, implement appropriate regulation.

For educators, I think we have to accept that students will be using these technologies. And we can either fight a losing battle of detection and prohibition with a technology that is likely always in advance of our ability to regulate... or we can [integrate these tools](#) into our teaching. Taking a pragmatic approach will mean [rethinking and being creative about our teaching](#), and intentionally using ChatGPT and other AI tools with our students, [helping them master their use and application](#) as part of a research or learning process. But this will obviously not be easy.

To maximize the benefits while minimizing the risks of the dual-use situations created by ChatGPT and other AI systems, we need (new) norms and cultures that are adapted to these realities, but that are also grounded in our fundamental values and principles.

À l'heure actuelle, tout le monde a entendu parler de [ChatGPT](#) (développé par [OpenAI](#)), le modèle linguistique d'IA ou "chatbot". Nous n'en sommes encore qu'aux premiers jours de cette technologie – la [version 4](#) de ChatGPT se profile à l'horizon – et d'autres grands acteurs de la technologie comme Alphabet/Google ([Bard](#)) et Microsoft ([Bing](#)) ont lancé leurs propres chatbots, dont certains ont connu des [échecs notables](#).

Ces technologies suscitent d'importantes inquiétudes quant à la fiabilité du contenu, à la possibilité simplifiée de tricher et à la difficulté de vérifier si un contenu particulier est généré par un humain ou une IA.

Les promesses d'un accès public généralisé à une IA avec laquelle nous pouvons avoir un dialogue significatif sont évidentes. Pour commencer, il y a un accès facile au contenu synthétisé en réponse à presque toutes les questions. Au lieu de faire une recherche sur le Web et de devoir patauger dans des pages d'articles, le chatbot permet d'accéder et de naviguer de manière ludique dans une mer d'informations en constante expansion. ChatGPT peut rapidement générer du contenu intelligent sur presque tous les sujets, qu'il s'agisse de créer un menu ou un plan de repas pour la semaine, d'écrire des haïkus ou de répondre à des questions générales ou spécifiques sur la vie, la politique, l'histoire, la science et même l'éthique!

Lorsque j'ai posé la question "**Quels sont les problèmes éthiques soulevés par ChatGPT?**" à ce chatbot, j'ai obtenu la réponse utile et précise (bien que générale) suivante :

En tant que modèle linguistique IA, ChatGPT soulève plusieurs questions éthiques, notamment :

1. **Biais** : ChatGPT est entraîné sur de grands ensembles de données de langage humain, ce qui peut inclure des biais qui se reflètent dans ses réponses. Ces préjugés peuvent affecter ses recommandations et ses interactions avec les utilisateurs, ce qui peut entraîner de la discrimination et de l'injustice.
2. **Vie privée** : ChatGPT collecte et traite de grandes quantités de données personnelles de ses utilisateurs, ce qui soulève des inquiétudes concernant la protection des données et la vie privée. Les utilisateurs peuvent être mal à l'aise à l'idée que leurs informations personnelles soient utilisées à des fins de marketing ou autres.
3. **Manipulation** : ChatGPT a le potentiel d'influencer les opinions et le comportement des utilisateurs grâce à ses recommandations et ses réponses. Cela soulève des inquiétudes quant à son utilisation pour manipuler les gens de quelque manière que ce soit, et aux garanties à mettre en place pour éviter que cela ne se produise.
4. **Transparence** : Les réponses de ChatGPT sont générées par des algorithmes complexes qui peuvent être difficiles à comprendre pour les humains. Il peut donc être difficile d'expliquer comment il prend ses décisions, ce qui rend difficile de le tenir pour responsable.
5. **Responsabilité** : ChatGPT est un outil puissant qui peut avoir des répercussions importantes sur la vie des gens. Il est nécessaire d'établir des règles et des règlements clairs autour de son utilisation pour s'assurer qu'il est utilisé de manière responsable et éthique.
6. **Propriété** : En tant que modèle linguistique IA, ChatGPT soulève des questions sur la propriété du contenu généré par le modèle. On peut se demander si les droits de propriété intellectuelle des données utilisées pour former le modèle sont respectés, et si le modèle lui-même peut être considéré comme une forme de propriété intellectuelle.

Dans l'ensemble, ces problèmes éthiques soulèvent d'importantes questions sur la façon dont nous devrions concevoir et réglementer les modèles linguistiques d'IA comme ChatGPT pour garantir qu'ils sont utilisés de manière responsable.

La simplicité et la clarté mêmes de la génération de contenu qui rendent ChatGPT si attrayant soulèvent également de sérieuses inquiétudes, en particulier dans l'éducation. Les collèges et les universités s'empressent de modifier [leurs politiques et leurs règlements](#) afin de faire face à la recrudescence redoutée du plagiat. De leur côté, les éditeurs scientifiques et les associations d'éditeurs de revues ([COPE](#), [WAME](#)) s'inquiètent de la responsabilité des auteurs et de la possibilité d'identifier

et d'empêcher la diffusion de contenus fabriqués et faux qui nuiraient à la [conduite de la recherche et à la prestation de services](#).

