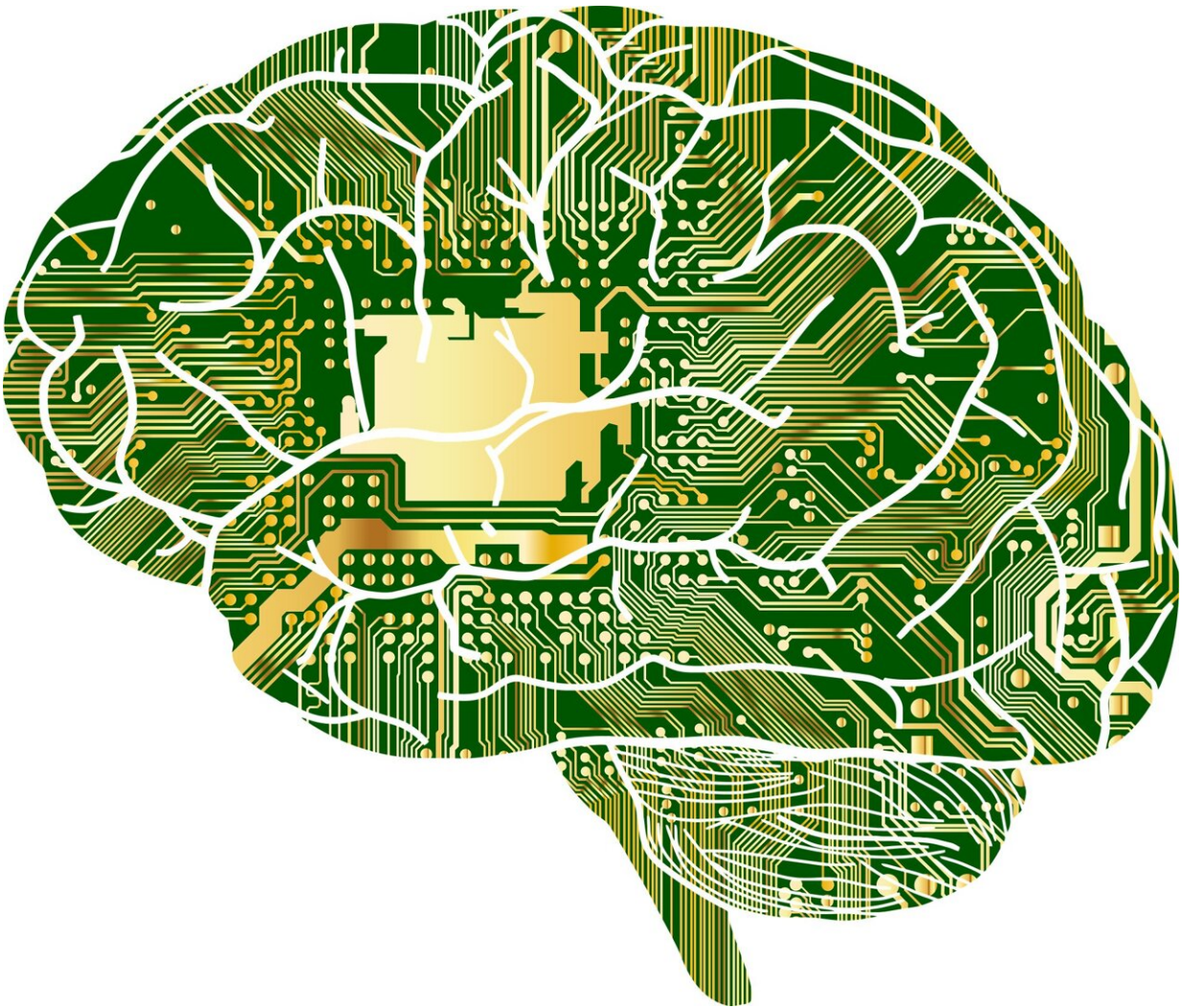


AI is an existential threat—just not the way you think

July 5 2023, by Nir Eisikovits



Credit: Pixabay/CC0 Public Domain

The rise of ChatGPT and similar artificial intelligence systems has been accompanied by a sharp [increase in anxiety about AI](#). For the past few months, executives and AI safety researchers have been offering predictions, dubbed "[P\(doom\)](#)," about the probability that AI will bring about a large-scale catastrophe.

Worries peaked in May 2023 when the nonprofit research and advocacy organization Center for AI Safety released [a one-sentence statement](#): "Mitigating the risk of extinction from A.I. should be a global priority alongside other societal-scale risks, such as pandemics and [nuclear war](#)." The statement was signed by many key players in the field, including the leaders of OpenAI, Google and Anthropic, as well as two of the so-called "godfathers" of AI: [Geoffrey Hinton](#) and [Yoshua Bengio](#).

