

Current benchmarks

GPT-3.5-turbo

GPT-4 (2023-06-13)

EvoBench

Developing

Fine-tuning

Sheared-LLaMA
Sheared-LLaMA-Pruned
Sheared-LLaMA-Pruned
Sheared-LLaMA

Llama-2

Llama-2-7b-chat
Vicuna
WizardMath

Pruning

LLama-3.1

LLama-3.2

gpt-4-1106

GPT-4

gpt-4-0613

GPT-4o

gpt-4o-mini

gpt-4-0409

gpt-4o-2024-11-20

gpt-4o-2024-08-06

gpt-4o-2024-05-13

Updating

