

Total #models: 169. Total #votes: 2,296,813. Last updated: 2024-11-21.

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: 169 (100%) #votes: 2,296,813 (100%)

Rank* (UB) ▲	Rank (StyleCtrl) ▲	Model ▲	Arena Score ▲	95% CI ▲	Votes ▲	Organization ▲	License ▲
1	2	Gemini-Exp-1121	1365	+8/-6	5625	Google	Proprietary
1	1	ChatGPT-4o-latest (2024-11-20)	1361	+4/-5	10658	OpenAI	Proprietary
3	5	Gemini-Exp-1114	1344	+4/-5	12778	Google	Proprietary
4	2	o1-preview	1334	+4/-4	27835	OpenAI	Proprietary
5	7	o1-mini	1308	+3/-4	31992	OpenAI	Proprietary
5	5	Gemini-1.5-Pro-002	1301	+5/-3	27336	Google	Proprietary
7	10	Grok-2-08-13	1289	+4/-3	52102	xAI	Proprietary
7	12	Yi-Lightning	1287	+4/-3	29336	01 AI	Proprietary
7	5	GPT-4o-2024-05-13	1285	+2/-2	111745	OpenAI	Proprietary
8	3	Claude 3.5 Sonnet (20241022)	1282	+4/-3	29454	Anthropic	Proprietary

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

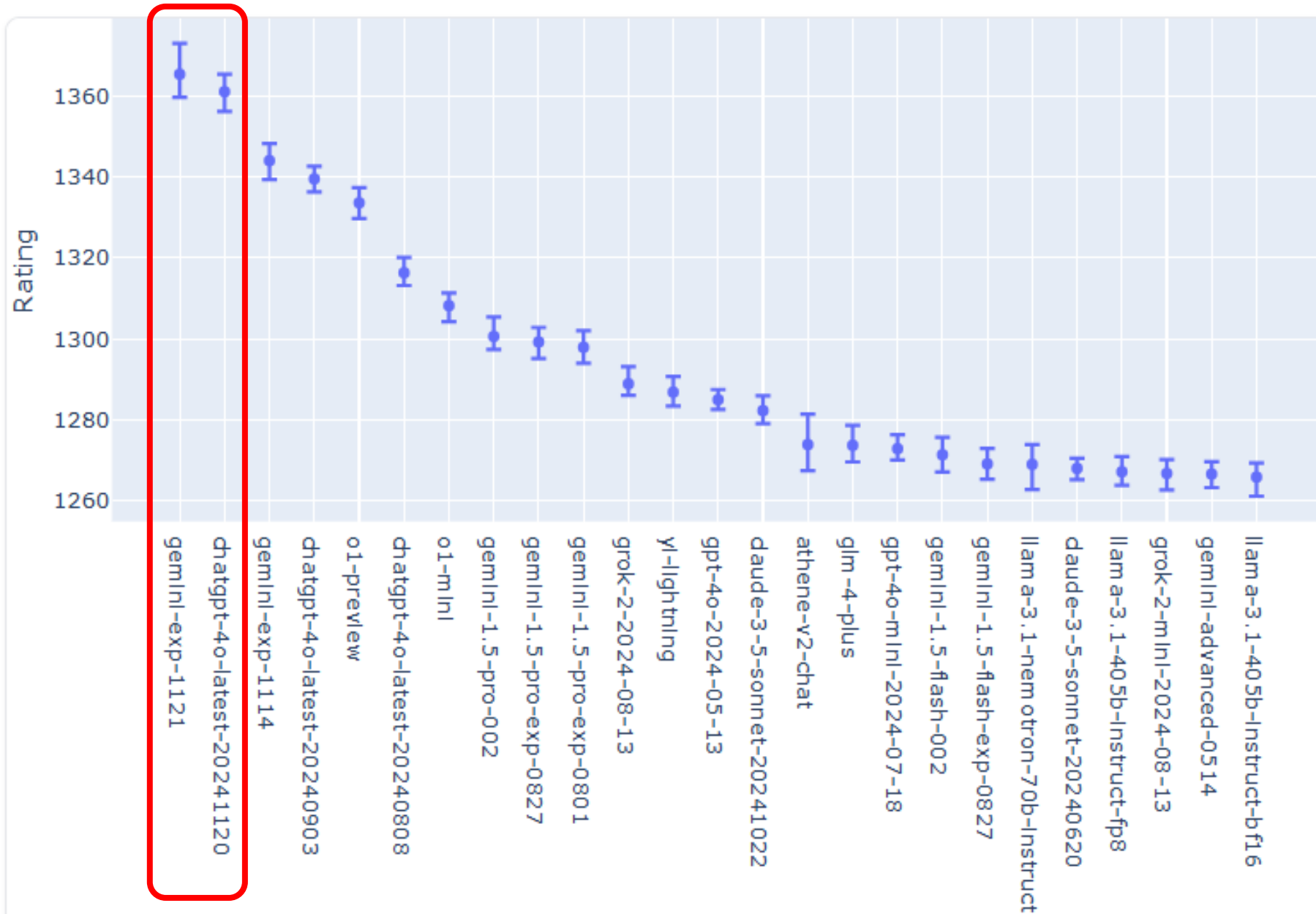


Figure 2: Average Win Rate Against All Other Models (Assuming Uniform Sampling and No Ties)