You might ask how such existential fears are supposed to play out. One famous scenario is the "[paper clip maximizer](#)" thought experiment articulated by Oxford philosopher [Nick Bostrom](#). The idea is that an AI system tasked with producing as many paper clips as possible might go to extraordinary lengths to find raw materials, like destroying factories and causing car accidents.

A [less resource-intensive variation](#) has an AI tasked with procuring a reservation to a popular restaurant shutting down cellular networks and traffic lights in order to prevent other patrons from getting a table.

Office supplies or dinner, the basic idea is the same: AI is fast becoming an alien intelligence, good at accomplishing goals but dangerous because it won't necessarily align with the moral values of its creators. And, in its most extreme version, this argument morphs into explicit anxieties about AIs [enslaving or destroying the human race](#).

Actual harm

In the past few years, my colleagues and I at [UMass Boston's Applied Ethics Center](#) have been studying the impact of engagement with AI on people's understanding of themselves, and I believe these catastrophic anxieties are [overblown and misdirected](#).

Yes, AI's ability to create convincing deep-fake video and audio is frightening, and it can be abused by people with bad intent. In fact, that is already happening: Russian operatives likely attempted to embarrass Kremlin critic [Bill Browder](#) by ensnaring him in a conversation with an avatar for former Ukrainian President Petro Poroshenko. Cybercriminals have been using AI voice cloning for a variety of crimes—from [high-tech heists](#) to ordinary scams.

AI decision-making systems that [offer loan approval and hiring recommendations](#) carry the risk of algorithmic bias, since the training data and decision models they run on reflect long-standing social prejudices.

These are big problems, and they require the attention of policymakers. But they have been around for a while, and they are hardly cataclysmic.

Not in the same league

The statement from the Center for AI Safety lumped AI in with pandemics and [nuclear weapons](#) as a major risk to civilization. There are problems with that comparison. COVID-19 resulted in almost [7 million deaths worldwide](#), brought on a [massive and continuing mental health crisis](#) and created [economic challenges](#), including chronic supply chain shortages and runaway inflation.

Nuclear weapons probably killed [more than 200,000 people](#) in Hiroshima and Nagasaki in 1945, claimed many more lives from cancer in the years that followed, generated decades of profound anxiety during

the Cold War and brought the world to the brink of annihilation during the Cuban Missile crisis in 1962. They have also [changed the calculations of national leaders](#) on how to respond to international aggression, as currently playing out with Russia's invasion of Ukraine.

AI is simply nowhere near gaining the ability to do this kind of damage. The paper clip scenario and others like it are science fiction. Existing AI applications execute [specific tasks](#) rather than making broad judgments. The technology is [far from being able to decide on and then plan out](#) the goals and subordinate goals necessary for shutting down traffic in order to get you a seat in a restaurant, or blowing up a car factory in order to satisfy your itch for paper clips.

Not only does the technology lack the complicated capacity for multilayer judgment that's involved in these scenarios, it also does not have autonomous access to sufficient parts of our critical infrastructure to start causing that kind of damage.

What it means to be human

Actually, there is an existential danger inherent in using AI, but that risk is existential in the philosophical rather than apocalyptic sense. AI in its current form can alter the way people view themselves. It can degrade abilities and experiences that people consider essential to being human.

For example, humans are judgment-making creatures. People rationally weigh particulars and make daily judgment calls at work and during leisure time about whom to hire, who should get a loan, what to watch and so on. But more and more of these judgments are [being automated and farmed out to algorithms](#). As that happens, the world won't end. But people will gradually lose the capacity to make these judgments themselves. The fewer of them people make, the worse they are likely to become at making them.

Or consider the role of chance in people's lives. Humans value serendipitous encounters: coming across a place, person or activity by accident, being drawn into it and retrospectively appreciating the role accident played in these meaningful finds. But the role of algorithmic recommendation engines is to reduce that kind of serendipity and replace it with planning and prediction.

Finally, consider ChatGPT's writing capabilities. The technology is in the process of eliminating the role of writing assignments in higher education. If it does, educators will lose a key tool for teaching students [how to think critically](#).

Not dead but diminished

So, no, AI won't blow up the world. But the increasingly uncritical embrace of it, in a variety of narrow contexts, means the gradual erosion of some of humans' most important skills. Algorithms are already undermining people's capacity to make judgments, enjoy serendipitous encounters and hone critical thinking.

The human species will survive such losses. But our way of existing will be impoverished in the process. The fantastic anxieties around the coming AI cataclysm, singularity, Skynet, or however you might think of it, obscure these more subtle costs. Recall T.S. Eliot's famous closing lines of "[The Hollow Men](#)": "This is the way the world ends," he wrote, "not with a bang but a whimper."

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

Provided by The Conversation

Citation: AI is an existential threat—just not the way you think (2023, July 5) retrieved 11 May 2024 from <https://techxplore.com/news/2023-07-ai-existential-threatjust.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.