# GOOGLE GEMINI

Relevant for: Science & Technology | Topic: Science and Technology- developments and their applications and effects in everyday life

To enjoy additional benefits

CONNECT WITH US

December 10, 2023 01:05 am | Updated 02:09 am IST

COMMents

SHARE

READ LATER

Gemini has shifted AI in a direction more expansive that just a talking chatbot | Photo Credit: Dado Ruvic

For a year now, Google has been playing catch-up with OpenAI. Since the release of ChatGPT marked a momentous occasion in what has become the age of AI, the lumbering search giant was seen scrambling to put their next foot forward. Google, a company that was aggressive in releasing AI research but slow at releasing tools to the public, had been outmanoeuvred by a nifty startup. The threat of the AI chatbot was great enough for CEO Sundar Pichai to pull the fire alarm and declare a 'Code Red' situation at the company. Founders Sergei Brin and Larry Page came out of retirement at Mr. Pichai's behest.

After reports of delays and a long wait, Google released their [new AI model Gemini](#) on Wednesday. And now was as opportune a moment as any. A couple of weeks ago, OpenAI had been caught in a board coup that had ended up temporarily ousting CEO Sam Altman. Google was certainly looking to capitalise on the ripple of uncertainty that had shaken up its competitor.

Google's treasure trove of multimodal data from search and YouTube had come to its rescue. Gemini had been trained to learn about the world like a baby — changing our perception of what a large language model is supposed to be. It didn't just read data and seemingly regurgitate it; it could understand what an image or an audio was. This multimodal ability was a much rounder way of "intelligence".

Where the standard approach to build multi-modal models usually means training the different components for different modalities, Gemini was trained on multiple modalities from the ground-up. Because of this Google termed Gemini "natively multimodal".

Demo videos of the model drew impressed reactions. There were things Gemini was seen doing in the videos that we haven't seen any AI model do as yet. Like it could figure out that a dot-to-dot picture was a crab even before it had been finished, or even track a ball of paper from under a plastic cup and spot sleight-of-hand tricks.

Unlike most models which are trained on graphics processing units or GPUs, Gemini was trained using Google's in-house designed tensor processing units or TPUs, which bodes well considering the overarching GPU shortages that plague most companies building their own AI models.

Gemini comes in three sizes meant for a range of platforms — Nano was designed for on-device tasks like summarising text and making suggestions in chat applications; Gemini Pro was the model currently underlying its AI-powered chatbot Bard; and Gemini Ultra, the multimodal version, will be released sometime next year once trust and safety checks are completed. The model will be made available to developers through Google Cloud's API from December 13. Gemini is also the most product-oriented than most models in the market as it is enmeshed in the Google ecosystem.

**Also read | What is multimodal artificial intelligence and why is it important?**

Some digging into Google's claims revealed some more truths. Wharton professor Ethan Mollick demonstrated that ChatGPT could comfortably replicate some of the tasks that had initially seemed impressive in the Gemini demo, like analysing an image step-by-step. Another associate professor from the University of Wisconsin-Madison, Dimitris Papailiopoulos, tried 14 examples of multimodal reasoning that the Gemini research paper had presented, on ChatGPT-4. GPT4V got 12 of these instances right with a couple of responses even better than Gemini's.

Google also admitted that the demo videos were edited to shorten the response time. Inquiries made by Bloomberg revealed that the seemingly flowing conversation between Gemini and the user in the video had been an inserted voice. In reality, the prompts were made via text while the model was shown images consecutively. The embarrassing gaffe made in the live demo during Bard's release was something that the company desperately would have wanted to avoid. But despite the caveat of good marketing, Gemini has shifted AI in a direction more expansive that just a talking chatbot.

COMMents

SHARE

Artificial Intelligence / internet / emerging technologies / The Hindu Profiles

BACK TO TOP

Comments have to be in English, and in full sentences. They cannot be abusive or personal. Please abide by our community guidelines for posting your comments.

We have migrated to a new commenting platform. If you are already a registered user of The Hindu and logged in, you may continue to engage with our articles. If you do not have an account please register and login to post comments. Users can access their older comments by logging into their accounts on Vuukle.