[Meta](https://de.wikipedia.org/wiki/Meta_Platforms) 's LLaMA (Large Language Model Meta AI) is a so-called **large**[language model](https://de.wikipedia.org/wiki/Gro%C3%9Fes_Sprachmodell) released by [Meta AI](https://de.wikipedia.org/wiki/Meta_AI) on February 24, 2023. [[ 1 ]](https://de.wikipedia.org/wiki/LLaMA-Sprachmodell#cite_note-1) It is a generative [language model](https://de.wikipedia.org/wiki/Sprachmodell) capable of generating human-like text, translating languages, writing various types of creative content, and answering questions informatively. [[ 2 ]](https://de.wikipedia.org/wiki/LLaMA-Sprachmodell#cite_note-:0-2)[[ 3 ]](https://de.wikipedia.org/wiki/LLaMA-Sprachmodell#cite_note-:1-3)[[ 4 ]](https://de.wikipedia.org/wiki/LLaMA-Sprachmodell#cite_note-4)

**publication**

[ [Edit](https://de.wikipedia.org/w/index.php?title=LLaMA-Sprachmodell&veaction=edit&section=1) | [Edit source](https://de.wikipedia.org/w/index.php?title=LLaMA-Sprachmodell&action=edit&section=1) ]

LLaMA was announced on February 23, 2023, in a blog post and a scientific paper describing the model's training, architecture, and performance. [[ 2 ]](https://de.wikipedia.org/wiki/LLaMA-Sprachmodell#cite_note-:0-2) The code used to run the model has been released under the [open-source GPL 3 license](https://de.wikipedia.org/wiki/GNU_General_Public_License) and is available via [Github](https://de.wikipedia.org/wiki/GitHub) . [[ 5 ]](https://de.wikipedia.org/wiki/LLaMA-Sprachmodell#cite_note-5) Access to the model is provided on a case-by-case basis to academic researchers, employees of government, civil society, and academic organizations, as well as industrial research laboratories around the world. [[ 6 ]](https://de.wikipedia.org/wiki/LLaMA-Sprachmodell#cite_note-6)

**Details**

[ [Edit](https://de.wikipedia.org/w/index.php?title=LLaMA-Sprachmodell&veaction=edit&section=2) | [Edit source](https://de.wikipedia.org/w/index.php?title=LLaMA-Sprachmodell&action=edit&section=2) ]

For the first version of LLaMA, four model sizes were trained: 7, 13, 33, and 65 billion [parameters](https://de.wikipedia.org/wiki/Parameter_(K%C3%BCnstliche_Intelligenz)) . [[ 7 ]](https://de.wikipedia.org/wiki/LLaMA-Sprachmodell#cite_note-:2-7) The LLaMA developers reported that the performance of the 13B-parameter model exceeded the performance of the much larger GPT - [3 (with 175 billion parameters) on most](https://de.wikipedia.org/wiki/Generative_Pre-trained_Transformer_3)[NLP](https://de.wikipedia.org/wiki/Neurolinguistik)[benchmarks](https://de.wikipedia.org/wiki/Benchmark) , and that the largest model was competitive with current models such as [PaLM](https://de.wikipedia.org/wiki/PaLM) and [Chinchilla](https://de.wikipedia.org/w/index.php?title=Chinchilla_Sprachmodell&action=edit&redlink=1) , a large language model unveiled by Google in March 2022. [[ 3 ]](https://de.wikipedia.org/wiki/LLaMA-Sprachmodell#cite_note-:1-3)[[ 7 ]](https://de.wikipedia.org/wiki/LLaMA-Sprachmodell#cite_note-:2-7)

The LLaMA model was trained on a dataset of 1.4 trillion tokens obtained from publicly available data sources, including: [[ 3 ]](https://de.wikipedia.org/wiki/LLaMA-Sprachmodell#cite_note-:1-3)

* Websites scraped by [Common Crawl](https://de.wikipedia.org/w/index.php?title=Common_Crawl&action=edit&redlink=1)
* Open source repositories with source code from [GitHub](https://de.wikipedia.org/wiki/GitHub)
* Wikipedia in 20 different languages
* Public domain books from [Project Gutenberg](https://de.wikipedia.org/wiki/Project_Gutenberg)
* The [LaTeX](https://de.wikipedia.org/wiki/LaTeX) source code for scientific papers uploaded to [ArXiv](https://de.wikipedia.org/wiki/ArXiv)
* Questions and answers from [Stack Exchange](https://de.wikipedia.org/wiki/Stack_Exchange) websites

LLaMa version 2 was introduced by Meta together with Microsoft on July 19, 2023. [[ 8 ]](https://de.wikipedia.org/wiki/LLaMA-Sprachmodell#cite_note-8) According to Meta, 40 percent more data was used for training and twice as much context was included. LLaMA 2 is available in three different sizes: with seven billion, 13 billion, and 70 billion parameters.

On July 23, 2024, Meta published the Llama 3.1 model with 8, 70, or 405 billion parameters as a freely accessible version. [[ 9 ]](https://de.wikipedia.org/wiki/LLaMA-Sprachmodell#cite_note-9) The latest model, Llama 3.2, released in September 2024, is not yet freely accessible in Europe. [[ 10 ]](https://de.wikipedia.org/wiki/LLaMA-Sprachmodell#cite_note-10)