Opinion: Let's base AI debates on reality, not extreme fears about the future

April 4 2023, by Fenwick McKelvey

Credit: AI-generated image ([disclaimer](#))

A recent open letter by computer scientists and tech industry leaders [calling for a six-month ban on artificial intelligence development](#) has [received widespread attention online](#). [Even Canada's Innovation Minister François-Philippe Champagne](#) has responded to the letter on Twitter.

The letter, published by the non-profit Future of Life Institute, has asked for all AI labs to stop training AI systems more powerful than GPT-4, the model behind ChatGPT. The letter argues that AI has been "locked in an out-of-control race to develop and deploy ever more powerful digital minds that no one—not even their creators—can understand, predict, or reliably control."

The letter assumes AI is becoming, or could become, "powerful digital minds"—a longtermist interpretation of AI's development that ignores important debates about AI today in lieu of future concerns.

Longtermism and AI

Longtermism is the belief that artificial intelligence poses long-term or existential risks to humanity's future by becoming an out-of-control superintelligence.

Worries about superintelligent AIs are usually the stuff of science fiction. AI fantasies are one of many fears in Silicon Valley that can lead to dark prophecies. But like the Torment Nexus meme, these worries translate into major investment not caution. Most major technology firms have cut their responsible AI teams.

ChatGPT is obviously not a path to superintelligence. The open letter sees AI language technology like ChatGPT as a cognitive breakthrough—something that allows an AI to compete with humans at general tasks. But that's only one opinion.

There are many others that see ChatGPT, its GPT-4 model and other language learning models as "stochastic parrots" that merely repeat what they learn online so they appear intelligent to humans.

Superintelligence's blind spots

Longtermism has direct policy implications that prioritize superintelligence over more pressing matter such as [AI's power imbalances](#). [Some proponents of longtermism](#) even consider regulation to stop superintelligence more urgent than addressing the climate emergency.

AI policy implications are immediate, not far off matters. Because GPT-4 is trained on the entire internet and has expressly commercial ends, [it raises questions about fair dealing and fair use](#).

[We still don't know if AI-generated texts and images are copyrightable in the first place](#), since machines and [animals](#) cannot hold copyright.

And when it comes to privacy matters, ChatGPT's approach is hard to distinguish from another AI application, [Clearview AI](#). Both AI models were trained using massive amounts of personal information collected on the [open internet](#). [Italy's data-protection authority has just banned ChatGPT over privacy concerns](#).

These immediate risks are left unmentioned in the open letter, which swings between wild philosophy and technical solutions, ignoring the issues that are right in front of us.

Drowning out pragmatism

The letter follows an old dynamic that my co-author and I identify in a [forthcoming peer-reviewed chapter about AI governance](#). There is a tendency to view AI as either an existential risk or something mundane and technical.

The tension between these two extremes is on display in the open letter. The letter begins by claiming "advanced AI could represent a profound change in the history of life on Earth" before calling for "robust public funding for technical AI safety research." The latter suggests the social harms of AI are merely technical projects to be solved.

The focus on these two extremes crowds out important voices trying to pragmatically discuss the immediate risks of AI mentioned above as well as labor issues and more.

The attention being given to the open letter is especially problematic in Canada because two other letters, written by artists and civil liberties organizations, have not received the same amount of attention. These letters call for reforms and a more robust approach to AI governance to protect those being affected by it.

An unneeded distraction toward AI legislation

Government responses to the open letter have stressed that Canada does have legislation—the Artificial Intelligence and Data Act (AIDA). The longterm risks of AI are being used to rush legislation now like AIDA.

AIDA is an important step toward a proper AI governance regime, but it needs to better consult with those affected by AI before being implemented. It cannot be rushed to respond to perceived longterm fears.

The letter's calls to rush AI legislation might end up advantaging the same few firms driving AI research today. Without time to consult, enhance public literacy and listen to those being affected by AI, AIDA risks passing on AI's accountability and auditing to institutions already well positioned to benefit from the technology, creating a market for a new AI auditing industry.

Humanity's fate might not be on the line, but AI's good governance certainly is.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation