🏆 Chatbot Arena LLM Leaderboard: Community-driven Evaluation for Best LLM and AI chatbots

Twitter | Discord | Blog | GitHub | Paper | Dataset | Kaggle Competition

Vote!

This is a mirror of the live leaderboard created and maintained at https://lmarena.ai/leaderboard. Please link to the original URL for citation purposes.

Chatbot Arena (lmarena.ai) is an open-source platform for evaluating AI through human preference, developed by researchers at UC Berkeley SkyLab and LMSYS. With over 1,000,000 user votes, the platform ranks best LLM and AI chatbots using the Bradley-Terry model to generate live leaderboards. For technical details, check out our paper.

Chatbot Arena thrives on community engagement — cast your vote to help improve AI evaluation!

New Launch! Copilot Arena: VS Code Extension to compare Top LLMs

Arena | 📣 NEW: Overview | Arena (Vision) | Arena-Hard-Auto

Full Leaderboard

Total #models: **203**. Total #votes: **2,645,647**. Last updated: 2025-02-11.

Code to recreate leaderboard tables and plots in this notebook. You can contribute your vote at lmarena.ai!

Category

Overall

Apply filter

Style Control | Show Deprecated

Overall Questions

#models: **203** (100%) #votes: **2,645,647** (100%)

Rank* (UB)	Rank (Style	Model ▲	Arena Score	95% CI	Votes	Organi	Lic
1	3	Gemini-2.0-Flash-Thinking-Exp-01-21	1384	+5/-5	11949	Google	Proprietary
1	2	Gemini-2.0-Pro-Exp-02-05	1379	+4/-7	9872	Google	Proprietary
1	1	ChatGPT-4o-latest (2025-01-29)	1377	+6/-6	8643	OpenAI	Proprietary
4	2	DeepSeek-R1	1361	+9/-9	4190	DeepSeek	MIT
4	7	Gemini-2.0-Flash-001	1355	+7/-6	7747	Google	Proprietary
4	2	o1-2024-12-17	1352	+5/-5	13976	OpenAI	Proprietary
7	5	o1-preview	1335	+4/-4	33164	OpenAI	Proprietary
7	7	Qwen2.5-Max	1332	+6/-6	5968	Alibaba	Proprietary
9	8	DeepSeek-V3	1316	+6/-5	16456	DeepSeek	Deepseek
9	9	Gemini-2.0-Flash-Lite-Preview-02-05	1309	+9/-6	7101	Google	Proprietary
				+8/-		Open	Pro

9	8	o3-mini	1307	7	7177	AI	iet y
9	15	GLM-4-Plus-0111	1304	+8/-8	3713	Zhipu	Pro iet y
9	15	Step-2-16K-Exp	1304	+8/-7	5135	StepFun	Pro iet y
10	15	o1-mini	1304	+3/-3	54641	OpenAI	Pro iet y
10	10	Gemini-1.5-Pro-002	1302	+3/-3	51685	Google	Pro iet y
16	18	Grok-2-08-13	1288	+3/-3	67067	xAI	Pro iet y
16	18	Yi-Lightning	1287	+4/-3	28956	01 AI	Pro iet y
16	11	GPT-4o-2024-05-13	1285	+2/-2	117728	OpenAI	Pro iet y
16	7	Claude 3.5 Sonnet (20241022)	1283	+3/-3	53136	Anthropic	Pro iet y
16	26	Qwen2.5-plus-1127	1282	+5/-6	10740	Alibaba	Pro iet y
17	20	Deepseek-v2.5-1210	1279	+6/-7	7248	DeepSeek	Dee eek
20	28	Athene-v2-Chat-72B	1275	+4/-3	26098	NexusFlow	Nex Flo

20	26	GLM-4-Plus	1274	+4/-3	27761	Zhipu AI	Proprietary
21	26	GPT-4o-mini-2024-07-18	1273	+3/-2	63623	OpenAI	Proprietary
21	29	Gemini-1.5-Flash-002	1271	+5/-3	37007	Google	Proprietary
21	41	Llama-3.1-Nemotron-70B-Instruct	1268	+7/-7	7597	Nvidia	Llama 3.1
21	15	Meta-Llama-3.1-405B-Instruct-bf16	1268	+4/-4	27932	Meta	Llama 3.1 Community
23	13	Claude 3.5 Sonnet (20240620)	1268	+2/-3	86160	Anthropic	Proprietary
24	15	Meta-Llama-3.1-405B-Instruct-fp8	1267	+3/-3	63168	Meta	Llama 3.1 Community
25	15	Gemini Advanced App (2024-05-14)	1267	+3/-3	52160	Google	Proprietary
25	36	Grok-2-Mini-08-13	1266	+3/-3	55436	xAI	Proprietary
25	28	Qwen-Max-0919	1263	+5/-4	17433	Alibaba	Qwen
26	16	GPT-4o-2024-08-06	1265	+3/-3	47959	OpenAI	Proprietary

