AI risk must be treated as seriously as climate crisis, says Google DeepMind chief

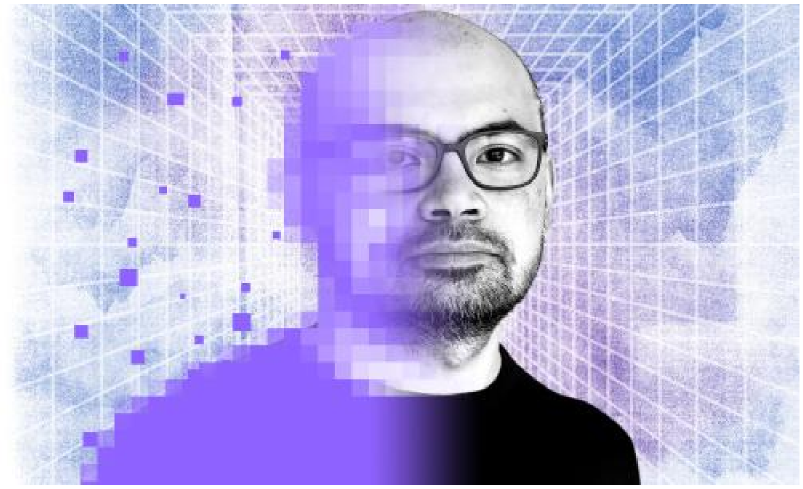
Publication Date: 2023-10-24

Author: Dan Milmo

Section: Technology

Tags: Technology, The AI race, Artificial intelligence (AI), Computing, Consciousness, news

Article URL: https://www.theguardian.com/technology/2023/oct/24/ai-risk-climate-crisis-google-deepmind-chief-demis-hassabis-regulation



The world must treat the risks from artificial intelligence as seriously as the climate crisis and cannot afford to delay its response, one of the technology's leading figures has warned. Speaking as the UK government prepares to host a summit on AI safety, Demis Hassabis said oversight of the industry could start with a body similar to the Intergovernmental Panel on Climate Change (IPCC). Hassabis, the British chief executive of Google's AI unit, said the world must act immediately in tackling the technology's dangers, which included aiding the creation of bioweapons and the existential threat posed by super-intelligent systems. "We must take the risks of AI as seriously as other major global challenges, like climate change," he said. "It took the international community too long to coordinate an effective global response to this, and we're living with the consequences of that now. We can't afford the same delay with AI." Hassabis, whose unit created the revolutionary AlphaFold program that depicts protein structures, said AI could be "one of the most important and beneficial technologies ever invented". However, he told the Guardian a regime of oversight was needed and governments should take inspiration from international structures such as the IPCC. "I think we have to start with something like the IPCC, where it's a scientific and research agreement with reports, and then build up from there." He added: "Then what I'd like to see eventually is an equivalent of a Cern for AI safety that does research into that – but internationally. And then maybe there's some kind of equivalent one day of the IAEA, which actually audits these things." The International Atomic Energy Agency (IAEA) is a UN body that promotes the secure and peaceful use of nuclear technology in an effort to prevent proliferation of nuclear weapons, including via inspections. However, Hassabis said none of the regulatory analogies used for AI were "directly applicable" to the technology, though "valuable lessons" could be drawn from existing institutions. Last week Eric Schmidt, the former Google chief executive, and Mustafa Suleyman, the co-founder of DeepMind, called for the creation of an IPCC-style panel on AI. Although UK officials support such a move, it is thought they believe its creation should be carried out under the auspices of the UN. Hassabis said AI could bring "incredible opportunities" in fields such as medicine and science but acknowledged existential concerns around the technology. Those centre on the possible development of artificial general intelligence (AGI) – systems with human or above-human levels of intelligence that could evade human control. Hassabis was one of the signatories in May of a statement warning that the threat of extinction from AI should be considered a societal-scale risk on a par with pandemics and nuclear war. "We should be starting that thinking and that research now. I mean yesterday, really," he said. "That's why I signed, and many people signed, that letter. It's because we wanted to give credibility to that being a reasonable thing to discuss." Some tech industry insiders are concerned that AGI or "god-like" AI could be only a few years away from emerging – though there is also a view that fears about the existential threat are being overplayed. Hassabis said the world was a long time away from AGI systems being developed but “we can see the path there, so we should be discussing it now”. He said current AI systems “aren’t of risk but the next few generations may be when they have extra capabilities like planning and memory and other things … They will be phenomenal for good use cases but also they will have risks.” The summit on 1 and 2 November at Bletchley Park, the base for second world war codebreakers including Alan Turing, will focus on the threat of advanced AI systems helping to create bioweapons, carry out crippling cyber-attacks or evading human control. Hassabis will be one of the attenders, along with the chief executives of leading AI firms including OpenAI, the San Francisco-based developer of ChatGPT. Hassabis’s unit has achieved significant breakthroughs in AI technology such as creating the AlphaGo AI program that defeated the world’s best player at Go, a Chinese board game, and the groundbreaking AlphaFold project that predicts how proteins fold into 3D shapes, a process that has paved the way for breakthroughs in a range of areas including tackling disease. Hassabis said he was optimistic about AI because of its potential to revolutionise fields such as medicine and science but a “middle way” needed to be found for managing the technology. AI has leapt up the political agenda after the public release last year of ChatGPT, a chatbot that became a sensation owing to its ability to produce highly plausible text responses to typed human prompts, from producing lengthy academic essays to recipes and job applications, and even helping people revoke parking tickets. AI image-generating tools such as Midjourney have also astonished observers by creating realistic images including a notorious “photograph” of the pope in a puffer jacket, giving rise to concerns that rogue actors could use AI tools to mass produce disinformation. Those fears have fed concerns about the potential power of the next generation of AI models. Hassabis, whose unit is working on a new AI model called Gemini that will generate image and text, said he envisaged a Kitemark-style system for models emerging. This year the UK government launched the Frontier AI taskforce, which aims to create guidelines for testing cutting-edge AI models and could become a benchmark for international-level testing efforts. Hassabis said: “I can imagine in future you’d have this battery of 1,000 tests, or 10,000 tests could be, and then you get safety Kitemark from that.”