Mais les utilisations potentielles de ChatGPT et les préoccupations qu'il suscite ne se limitent pas à la communauté universitaire. Et je pense qu'il ne faut pas les considérer exclusivement comme des "problèmes".

Cette technologie présente une situation classique de [double usage](#). Les connaissances et les technologies peuvent être utilisées pour le bien ou pour le mal, selon la façon dont elles sont développées (les valeurs qui y sont intégrées) et utilisées (par qui et avec quelle intention). Et les risques et avantages associés diffèrent selon l'utilisateur et les autres parties prenantes, le contexte de mise en œuvre, et même selon qui ou comment on perçoit ou apprécie ces avantages ou risques – le risque d'une personne peut être l'avantage d'une autre.

Heureusement, ces questions et d'autres ont également suscité une quantité impressionnante d'attention de la part du public et des universitaires (voir ces articles dans [The Conversation](#)), et la communauté des chercheurs s'est mobilisée pour réfléchir plus profondément à ces questions et proposer des solutions. Par exemple, certains d'entre eux ont été abordés lors d'une [récente conférence](#) organisée par des collègues de l'[Observatoire international sur les impacts sociétaux de l'IA et du numérique](#) (OBVIA), qui travaillent également sur des déclarations de position et des mémoires politiques.

Étant donné la sophistication de ChatGPT, qu'est-ce que les éthiciens "en chair et en os" pourraient ajouter à l'analyse? La réponse est la nuance, l'argumentation rigoureuse et surtout la crédibilité. Les arguments et les analyses proposés par les chercheurs (qui suivent les normes de conduite responsable en recherche) sont soutenus par la littérature scientifique et politique et y sont ancrés. Et comme les auteurs sont clairs, la responsabilité de la qualité et de la véracité du contenu produit l'est aussi.

En revanche, ChatGPT et d'autres IA ont un "[problème de boîte noire](#)" – même si nous pouvons faire en sorte que ces technologies produisent des résultats significatifs, nous sommes souvent incapables (ou peu désireux) d'expliquer "comment" elles fonctionnent. Le risque, comme le dit le vieil adage informatique, est donc "[Garbage in, garbage out](#)". Des données (mauvais) biaisées conduiront à des résultats (mauvais) biaisés et à des recommandations erronées. Et cela entraîne ensuite des inquiétudes très légitimes quant à l'imputabilité, la responsabilité et la fiabilité de ces systèmes et des organisations (entreprises) qui les produisent.

Pour beaucoup d'entre nous, ChatGPT est peut-être notre première rencontre réelle et significative avec une IA intelligente. Les futures versions seront encore plus intelligentes et pourraient permettre, par exemple, de créer des bibliothécaires personnels à qui nous pourrions poser nos questions ou des avatars qui gèreraient nos courriels en notre nom (oui, s'il vous plaît!). Mais qu'arrivera-t-il alors à toutes ces personnes qui remplissent actuellement ces rôles que ChatGPT pourrait remplacer? Quelle est l'utilité des bibliothécaires, des administrateurs, des avocats, des décideurs, des éducateurs... des éthiciens? Et quelles sont les spécificités et les responsabilités des ingénieurs en IA et des entreprises qui produisent ces technologies, et des gouvernements qui approuvent (et réglementent) leur mise en œuvre?

Le génie de l'IA est sorti de la bouteille et ce depuis un moment. Les nombreuses questions soulevées par ChatGPT sont la raison pour laquelle nous avons des institutions comme l'[OBVIA](#) qui mobilisent des centaines de chercheurs universitaires pour réfléchir en profondeur aux défis posés par l'IA, dans toutes ses incarnations, et pour proposer des solutions pragmatiques.

Collectivement, nous devons trouver comment profiter des aspects positifs des innovations de l'IA tout en étant conscients de leurs risques potentiels (personnels, sociaux, géopolitiques, écologiques, etc.). Et les avantages et les risques ne seront pas forcément répartis équitablement, ce qui renforcera les inégalités existantes. Nous devons donc continuer à sensibiliser le public et les décideurs, et si nécessaire, mettre en place une réglementation appropriée.

Pour les éducateurs, je pense que nous devons accepter que les étudiants utiliseront ces technologies. Et nous pouvons soit mener une bataille perdue d'avance de détection et d'interdiction avec une technologie qui est probablement toujours en avance sur notre capacité à réguler... soit [intégrer ces outils](#) dans notre enseignement. Adopter une approche pragmatique signifiera [repenser notre enseignement et faire preuve de créativité](#), et utiliser intentionnellement ChatGPT et d'autres outils d'IA avec nos étudiants, les [aider à maîtriser leur utilisation et leur application](#) dans le cadre d'un processus de recherche ou d'apprentissage. Mais cela ne sera évidemment pas facile.

Pour maximiser les avantages tout en minimisant les risques des situations à double usage créées par ChatGPT et d'autres systèmes d'IA, nous avons besoin de (nouvelles) normes et cultures adaptées à ces réalités, mais qui sont également ancrées dans nos valeurs et principes fondamentaux.