Sections ▉

Newsletters ✉

f 𝕏 ▶ 📷

EDITION
INDIA ⌄

The Indian EXPRESS
JOURNALISM OF COURAGE

Monday, Apr 22, 2024

EPAPER   TODAY'S PAPER

TRENDING | UPSC Pack | IPL | Premium Stories | Express Shorts | Mini Games | 🎙 Podcast | Health & Wellness | Board Exam
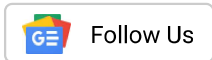
News / Explained / Explained Sci-Tech / What is Llama 3, Meta's most sophisticated and capable large language model yet?

What is Llama 3, Meta's most sophisticated and capable large language model yet?

Meta will be integrating its latest model into its proprietary virtual assistant — Meta AI. It claims Llama 3 outperformed Gemma 7B, Claude 3 Sonnet, and Mistral 7B in various benchmarks.

Written by **Bijin Jose**

New Delhi | Updated: April 21, 2024 12:58 IST

✓NewsGuard

Llama or Large Language Model Meta AI is a family of LLMs introduced by Meta AI in February 2023. (Photo: Meta)

Meta on Thursday (April 18) **introduced its most capable Large Language Model (LLM)** , the Meta Llama 3. The company also introduced an image generator, which updates pictures in real-time even as the user types out the prompt. Meta will be integrating its latest model into its proprietary virtual assistant — Meta AI.

Meta is pitching its latest models as the most sophisticated AI models, steering way ahead of its peers such as Google, Mistral, etc., in terms of performance and capabilities. The updated Meta AI assistant will be integrated into Facebook, Instagram, WhatsApp, Messenger, and a standalone website much like OpenAI's ChatGPT.

Here we take a look at what exactly Meta's Llama 3 model is, and how it is different, and why Meta claims it to be its most capable model yet.

What is Llama 3?

Llama or Large Language Model Meta AI is a family of LLMs introduced by Meta AI in February 2023. The first version of the model was released in four sizes — 7B,

13B, 33B, and 65 billion parameters. Reportedly, the 13B model of Llama outperformed OpenAI's GPT-3 which had 135 billion parameters.

To simplify, parameters here are a measure of the size and complexity of an AI model and generally, a larger number of parameters means an AI model is more complex and powerful. Meta released Llama 2 in July last year, a significantly upgraded version of its first LLM. Llama 2 was released in 7B, 13B, and 70B parameters and it was trained on 40 per cent more data when compared to its predecessor.

Now, Meta is back with Llama 3, the latest iteration of its LLM which is claimed to be the most sophisticated model with significant progress in terms of performance and AI capabilities. Llama 3, which is based on the Llama 2 architecture, has been released in two sizes, 8B and 70B parameters. Both sizes come with a base model and an instruction-tuned version that has been designed to augment performance in specific tasks. Reportedly, the instruction-tuned version is meant for powering AI chatbots that are meant to hold conversations with users.

According to Meta, with Llama 3, the company has built the best open-source models that are on par with the best proprietary models available today. It has also embraced an open-source ethos of releasing early and enabling the dev community — a community of software engineers — to get access to the models while they are still in development. F

or now, Meta has released text-based models in the Llama 3 collection of models. However, the company has plans to make Llama 3 multilingual and multimodal,

accept longer context, all while continuing to improve performance across LLM abilities such as coding and reasoning.

All models of Llama 3 support context lengths of 8,000 tokens. This allows for interactions, and complex input handling compared to Llama 2 or 1. More t[...] here mean more content input or prompts from users and more content as a[...] response from the model. When it comes to safety, Meta has said that it is dedicated to developing Llama 3 in a responsible way. "We're offering various resources to help others use it responsibly as well. This includes introducing new trust and safety tools with Llama Guard 2, Code Shield, and CyberSec Eval 2," the company said in its blog.

Also in Explained | **What is OpenAI's GPT-4 Vision and how can it help you interpret images, charts?**

How good is Llama 3?

Meta claims that the 8B and 70B parameter Llama 3 models are a giant leap from Llama 2. This is possible owing to the improvements in pretraining and post-training. "Our pretrained and instruction-fine-tuned models are the best models existing today at the 8B and 70B parameter scale," the company said on its website. According to the company, the post-training processes have led to greatly improved capabilities like reasoning, code generation, and instruction following making Llama 3 more steerable.

Meta claims that in benchmark evaluations, Llama 3 8B surpassed other open-source AIs like Mistral 7B and Gemma 7B. Llama 3 outperformed Google's Gemma 7B and Mistral's Mistral 7B, Anthropic's Claude 3 Sonnet in benchmarks such as MMLU 5-shot (Massive Multitask Language Understanding), GPQA 0-shot (A Graduate-Level Google-Proof Q&A Benchmark), HumanEval 0-shot (a benchmark for evaluating the multilingual ability of code generative models), GSM-8K 8-shot and Math 4-shot, CoT (maths and word problems).

Although Meta has not officially stated the use cases of Llama 3. Considering that it is similar to existing AI chatbots, Llama 3 can be used to create different forms of

texts such as poems, code, scripts, and musical pieces. It can summarise factual topics and can also be used to translate languages.

How to try Llama 3

Meta said that it has integrated Llama 3 into Meta AI which can be used on Facebook, Instagram, WhatsApp, Messenger, and the web. It is readily available for developers as Meta has integrated the LLM into the Hugging Face ecosystem. It is also available via Perplexity Labs, Fireworks AI, and on cloud provider platforms such as Azure ML and Vertex AI.

Llama 3 models will soon be available on AWS, Google Cloud, Hugging Face, Databricks, Kaggle, IBM WatsonX, Microsoft Azure, NVIDIA NIM, Snowflake, etc.

At present, Meta AI is available in English across the US on WhatsApp. Meta is also expanding to more countries including Australia, Canada, Ghana, Jamaica, Malawi, New Zealand, Nigeria, Pakistan, Singapore, South Africa, Uganda, Zimbabwe, and Zambia.

Bijin Jose   Follow   X

Bijin Jose, an Assistant Editor at Indian Express Online in New Delhi, is a technology journalist with a portfolio spanning various prestigious publications. Starting as a **... Read More**

First uploaded on: 20-04-2024 at 13:12 IST

TAGS: Explained Sci-Tech   Express Explained

EXPRESS *Shorts*