30	22	Gemini-1.5-Pro-001	1260	+3/-3	82433	Google	Proprietary	
31	32	Deepseek-v2.5	1258	+4/-3	26337	DeepSeek	Deepseek	
33	39	Qwen2.5-72B-Instruct	1257	+3/-3	41532	Alibaba	Qwen	
33	22	GPT-4-Turbo-2024-04-09	1256	+3/-2	102121	OpenAI	Proprietary	
33	27	Llama-3.3-70B-Instruct	1255	+4/-5	22318	Meta	Llama-3.3	
35	37	Athene-70B	1250	+5/-4	20598	NexusFlow	CC-BY-NC-4.0	
35	42	Llama-3.1-Tulu-3-70B	1244	+12/-9	3018	Ai2	Llama 3.1	
37	28	Mistral-Large-2407	1252	+3/-3	48183	Mistral	Mistral Research	
38	27	GPT-4-1106-preview	1250	+2/-2	103718	OpenAI	Proprietary	
38	44	Meta-Llama-3.1-70B-Instruct	1248	+3/-3	58760	Meta	Llama 3.1 Community	
38	40	Mistral-Large-2411	1247	+5/-4	17377	Mistral	MRL	
39	35	Claude-3-Opus	1247	+2/-	2026	Anthropic	Proprietary	

39	25	Claude 3 Opus	1247	2	69	Topic	Proprietary
39	45	Amazon Nova Pro 1.0	1245	+5/-4	16336	Amazon	Proprietary
40	28	GPT-4-0125-preview	1245	+3/-2	97048	OpenAI	Proprietary
45	45	Reka-Core-20240904	1235	+8/-6	7936	Reka AI	Proprietary
46	27	Claude 3.5 Haiku (20241022)	1236	+4/-5	17502	Anthropic	Proprieta
49	47	Gemini-1.5-Flash-001	1227	+3/-3	65652	Google	Proprietary
50	45	Jamba-1.5-Large	1221	+5/-6	9118	AI21 Labs	Jam Ope
50	53	Mistral-Small-24B-Instruct-2501	1218	+9/-9	2455	Mistral	Apa e 2
51	46	Gemma-2-27B-it	1220	+3/-2	75867	Google	Gem lic se
51	60	Amazon Nova Lite 1.0	1219	+4/-5	14220	Amazon	Proprietary
51	53	Qwen2.5-Coder-32B-Instruct	1217	+7/-8	5723	Alibaba	Apa e 2
51	47	Gemma-2-9B-it-SimPO	1216	+5/-4	10555	Princeton	MIT

51	48	Command R+ (08-2024)	1215	+6/-6	10546	Cohere	CC-BY-NC-4.0
51	45	Llama-3.1-Nemotron-51B-Instruct	1211	+7/-6	3895	Nvidia	Llama 3.1
52	61	Gemini-1.5-Flash-8B-001	1213	+4/-4	37693	Google	Proprietary
54	60	Aya-Expanse-32B	1209	+4/-4	28767	Cohere	CC-BY-NC-4.0

