

What if ChatGPT were good for ethics?

November 29 2023, by Virginie Soffer



Credit: Sanket Mishra from Pexels

Many people use ChatGPT: computer programmers write code with it, students do their homework with it and teachers plan their lessons with it. And yet the Open AI chatbot's rise has also prompted many ethical concerns.

To find out more, we talked to Marc-Antoine Dilhac, a philosophy professor at Université de Montréal's Faculty of Arts and Science who helped write the [Montréal Declaration for a Responsible Development of Artificial Intelligence](#).

Does ChatGPT reinforce discrimination?

Discrimination is an issue for artificial intelligence in general. It's not limited to ChatGPT. The gender-related bias seen in ChatGPT is similar to what we've already seen in traditional natural language processing and Google's autocomplete predictions. For example, in machine translations, AIs tend to use the masculine for certain professions and the feminine for others, so doctors can be automatically referred to as "he" and nurses as "she."

ChatGPT reproduces the gender bias in the pre-existing tests it's trained on, since they're mostly grounded in social norms. If we want that to change, we as human beings need to change how we usually think. When ChatGPT refers to doctors as "he," that should be recognized as a reflection of what we ourselves say. Hopefully that could push us to do something to reduce these biases. By exposing these biases, ChatGPT is actually doing us a favor: it reflects our own prejudices back at us, which lets us know we have them in the first place.

What are the main ethical challenges ChatGPT poses to society?

I see three big challenges.

The first is educational, and it's something Université de Montréal has already started thinking about. This concerns the future of learning in an environment where students can use ChatGPT to write texts and collect

information intended for them. Teachers could also use ChatGPT to correct assignments automatically, something [educational institutions](#) may encourage to reduce the time spent on corrections. But this raises the question of responsibility for evaluating student work. How does feedback fit in? What kind of student-teacher relationship do we want to build?

The second challenge involves the risk to intellectual property. ChatGPT and other generative AIs are trained on original works such as text and images (including photographs and paintings) to create synthetic content that the creators of those original works aren't compensated for. This issue has real legal and economic implications that may not only discourage people from producing artistic work, but could also discourage institutions like universities from producing knowledge.

Finally, the third challenge that could be exacerbated by ChatGPT is related to democracy and election integrity. This has to do with the potential to produce texts that target individuals to influence or manipulate their political beliefs. I'm not entirely convinced this is much of a risk, because I believe people are more easily convinced by the opinions of other human beings than by those generated by machine.

But it's true that we can't always identify the source of Internet content, and it's becoming increasingly easy to mass-produce articles. Internet users could be overwhelmed with information, which may end up affecting how they think. Massive amounts of text could be produced and used to microtarget individuals in a way that's much more precise than in the past.

You were involved in putting together the Montreal Declaration. How could its principles lead to the ethical use of ChatGPT?

There are at least three principles that could be followed to ensure ChatGPT is used ethically and responsibly.

The first fundamental principle to consider is respect for autonomy. This principle can be adapted to different levels of AI use. For example, when people like students, teachers, journalists or lawyers use ChatGPT to do their work, they put their own autonomy at risk. The issue here is that people are delegating their tasks, which could lead them to lose their autonomy. When we stop doing certain things ourselves and have other people or technology do them, we become dependent on the work completed by the other party, which in this case is AI.

The use of ChatGPT in education raises questions about students' ability to think critically on their own and on [teacher training](#), since teachers may be less likely to check or understand sources themselves if they can rely on the summaries provided by ChatGPT. Some uses of ChatGPT could endanger our cognitive abilities and therefore our autonomy.

The second principle is solidarity. As stated in the Montreal Declaration, this principle says we must constantly work toward maintaining quality human relationships and that AI should only be used to develop them. This means that we need to work with AI rather than delegating tasks to it. We also need to maintain meaningful interpersonal relationships that are sometimes necessary for certain roles, such as the caring professions.

You might think that providing [mental health services](#) through ChatGPT goes against the solidarity principle. But ChatGPT is already being used in this way. This poses a real problem since it obviously loses sight of what a therapeutic relationship is like. Uncontrolled commercial applications could be disastrous since they are built without an understanding of the ethical principles of responsible AI development. In this case, the principle in question involves what a therapeutic relationship, and more broadly a quality human relationship, is like.

The third principle is democratic participation. If we don't understand how AI works, if we don't have any control on content production, and if the content produced disrupts interactions between human beings and lowers the quality of discourse, then we undermine one of the foundations of democracy: the ability to make informed decisions based on reasonable debate with our fellow citizens.

How the principle of democratic participation is applied is crucial in this context. For humans to maintain control over AI, a certain level of transparency is required, and the public's use of the technology should be limited. Application programming interfaces (APIs) that allow programs like ChatGPT to be used by a third-party application (such as mental-health counseling apps) should be placed under strict control.

Provided by University of Montreal

Citation: What if ChatGPT were good for ethics? (2023, November 29) retrieved 27 April 2024 from <https://techxplore.com/news/2023-11-chatgpt-good-ethics.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.