A.I. Poses 'Risk of Extinction,' Industry Leaders Warn

Leaders from OpenAI, Google Deepmind, Anthropic and other A.I. labs warn that future systems could be as deadly as pandemics and nuclear weapons.

By Kevin Roose

May 30, 2023 Updated 10:16 a.m. ET

A group of industry leaders warned on Tuesday that the artificial intelligence technology they are building may one day pose an existential threat to humanity and should be considered a societal risk on par with pandemics and nuclear wars.

"Mitigating the risk of extinction from A.I. should be a global priority alongside other societal-scale risks, such as pandemics and nuclear war," reads a one-sentence statement released by the Center for AI Safety, a nonprofit organization. The open letter has been signed by more than 350 executives, researchers and engineers working in A.I.

The signatories included top executives from three of the leading A.I. companies: Sam Altman, chief executive of OpenAI; Demis Hassabis, chief executive of Google DeepMind; and Dario Amodei, chief executive of Anthropic.

Geoffrey Hinton and Yoshua Bengio, two of the three researchers who won a Turing Award for their pioneering work on neural networks and are often considered "godfathers" of the modern A.I. movement, signed the statement, as did other prominent researchers in the field. (The third Turing Award winner, Yann LeCun, who leads Meta's A.I. research efforts, had not signed as of Tuesday.)

The statement comes at a time of growing concern about the potential harms of artificial intelligence. Recent advancements in so-called large language models — the type of A.I. system used by ChatGPT and other chatbots — have raised fears that A.I. could soon be used at scale to spread misinformation and propaganda, or that it could eliminate millions of white-collar jobs.

Eventually, some believe, A.I. could become powerful enough that it could create societal-scale disruptions within a few years if nothing is done to slow it down, though researchers sometimes stop short of explaining how that would happen.

These fears are shared by numerous industry leaders, putting them in the unusual position of arguing that a technology they are building — and, in many cases, are furiously racing to build faster than their competitors — poses grave risks and should be regulated more tightly.

This month, Mr. Altman, Mr. Hassabis and Mr. Amodei met with President Biden and Vice President Kamala Harris to talk about A.I. regulation. In a Senate testimony after the meeting, Mr. Altman warned that the risks of advanced A.I. systems were serious enough to warrant government intervention and called for regulation of A.I. for its potential harms.

Dan Hendrycks, the executive director of the Center for AI Safety, said in an interview that the open letter represented a "coming-out" for some industry leaders who had expressed concerns — but only in private — about the risks of the technology they were developing.

"There's a very common misconception, even in the A.I. community, that there only are a handful of doomers," Mr. Hendrycks said. "But, in fact, many people privately would express concerns about these things."

Some skeptics argue that A.I. technology is still too immature to pose an existential threat. When it comes to today's A.I. systems, they worry more about short-term problems, such as biased and incorrect responses, than longer-term dangers.

But others have argued that A.I. is improving so rapidly that it has already surpassed human-level performance in some areas, and it will soon surpass it in others. They say the technology has showed signs of advanced capabilities and understanding, giving rise to fears that "artificial general intelligence," or A.G.I., a type of artificial intelligence that can match or exceed human-level performance at a wide variety of tasks, may not be far-off.

In a blog post last week, Mr. Altman and two other OpenAI executives proposed several ways that powerful A.I. systems could be responsibly managed. They called for cooperation among the leading A.I. makers, more technical research into large language models and the formation of an international A.I. safety organization, similar to the International Atomic Energy Agency, which seeks to control the use of nuclear weapons.

Mr. Altman has also expressed support for rules that would require makers of large, cutting-edge A.I. models to register for a government-issued license.

In March, more than 1,000 technologists and researchers signed another open letter calling for a six-month pause on the development of the largest A.I. models, citing concerns about "an out-of-control race to develop and deploy ever more powerful digital minds."

That letter, which was organized by another A.I.-focused nonprofit, the Future of Life Institute, was signed by Elon Musk and other well-known tech leaders, but it did not have many signatures from the leading A.I. labs.

The brevity of the new statement from the Center for AI Safety — just 22 words in all — was meant to unite A.I. experts who might disagree about the nature of specific risks or steps to prevent those risks from occurring, but who shared general concerns about powerful A.I. systems, Mr. Hendrycks said.

“We didn’t want to push for a very large menu of 30 potential interventions,” Mr. Hendrycks said. “When that happens, it dilutes the message.”

The statement was initially shared with a few high-profile A.I. experts, including Mr. Hinton, who quit his job at Google this month so that he could speak more freely, he said, about the potential harms of artificial intelligence. From there, it made its way to several of the major A.I. labs, where some employees then signed on.

The urgency of A.I. leaders’ warnings has increased as millions of people have turned to A.I. chatbots for entertainment, companionship and increased productivity, and as the underlying technology improves at a rapid clip.

“I think if this technology goes wrong, it can go quite wrong,” Mr. Altman told the Senate subcommittee. “We want to work with the government to prevent that from happening.”

Kevin Roose is a technology columnist and the author of “Futureproof: 9 Rules for Humans in the Age of Automation.”

More from Kevin Roose

The Surgeon General’s Social Media Warning and A.I.’s Existential Risks
May 26, 2023

Mr. Altman Goes to Washington, and Casey Goes on This American Life
May 19, 2023