

Let's focus on AI's tangible risks rather than speculating about its potential to pose an existential threat

June 21 2023, by Nuria Oliver, Bernhard Schölkopf, Florence d'Alché-Buc, Nicolò Cesa-Bianchi, Sepp Hochreiter and Serge Belongie



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

Over the past few months, artificial intelligence (AI) has entered the global conversation as a result of the widespread adoption of generative AI-based tools such as chatbots and automatic image generation

programs. Prominent AI scientists and technologists have raised concerns about the [hypothetical existential risks](#) posed by these developments.

Having worked in AI for decades, this surge in popularity and the sensationalism that has followed have caught us by surprise. Our goal with this article is not to antagonize, but to balance the [public perception](#) which seems disproportionately dominated by fears of speculative AI-related existential threats.

It's not our place to say one cannot, or should not, worry about the more exotic risks. As members of the [European Laboratory for Learning and Intelligent Systems](#) (ELLIS), a research-anchored organization focused on machine learning, we do feel it is our place to put these risks into perspective, particularly in the context of governmental organizations contemplating regulatory actions with input from tech companies.

What is AI?

AI is a discipline within computer science or engineering that took shape [in the 1950s](#). Its aspiration is to build intelligent computational systems, taking as a reference human intelligence. In the same way as human intelligence is complex and diverse, there are many areas within [artificial intelligence](#) that aim to emulate aspects of human intelligence, from perception to reasoning, planning and decision-making.

Depending on the level of competence, AI systems can be divided into three levels:

1. Narrow or weak AI, which refers to AI systems that are able to perform specific tasks or solve particular problems, nowadays often with a level of performance superior to humans. All AI systems today are narrow AI. Examples include chatbots like

[chatGPT](#), voice assistants like Siri and Alexa, image recognition systems, and [recommendation algorithms](#).

2. General or strong AI, which refers to AI systems that exhibit a level of intelligence similar to that of humans, including the ability to understand, learn and apply knowledge across a wide range of tasks and incorporating concepts such as [consciousness](#). General AI is largely hypothetical and has not been achieved to date.
3. [Super AI](#), which refers to AI systems with an intelligence superior to [human intelligence](#) on all tasks. By definition, we are unable to understand this kind of intelligence in the same way an ant is not able to understand our intelligence. Super AI is an even more speculative concept than general AI.

AI can be applied to any field from education to transportation, healthcare, law or manufacturing. Thus, it is profoundly changing all aspects of society. Even in its "narrow AI" form, it has a significant potential to generate [sustainable economic growth](#) and help us [tackle the most pressing challenges](#) of the 21st century, such as climate change, pandemics, and inequality.

Challenges posed by today's AI systems

The adoption of AI-based decision-making systems over the last decade on a wide range of domains, from [social media](#) to the labor market, also poses significant societal risks and challenges that need to be understood and addressed.

The recent emergence of highly capable large, generative pre-trained

transformer (GPT) models exacerbates many of the existing challenges while creating new ones that deserve careful attention. The unprecedented scale and speed with which these tools have been adopted by hundreds of millions of people worldwide is placing further stress on our societal and regulatory systems.

There are some critically important challenges that should be our priority:

- The manipulation of human behavior by AI algorithms with potentially devastating social consequences in the spread of false information, the formation of public opinions and the outcomes of democratic processes.
- Algorithmic [biases and discrimination](#) that not only perpetuate but exacerbate stereotypes, patterns of discrimination, or even oppression.
- The lack of transparency in both models and their uses.
- The violation of privacy and the use of massive amounts of training data without consent by or compensation for its creators.
- The [exploitation of workers](#) annotating, training, and correcting AI systems, many of whom are in developing countries with meager wages.
- The [massive carbon footprint](#) of the large data centers and neural networks that are needed to build these AI systems.
- The lack of truthfulness in generative AI systems that invent believable content (images, texts, audios, videos...) without correspondence to the real world.
- The fragility of these large models that can make mistakes and be deceived.
- The displacement of jobs and professions.
- The concentration of power in the hands of an oligopoly of those controlling today's AI systems.

Is AI really an existential risk for humanity?

Unfortunately, rather than focusing on these tangible risks, the public conversation—most notably the recent open letters—has mainly focused on hypothetical existential risks of AI.

An existential risk refers to a potential event or scenario that represents a threat to the continued existence of humanity with consequences that could irreversibly damage or destroy human civilization, and therefore lead to the extinction of our species. A global catastrophic event (such as an asteroid impact or a pandemic), the destruction of a livable planet (due to [climate change](#), deforestation or depletion of critical resources like water and clean air), or a worldwide nuclear war are examples of existential risks.

Our world certainly faces a number of risks, and future developments are hard to predict. In the face of this uncertainty, we need to prioritize our efforts. The remote possibility of an uncontrolled super-[intelligence](#) thus needs to be viewed in context, and this includes the context of 3.6 billion people in the world who are [highly vulnerable due to climate change](#); the roughly 1 billion people who live on less than 1 US dollar a day; or the 2 billion people who are [affected by conflict](#). These are real human beings whose lives are in severe danger today, a danger certainly not caused by super AI.

Focusing on a hypothetical existential risk deviates our attention from the documented severe challenges that AI poses today, does not encompass the different perspectives of the broader research community, and contributes to unnecessary panic in the population.

Society would surely benefit from including the necessary diversity, complexity, and nuance of these issues, and from designing concrete and coordinated actionable solutions to address today's AI challenges,

including [regulation](#). Addressing these challenges requires the collaboration and involvement of the most impacted sectors of society together with the necessary technical and governance expertise. It is time to act now with ambition and wisdom—and in cooperation.

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

Provided by The Conversation

Citation: Let's focus on AI's tangible risks rather than speculating about its potential to pose an existential threat (2023, June 21) retrieved 24 February 2024 from <https://techxplore.com/news/2023-06-focus-ai-tangible-speculating-potential.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.