LMArena (formerly Chatbot Arena) is a public, web-based platform that evaluates large language models (LLMs) through anonymous, crowd-sourced pairwise comparisons. Users enter prompts for two anonymous models to respond to and vote on the model that gave the better response, in which the model's identities are then revealed. Users can also choose models to test themselves. [1] [2]

LMArena is popular within the artificial intelligence industry, with major companies supplying their large language models, such as OpenAI 's GPT-4o and o1 , Google DeepMind 's Gemini , [3] and Anthropic 's Claude , [4] and using their subsequent rankings to promote them.

The website has even been used for preview releases of upcoming models. Notably, Chinese company DeepSeek tested its prototype models in the LMArena months before its R1 model gained attention in Western media. [5] Other notable pre-release models include OpenAI's GPT-5 under the codename "summit" and Google DeepMind's Gemini 2.5 Flash Image, an image generation and editing model, under the codename "nano-banana" . [6] [7]

LMArena's evaluation methodology for large language models has been examined in academic analyses, which have identified specific limitations and suggested areas for improvement. The platform is an active contributor of the AI research ecosystem and has since implemented methodological updates in coordination with ongoing research through its policy updates. [8] [9]

References

External links

Official website

v

t

e

LMArena

List of chatbots

List of LLMs

character.ai

ChatGPT

Claude

Command

Copilot

DeepSeek

Ernie

Gemini

GLM

Grok

Hunyuan

Kimi

Llama

Mistral

Perplexity

Poe

Qwen

You.com

Category

This large language model -related article is a stub . You can help Wikipedia by expanding it .

v

t

e