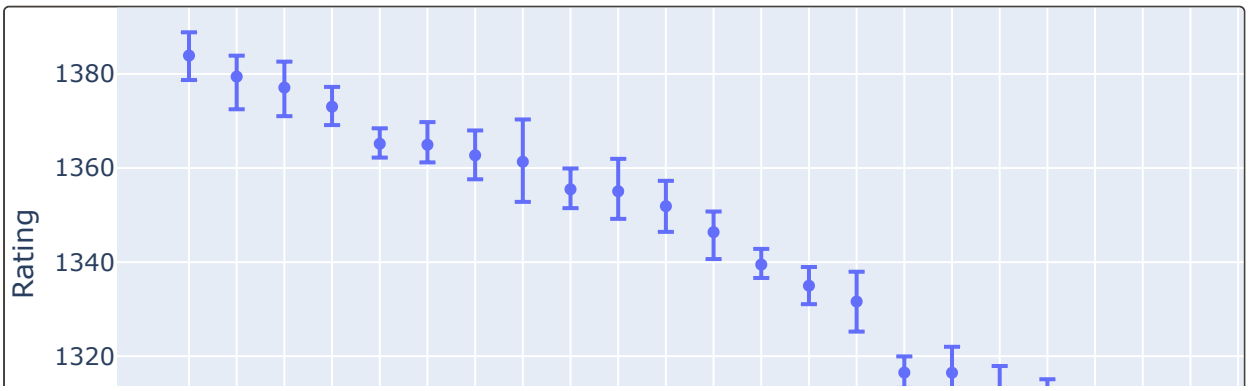
*__Rank (UB)__: model's ranking (upper-bound), defined by one + the number of models that are statistically better than the target model. Model A is statistically better than model B when A's lower-bound score is greater than B's upper-bound score (in 95% confidence interval). See Figure 1 below for visualization of the confidence intervals of model scores.

__Rank (StyleCtrl)__: model's ranking with style control, which accounts for factors like response length and markdown usage to decouple model performance from these potential confounding variables. See blog post for further details.

Note: in each category, we exclude models with fewer than 300 votes as their confidence intervals can be large.

More Statistics for Chatbot Arena (Overall)

Figure 1: Confidence Intervals on Model Strength (via Bootstrapping)



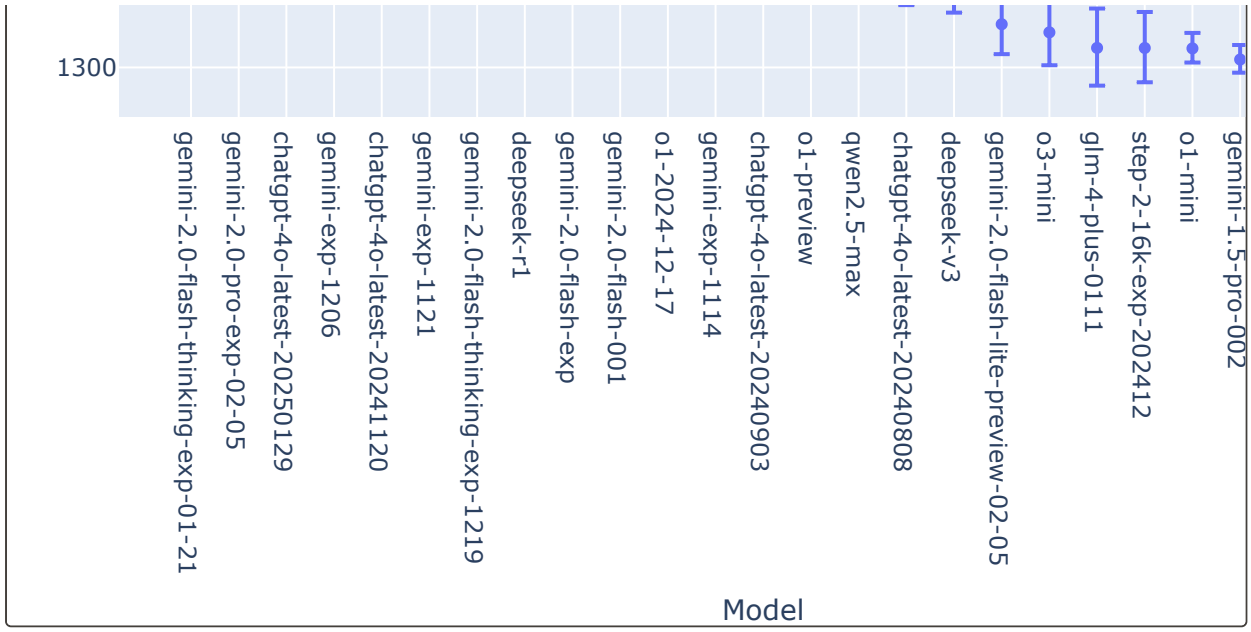


Figure 2: Average Win Rate Against All Other Models (Assuming Uniform Sampling and No Ties)

