If we're going to label AI an 'extinction risk', we need to clarify how it could happen

May 31 2023, by Nello Cristianini

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

This week a group of well-known and reputable AI researchers [signed a statement](#) consisting of 22 words:

"Mitigating the risk of [extinction](#) from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war."

As a professor of AI, I am also in favor of reducing any risk, and prepared to work on it personally. But any statement worded in such a way is bound to create alarm, so its authors should probably be more specific and clarify their concerns.

As defined by Encyclopedia Britannica, extinction is "[the dying out or extermination of a species](#)". I have met many of the statement's signatories, who are among the most reputable and solid scientists in the field—and they certainly mean well. However, they have given us no tangible scenario for how such an extreme event might occur.

It is not the first time we have been in this position. On March 22 this year, a [petition](#) signed by a different set of entrepreneurs and researchers [requested a pause in AI deployment of six months](#). In the petition, on the website of the Future of Life Institute, they set out as their reasoning: "Profound risks to society and humanity, as shown by extensive research and acknowledged by top AI labs"—and accompanied their request with a list of rhetorical questions:

"Should we let machines flood our information channels with propaganda and untruth? Should we automate away all the jobs, including the fulfilling ones? Should we develop non-human minds that might eventually outnumber, outsmart, obsolete and replace us? Should we risk loss of control of our civilization?"

A generic sense of alarm

It is certainly true that, along with many benefits, this technology comes with risks that we need to take seriously. But none of the aforementioned scenarios seem to outline a specific pathway to extinction. This means we are left with a generic sense of alarm, without any possible actions we can take.

The website of the [Centre for AI Safety](#), where the latest statement appeared, outlines in a separate section [eight broad risk categories](#). These include the "weaponisation" of AI, its use to manipulate the news system, the possibility of humans eventually becoming unable to self-govern, the facilitation of oppressive regimes, and so on.

Except for weaponisation, it is unclear how the other—still awful—risks could lead to the extinction of our species, and the burden of spelling it out is on those who claim it.

Weaponisation is a real concern, of course, but what is meant by this should also be clarified. On its website, the Centre for AI Safety's main worry appears to be the use of AI systems to design chemical weapons. This should be prevented at all costs—but chemical weapons are already banned. Extinction is a very specific event which calls for very specific explanations.

On May 16, [at his US Senate hearing](#), Sam Altman, the CEO of OpenAI—which developed the ChatGPT AI chatbot—was twice asked to spell out his [worst-case scenario](#). He finally replied:

"My worst fears are that we—the field, the technology, the industry—cause significant harm to the world … It's why we started the company [to avert that future] … I think if this technology goes wrong, it can go quite wrong."

But while I am strongly in favor of being as careful as we possibly can be, and have been saying so publicly for the past ten years, it is important to maintain a sense of proportion—particularly when discussing the extinction of a species of eight billion individuals.

AI can create social problems that must really be averted. As scientists, we have a duty to understand them and then do our best to solve them.

But the first step is to name and describe them—and to be specific.

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

Provided by The Conversation