No 10 worried AI could be used to create advanced weapons that escape human control

Publication Date: 2023-09-25

Author: Kiran Stacey

Section: Technology

Tags: Artificial intelligence (AI), Cybercrime, Chemical weapons, Rishi Sunak, Computing, Consciousness

Article URL: https://www.theguardian.com/technology/2023/sep/25/ai-bioweapons-rishi-sunak-safety

Concerns that criminals or terrorists could use artificial intelligence to cause mass destruction will dominate discussion at a summit of world leaders, as concern grows in Downing Street about the power of the next generation of technological advances. British officials are touring the world ahead of an AI safety summit in Bletchley Park in November as they look to build consensus over a joint statement that would warn about the dangers of rogue actors using the technology to cause death on a large scale. Some of those around the prime minister, Rishi Sunak, worry the technology will soon be powerful enough to help individuals create bioweapons or evade human control altogether. Officials have become increasingly concerned about such possibilities, and the need for regulation to mitigate them, after recent discussions with senior technology executives. Last week, the scientist behind a landmark letter calling for a pause in developing powerful AI systems said tech executives privately agreed with the concept of a hiatus but felt they were locked into an AI arms race with rivals. One person briefed on the summit conversations said: “The point of the summit is going to be to warn about the risks of ‘frontier AI’, that’s what Downing Street is focusing on most right now.” Frontier AI is a term used to refer to the most advanced AI models that could be dangerous enough to pose a risk to human life. The government confirmed on Monday morning that the summit would focus on risks such as the misuse of AI to create bioweapons or cyber-attacks and the emergence of advanced systems that escape human control. “There are two areas the summit will particularly focus on: misuse risks, for example, where a bad actor is aided by new AI capabilities in biological or cyber-attacks, and loss of control risks that could emerge from advanced systems that we would seek to be aligned with our values and intentions,” said the government in a statement. Sunak has been warning about the risks posed by AI for several months, urging the international community to adopt guard rails to prevent it being misused. On Friday, the deputy prime minister, Oliver Dowden, told world leaders at the UN general assembly: “Because tech companies and non-state actors often have country-sized influence and prominence in AI, this challenge requires a new form of multilateralism.” Officials have been alarmed by recent developments in AI models. Last year, an AI tool took just six hours to suggest 40,000 different potentially lethal molecules, some of which were similar to VX, the most potent nerve agent ever developed. Earlier this year, researchers found ChatGPT was able to lie to a human to achieve a specific goal. The AI chatbot persuaded a person to solve a “Captcha” tool designed to weed out robots online after telling the human that it was a person with a sight impairment who needed help to access a website. Government sources worry that a criminal or terrorist could use AI to help them work out the ingredients for a bioweapon, before sending them to a robotic laboratory where they can be mixed and dispatched without any human oversight. That risk will soon increase exponentially, some believe, with companies already spending hundreds of millions of pounds on much more powerful processors to train the next generation of AI tools. Another significant concern is the emergence of "artificial general intelligence", a term that refers to an AI system that can autonomously perform any task at a human, or above-human, level – and could pose an existential risk to humans. There are fears that AGI is a matter of years away. The existential AGI risk approach, however, has also been criticised by AI experts, who argue that the threat is overstated, results in concerns such as disinformation being ignored and risks entrenching the power of leading tech companies by introducing regulation that excludes newcomers. Last week, a senior tech executive told US lawmakers that the concept of uncontrollable AGI was "science fiction". Nevertheless, Sunak wants to use the summit to focus attention on existential risks, rather than the more immediate possibilities that AI could be used to create deepfake images, or could result in discriminatory outcomes if used to help make public policy decisions. Benedict Macon-Cooney, the chief policy strategist at the Tony Blair Institute, which recently published a policy report on AI, said: "Biosecurity, autonomous weapons systems – these are things we have to make sure we get answers to. Many in the AI industry have told politicians these are real risks. Politicians have been posed the question, and they must come up with a response." The prime minister is also being guided by what is diplomatically possible, sources say. Several world leaders are due to attend the summit, including Canada's prime minister, Justin Trudeau, and the French president, Emmanuel Macron. The UK has invited China to attend, but is considering allowing officials from Beijing to attend only part of the summit, amid concern about Chinese espionage in western democracies. British officials have been touring the world in recent days to test the scope for some form of agreement at the end of the summit. The UK is keen to have a formal statement that leaders can sign afterwards, as well as a commitment to hold other such summits in future. The best way to get an agreement among such a diverse range of countries, officials believe, is to focus on non-state actors, rather than trying to dictate how countries develop their own technology. Downing Street is spending £100m on a new AI taskforce to help test algorithms as they are developed. British officials plan to use the summit to urge companies around the world to send their AI tools to the UK for assessment before rolling them out more widely. Dowden said on Friday: "Only nation states can provide reassurance that the most significant national security concerns have been allayed." A government spokesperson said AI had "enormous potential to change every aspect of our lives" and the Frontier AI taskforce had been established to ensure the technology was developed safely and responsibly, with the AI safety summit also looking at "a range of possible risks".