Google's Gemini showcases more powerful technology, but we're still not close to superhuman AI

March 5 2024, by Christian Gagné

Credit: Pixabay/CC0 Public Domain

In December 2023, Google announced the launch of its new large language model (LLM) named Gemini. Gemini now provides the

artificial intelligence (AI) foundations of Google products; it is also a direct rival to [OpenAI's GPT-4](#).

But why is Google considering Gemini as such an important milestone, and what does this mean for users of Google's services? And generally speaking, what does it mean in the context of the current hyperfast-paced developments of AI?

AI everywhere

Google is betting on Gemini to transform most of its products by enhancing current functionalities and creating new ones for services such as search, Gmail, YouTube and its office productivity suite. This would also allow improvements to their online advertising business—their main source of revenue—as well as for Android phone software, with trimmed versions of Gemini running on limited capacity hardware.

For users, Gemini means new features and improved capacities that would make Google services harder to shun, strengthening an already dominant position in areas such as search engines. The potential and opportunities for Google are considerable, given the bulk of their software is easily upgradable cloud services.

But the huge [and unexpected success](#) of ChatGPT attracted a lot of attention and enhanced the credibility of OpenAI. Gemini will allow Google to reinstate itself as a major player in AI in the [public view](#). Google is a powerhouse in AI, with large and strong research teams at the origin of many major advances of the last decade.

There is public discussion about these new technologies, both on the benefits they provide and the disruption they create in fields such as education, design and health care.

Strengthening AI

At its core, Gemini relies on transformer networks. Originally devised by a research team at Google, the same technology is used to power other LLMs such as GPT-4.

A distinctive element of Gemini is its capacity to deal with different data modalities: text, audio, image and video. This provides the AI model with the capacity to execute tasks over several modalities, like answering questions regarding the content of an image or conducting a keyword search on specific types of content discussed in podcasts.

But more importantly, that the models can handle distinct modalities enables the training of globally superior AI models, compared to distinct models trained independently for each modality. Indeed, such multimodal models are deemed to be stronger since they are exposed to different perspectives of the same concepts.

For example, the concept of birds may be better understood through learning from a mix of birds' textual descriptions, vocalizations, images and videos. This idea of multimodal transformer models has been explored in previous research at Google, Gemini being the first full-fledged commercial implementation of the approach.

Such a model is seen as a step in the direction of stronger generalist AI models, also known as artificial general intelligence (AGI).

Risks of AGI

Given the rate at which AI is advancing, the expectations that AGI with superhuman capabilities will be designed in the near future generates discussions in the research community and more broadly in the society.

On one hand, some anticipate the risk of catastrophic events if a powerful AGI falls into the hands of ill-intentioned groups, and request that developments be slowed down.

Others claim that we are still very far from such actionable AGI, that the current approaches allow for a shallow modeling of intelligence, mimicking the data on which they are trained, and lack an effective world model—a detailed understanding of actual reality—required to achieve human-level intelligence.

On the other hand, one could argue that focusing the conversation on existential risk is distracting attention from more immediate impacts brought on by recent advances of AI, including perpetuating biases, producing incorrect and misleading content—prompting Google to pause its Gemini image generator, increasing environmental impacts and enforcing the dominance of Big Tech.

The line to follow lies somewhere in between all of these considerations. We are still far from the advent of actionable AGI—additional breakthroughs are required, including introducing stronger capacities for symbolic modeling and reasoning.

In the meantime, we should not be distracted from the important ethical and societal impacts of modern AI. These considerations are important and should be addressed by people with diverse expertise, spanning technological and social science backgrounds.

Nevertheless, although this is not a short-term threat, achieving AI with superhuman capacity is a matter of concern. It is important that we, collectively, become ready to responsibly manage the emergence of AGI when this significant milestone is reached.

This article is republished from The Conversation under a Creative

Commons license. Read the original article.

Provided by The Conversation