

Sunak's global AI safety summit risks achieving very little, warns tech boss

Publication Date: 2023-10-20

Author: Kiran Stacey

Section: Technology

Tags: Artificial intelligence (AI), Technology sector, Rishi Sunak, Regulators, news

Article URL: <https://www.theguardian.com/technology/2023/oct/20/rishi-sunak-global-ai-safety-summit-connor-leahy>



One of the executives invited to Rishi Sunak's international AI safety summit next month has warned that the conference risks achieving very little, accusing powerful tech companies of attempting to "capture" the landmark meeting. Connor Leahy, the chief executive of the AI safety research company Conjecture, said he believed heads of government were poised to agree a style of regulation that would allow companies to continue developing "god-like" AI almost unchecked. Leahy is one of just 100 people, including foreign government ministers, tech executives and civil society figures, who have been invited to November's summit at Bletchley Park, which Downing Street is hoping will mark a turning point in how advanced AI technology is developed. Officials have published an agenda for the summit that refers to the importance of "responsible capability scaling" – the idea that companies should develop their cutting-edge models according to a set of guidelines. However Leahy and others believe that there should be a complete moratorium on developing artificial general intelligence – AI models that can accomplish tasks at a human or beyond-human level of intelligence. Leahy said: "The primary aim of responsible scaling is to provide a framework which looks like something was done so that politicians can go home and say: 'We have done something.' But the actual policy is nothing." Leahy and others have launched a campaign called Control AI to urge policymakers to go further and use the Bletchley Park summit to implement a pause on developing the most sophisticated forms of AI. Like some others in the industry, Leahy believes humanity is at risk if humans develop an AI system that learns how to evade human control. "If you build systems that are more capable than humans at manipulation, business, politics, science and everything else, and we do not control them, then the future belongs to them, not us," he said. "The AI companies in particular and other organisations around them are trying to capture the summit, lock in a status quo of an unregulated race to disaster," he added. Leahy cited a recent interview with the CEO of the AI firm Anthropic, Dario Amodei, who said the chances of an AI system going "catastrophically wrong on the scale of ... human civilisation" was between 10% and 25%. Leahy said: "If you are building a machine that has such a chance then my suggestion is: don't do it." A government spokesperson said the summit was the start of a "global conversation" on AI and would welcome "diverse viewpoints". However, Leahy's comments echo recent warnings from others in the world of AI. Geoffrey Hinton, the man known as the "godfather of AI", recently quit Google to sound a warning about what he called the "existential risk" posed by digital intelligence. Weeks later, a group of senior executives from the AI industry including Leahy released a one-sentence statement saying: "Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks, such as pandemics and nuclear war." Warnings such as this prompted Sunak to call the summit for 1-2 November. It is designed to act as a forum for heads of state and technology executives to discuss these issues in person for the first time. British

officials have been touring the globe encouraging heads of state to attend the conference, which they hope will be the first in a series of such summits. Eric Schmidt, the former Google chief executive, and Mustafa Suleyman, the co-founder of Inflection and DeepMind, called this week for international leaders to set up a global panel of experts on AI akin to the Intergovernmental Panel on Climate Change. UK officials say they support such a body but that agreeing the finer points of how it should work would be better left for discussions at the UN. Instead, UK government sources say they are hoping to see any sign of the AI industry slowing down its development of artificial general intelligence, saying such a move would be unprecedented in the technology's history. Leahy, however, argues that vague promises to proceed with responsible development of advanced AI will not be enough, adding that doing so would amount to a victory for technology companies over regulators. A government spokesperson said: "The AI safety summit will bring together a wide array of attendees including international governments, academia, industry and civil society, as part of a collaborative approach to drive targeted, rapid international action on the safe and responsible development of AI." "As is entirely normal for summits of this nature, we do not confirm attendees this far in advance. This is the start of the global conversation on frontier AI risks and we welcome diverse viewpoints – this is a strength of the summit."