

Meta to make new version of AI model available free of charge on Microsoft

Publication Date: 2023-07-18

Author: Dan Milmo

Section: Technology

Tags: Meta, Microsoft, Artificial intelligence (AI), Mark Zuckerberg, Computing, news

Article URL: <https://www.theguardian.com/technology/2023/jul/18/ai-free-open-source-microsoft>



Mark Zuckerberg's Meta is making a commercial version of its artificial intelligence model freely available, in a move that gives startups and other businesses a low-cost opportunity compete with OpenAI's ChatGPT and Google's Bard. A new version of a Meta large language model (LLM), called Llama 2, will be distributed by Microsoft through its Azure cloud service and will run on the Windows operating system, Meta said in a blogpost, referring to Microsoft as "our preferred partner" for the release. LLMs underpin generative AI products like the ChatGPT chatbot, although ChatGPT's owner has not open-sourced – or made widely available to others – its LLM, called GPT-4. The model, which Meta previously provided only to select academics for research purposes, also will be made available via direct download and through Amazon Web Services, Hugging Face and other providers. "Open source drives innovation because it enables many more developers to build with new technology," Zuckerberg wrote in a Facebook post. "I believe it would unlock more progress if the ecosystem were more open." However, one expert said open source AI was a divisive issue, with some arguing that AI systems should not be widely available. "There are competing schools of thought about open-sourcing AI," said Dr Andrew Rogoyski, of the Institute for People-Centred AI at the University of Surrey. "There are people who think AI should be accessible to anyone, democratising AI and keeping it available for independent scrutiny, and there are people who believe that AI should be withheld, like nuclear secrets, to avoid powerful technology getting into the wrong hands." In April, researchers at Stanford University in California took down a chatbot they had built for \$600 using a version of the first Llama model after it generated unsavory text. Meta executives say they believe public releases of technologies actually reduce safety risks by harnessing the wisdom of the crowd to identify problems and build resilience into the systems. The company also says it has put in place an "acceptable use" policy for commercial Llama that prohibits "certain use cases", including violence, terrorism, child exploitation and other criminal activities. Making a model as sophisticated as Llama widely available and free for businesses to build atop threatens to upend the early dominance established in the nascent market for generative AI software by players like OpenAI, which Microsoft backs and whose models it already offers to business customers via Azure. The first Llama was already competitive with models that power OpenAI's ChatGPT and Google's Bard chatbot, while the new Llama has been trained on 40% more data than its predecessor, with more than 1m annotations by humans to fine-tune the quality of its outputs, Zuckerberg said. "Commercial Llama could change the picture," said Amjad Masad, the chief executive at software developer platform Replit, who said more than 80% of projects there use OpenAI's models. "Any incremental improvement in open-source models is eating into the market share of closed-source models because you can run them cheaply and have less dependency," said Masad. The announcement follows plans by Microsoft's largest cloud rivals,

Alphabet's Google and Amazon, to give business customers a range of AI models from which to choose. Until now, Microsoft has focused on making technology available from OpenAI in Azure. Asked why Microsoft would support an offering that might degrade OpenAI's value, a Microsoft spokesperson said giving developers choice in the types of models they use would help extend its position as the go-to cloud platform for AI work. For Meta, a flourishing open-source ecosystem of AI tech built using its models could stymie rivals' plans to earn revenue off their proprietary technology, the value of which would evaporate if developers could use equally powerful open-source systems for free. A leaked internal Google memo titled "We have no moat, and neither does OpenAI" lit up the tech world in May after it forecast just such a scenario. Meta is also betting that it will benefit from the advances, bug fixes and products that may grow out of its model becoming the go-to default for AI innovation, as it has over the past several years with its widely adopted open-source AI framework PyTorch. As a social media company, Zuckerberg told investors in April, Meta has more to gain by effectively crowd-sourcing ways to reduce infrastructure costs and maximize creation of new consumer-facing tools that might draw people to its ad-supported services than it does by charging for access to its models. "Unlike some of the other companies in the space, we're not selling a cloud computing service where we try to keep the different software infrastructure that we're building proprietary," Zuckerberg said. "For us, it's way better if the industry standardizes on the basic tools that we're using and therefore we can benefit from the improvements that others make." Releasing Llama into the wild also comes with risks, however, as it supercharges the ease with which unscrupulous actors may build products with little regard for safety controls.