

Open LLM Leaderboard:

T	Model	Average	ARC	HellaSwag	MMLU	TruthfulQA	Winogrande	GSMBK
◆	daavidkim285/Rhea-72b-v0.5	81.22	79.78	91.15	77.95	74.5	87.85	76.12
◆	SF-Foundation/Ein-72B-v0.11	80.81	76.79	89.82	77.2	79.82	84.86	78.77
◆	SF-Foundation/Ein-72B-v0.13	80.79	76.19	89.44	77.87	77.82	84.93	79.3
◆	SF-Foundation/Ein-72B-v0.12	80.72	76.19	89.46	77.17	77.78	84.45	79.23
◆	abacusai/Smaug-72B-v0.1	80.48	76.82	89.27	77.15	76.67	85.88	78.7
◆	ibivibiv/alpaca-dragon-72b-v1	79.3	73.89	88.16	77.4	72.69	86.83	77.63
●	moreh/MoMo-72B-lora-1.8.7-DPO	78.55	78.82	85.96	77.13	74.71	84.86	78.62
◆	cloudyu/TomGrc_FusionNet_34Bx2_MoE_v0.1_DP0_f16	77.91	74.86	86.74	76.65	72.24	83.35	74.45
◆	saltlux/luxia-21.4b-alignment-v1.0	77.74	77.47	91.88	68.1	79.17	87.45	62.4
◆	cloudyu/TomGrc_FusionNet_34Bx2_MoE_v0.1_full_linear_DP0	77.52	74.86	86.67	76.69	71.32	83.43	72.93
◆	zhengx/MixTAO-7Bx2-MoE-v0.1	77.5	73.81	89.22	64.92	78.57	87.37	71.11
●	yunconglong/Truthful_DP0_TomGrc_FusionNet_7Bx2_MoE_13B	77.44	74.91	89.3	64.67	78.82	88.24	69.52

LMSYS Chatbot Arena Leaderboard:

<https://github.com/lm-sys/FastChat>

<https://lmsys.org/blog/2023-05-03-arena/>

Model	Arena Elo	MT-bench	MMLU	Organization	License
Claude 3 Opus	1253		86.8	Anthropic	Proprietary
GPT-4-1106-preview	1251	9.32		OpenAI	Proprietary
GPT-4-0125-preview	1248			OpenAI	Proprietary
Bard (Gemini Pro)	1203			Google	Proprietary
Claude 3 Sonnet	1198		79	Anthropic	Proprietary
GPT-4-0314	1185	8.96	86.4	OpenAI	Proprietary
Claude 3 Haiku	1179		75.2	Anthropic	Proprietary
GPT-4-0613	1158	9.18		OpenAI	Proprietary
Mistral-Large-2402	1157		81.2	Mistral	Proprietary
Qwen1.5-72B-Chat	1148	8.61	77.5	Alibaba	Qianwen LICENSE
Claude-1	1146	7.9	77	Anthropic	Proprietary
Mistral Medium	1145	8.61	75.3	Mistral	Proprietary
Starling-LM-7B-beta	1127	8.12		UC Berkeley	Apache-2.0
Claude-2.0	1126	8.86	78.5	Anthropic	Proprietary
Gemini Pro (Dev API)	1125		71.8	Google	Proprietary
Mistral-Next	1122			Mistral	Proprietary