

# Experts disagree over threat posed but artificial intelligence cannot be ignored

Publication Date: 2023-09-25

Author: Dan Milmo

Section: Technology

Tags: Artificial intelligence (AI), Computing, analysis

Article URL: <https://www.theguardian.com/technology/2023/sep/25/experts-disagree-over-threat-posed-but-artificial-intelligence-cannot-be-ignored-ai>



For some AI experts, a watershed moment in artificial intelligence development is not far away. And the global AI safety summit, to be held at Bletchley Park in Buckinghamshire in November, therefore cannot come soon enough. Ian Hogarth, the chair of the UK taskforce charged with scrutinising the safety of cutting-edge AI, raised concerns before he took the job this year about artificial general intelligence, or “God-like” AI. Definitions of AGI vary but broadly it refers to an AI system that can perform a task at a human, or above human, level – and could evade our control. Max Tegmark, the scientist behind a headline-grabbing letter this year calling for a pause in large AI experiments, told the Guardian that tech professionals in California believe AGI is close. “A lot of people here think that we’re going to get to God-like artificial general intelligence in maybe three years. Some think maybe two years.” He added: “Some think it’s going to take a longer time and won’t happen until 2030.” Which doesn’t seem very far away either. There are also respected voices who think the clamour over AGI is being overplayed. According to one of those counterarguments, the noise is a cynical ploy to regulate and fence off the market and consolidate the position of big players like ChatGPT developer OpenAI, Google and Microsoft. The Distributed AI Research Institute has warned that focusing on existential risk ignores immediate impacts from AI systems such as: using artists’ and authors’ work without permission in order to build AI models; and using low-paid workers to carry out some of the model-building tasks. Timnit Gebru, founder and executive director of DAIR, last week praised a US senator for raising concerns over working conditions for data workers rather than focusing on “existential risk nonsense”. Another view is that uncontrollable AGI simply won’t happen. “Uncontrollable artificial general intelligence is science fiction and not reality,” said William Dally, the chief scientist at the AI chipmaker Nvidia, at a US senate hearing last week. “Humans will always decide how much decision-making power to cede to AI models.” However, for those who disagree, the threat posed by AGI cannot be ignored. Fears about such systems include refusing – and evading – being switched off, combining with other AIs or being able to improve themselves autonomously. Connor Leahy, the chief executive of the AI safety research company Conjecture, said the problem was more simple than that. “The deep issue with AGI is not that it’s evil or has a specifically dangerous aspect that you need to take out. It’s the fact that it is competent. If you cannot control a competent, human-level AI then it is by definition dangerous,” he said. Other concerns held by UK government officials are that the next iteration of AI models, below the AGI level, could be manipulated by rogue actors to produce serious threats such as bioweapons. Open source AI, where the models underpinning the technology are freely available and modifiable, is a related concern. Civil servants say they are also working on combating nearer-term risks, such as disinformation and copyright infringements. But with

international leaders arriving at Bletchley Park in a few weeks' time, Downing Street wants to focus the world's attention on something officials believe is not being taken seriously enough in policy circles: the chance that machines could cause serious damage to humanity.