

Google says new AI model Gemini outperforms ChatGPT in most tests

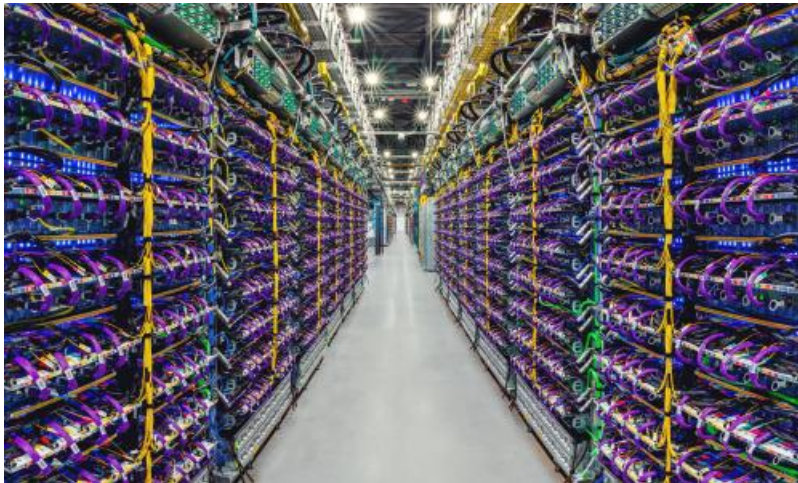
Publication Date: 2023-12-06

Author: Dan Milmo

Section: Technology

Tags: Artificial intelligence (AI), DeepMind, Google, Computing, news

Article URL: <https://www.theguardian.com/technology/2023/dec/06/google-new-ai-model-gemini-bard-upgrade>



Google has unveiled a new artificial intelligence model that it claims outperforms ChatGPT in most tests and displays “advanced reasoning” across multiple formats, including an ability to view and mark a student’s physics homework. The model, called Gemini, is the first to be announced since last month’s global AI safety summit, at which tech firms agreed to collaborate with governments on testing advanced systems before and after their release. Google said it was in discussions with the UK’s newly formed AI Safety Institute over testing Gemini’s most powerful version, which will be released next year. Google said Ultra outperformed “state-of-the-art” AI models including ChatGPT’s most powerful model, GPT-4, on 30 out of 32 benchmark tests including in reasoning and image understanding. The Pro model outperformed GPT-3.5, the technology that underpins the free-access version of ChatGPT, in six out of eight tests. The model comes in three versions and is “multimodal”, which means it can comprehend text, audio, images, video and computer code simultaneously. Gemini, which will be folded into Google products including its search engine, is being released initially in more than 170 countries including the US on Wednesday in the form of an upgrade to Google’s chatbot, Bard. However, the Bard upgrade will not be released in the UK and Europe as Google seeks clearance from regulators. Demis Hassabis, the chief executive of DeepMind, the London-based Google unit that developed Gemini, said: “It’s been the most complicated project we’ve ever worked on, I would say the biggest undertaking. It’s been an enormous effort.” Two smaller versions of Gemini, Pro and Nano, will be released on Wednesday. The Pro model can be accessed on Google’s Bard chatbot and the Nano version will be on mobile phones using Google’s Android system. The most powerful iteration, Ultra, is being tested externally and will not be released publicly until early 2024, when it will also be integrated into a version of Bard called Bard Advanced. Google said Ultra was the first AI model to outperform human experts, with a score of 90%, on a multitasking test called MMLU, which covers 57 subjects including maths, physics, law, medicine and ethics. Ultra will now power a new code-writing tool called AlphaCode2, which Google claimed could outperform 85% of competition-level human computer programmers. Hassabis said the Ultra model would undergo external “red team” testing – where experts test the security and safety of a product – and Google would share the results with the US government, in line with an executive order issued by Joe Biden in October. Asked if Gemini had been tested in collaboration with the US or UK governments, as set out at the AI safety summit at Bletchley Park, Hassabis said Google was in discussions with the UK government about the AI Safety Institute carrying out tests on the model. “We’re discussing with them how we want them to do that,” he said. The Pro and Nano models will not be part of the tests, which are for the most advanced, or “frontier”, models. Sissie Hsiao, the general manager for Bard at Google, said the Pro-powered version of Bard would not be released in the UK yet. It is also not being released in the

European Economic Area, which includes the EU and Switzerland. She said: "We are working with local regulators." Google did not specify the regulatory issues behind the delays in the UK and EU. However, Google indicated that "hallucinations", or false answers, were still a problem with the model. "It's still, I would say, an unresolved research problem," said Eli Collins, the head of product at Google DeepMind. Although all of the Gemini versions are multimodal in terms of the prompts they can comprehend, the Pro and Nano iterations being released publicly this month can currently respond only in text or code format. Google released promotional videos of Gemini's capabilities, which included showing the Ultra model understanding a student's handwritten physics homework answers and giving detailed tips on how to solve the questions, including displaying equations. Other videos showed Gemini's Pro version analysing and identifying a drawing of a duck as well as answering correctly which film a person was enacting in a smartphone video – in this case, an amateurish take on the famous "bullet time" scene in The Matrix. Collins said Gemini's most powerful mode had shown "advanced reasoning" and could show "novel capabilities" – an ability to perform tasks not shown by AI models before. Concerns over AI – the term for computer systems that can perform tasks normally requiring human intelligence – range from mass-produced disinformation to the creation of "superintelligent" systems that evade human control. Some experts are concerned about the development of artificial general intelligence, which refers to an AI that can perform an array of tasks at a human or above-human level of intelligence. Asked whether Gemini represented an important step towards AGI, Hassabis said: "I think these multimodal foundational models are going to be key component of AGI, whatever that final system turns out to be. But there's still things that are missing, which we're still researching and innovating on now." Hassabis said data used to train Gemini had been taken from a range of sources including the open web. The publishing and creative industries have protested against AI companies using copyrighted content available online to build models.