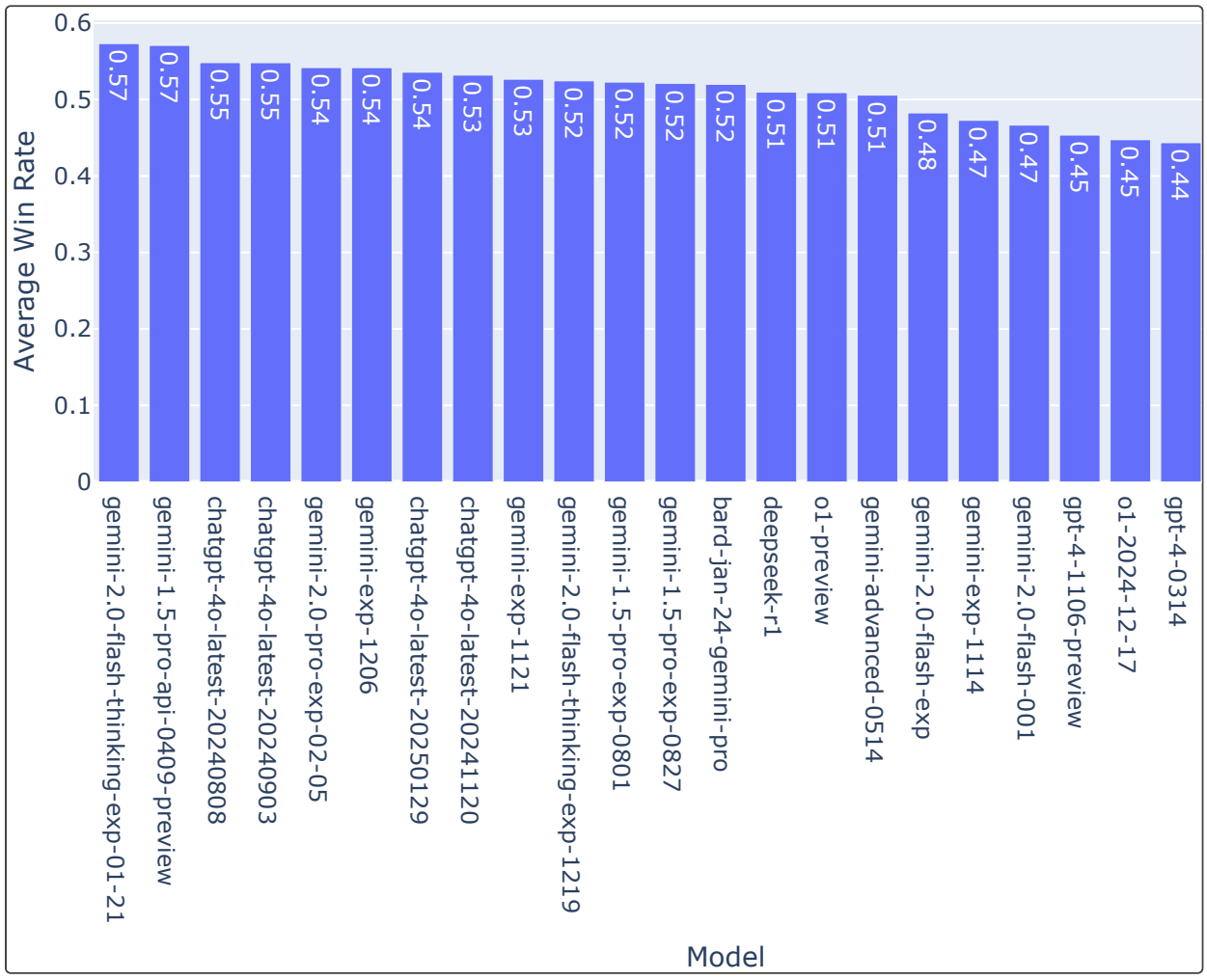


Figure 3: Fraction of Model A Wins for All Non-tied A vs. B Battles

Model A
gemini-2.0-flash-thinking-exp-01-21
gemini-2.0-pro-exp-02-05
chatgpt-4o-latest-20250129
gemini-exp-1206
chatgpt-4o-latest-20241120
gemini-exp-1121
gemini-2.0-flash-thinking-exp-1219
deepseek-r1
gemini-2.0-flash-exp
gemini-2.0-flash-001
o1-2024-12-17
gemini-exp-1114
chatgpt-4o-latest-20240903
o1-preview
qwen2.5-max
chatgpt-4o-latest-20240808
deepseek-v3
gemini-2.0-flash-lite-preview-02-05
o3-mini
glm-4-plus-0111
step-2-16k-exp-202412
o1-mini
gemini-1.5-pro-002
gemini-1.5-pro-exp-0827
gemini-1.5-pro-exp-0801