

Exciting news - the latest Arena result are out!

@cohere's Command R+ has climbed to the 6th spot, matching GPT-4-0314 level by 13K+ human votes! It's undoubtedly the **\*\*best\*\*** open model on the leaderboard now 🔥

Big congrats to @cohere's incredible work & valuable contribution to the open community!

More exciting updates:

- Qwen1.5-32B-Chat almost top-10
- Gemma-1.1-7B-it shows great improvement (1044 -> 1088, on par with Llama-2-70b)
- Starling-7B-Beta still the best 7B with over 13K votes!

Total #models: 80. Total #votes: 599638. Last updated: April 9, 2024.

Contribute your vote at [chat.lmsys.org](https://chat.lmsys.org)! Find more analysis in the [notebook](#).

Rank	Model	Arena Elo	95% CI	Votes	Organization	License
1	<a href="#">Claude 3 Opus</a>	1256	+3/-4	44882	Anthropic	Proprietary
1	<a href="#">GPT-4-1106-preview</a>	1254	+3/-3	61868	OpenAI	Proprietary
1	<a href="#">GPT-4-0125-preview</a>	1250	+3/-4	45101	OpenAI	Proprietary
4	<a href="#">Bard (Gemini Pro)</a>	1208	+5/-5	12468	Google	Proprietary
4	<a href="#">Claude 3 Sonnet</a>	1204	+3/-3	54349	Anthropic	Proprietary
6	<a href="#">Command R+</a>	1194	+4/-6	13711	Cohere	CC-BY-NC-4.0
6	<a href="#">GPT-4-0314</a>	1188	+4/-4	40321	OpenAI	Proprietary
8	<a href="#">Claude 3 Haiku</a>	1182	+3/-4	45879	Anthropic	Proprietary
9	<a href="#">GPT-4-0613</a>	1163	+3/-3	59434	OpenAI	Proprietary
9	<a href="#">Mistral-Large-2402</a>	1158	+3/-4	32813	Mistral	Proprietary
10	<a href="#">Qwen1.5-72B-Chat</a>	1153	+4/-6	26314	Alibaba	Qianwen LICENSE
10	<a href="#">Claude-1</a>	1150	+5/-6	21868	Anthropic	Proprietary
11	<a href="#">Mistral Medium</a>	1148	+4/-4	29226	Mistral	Proprietary
11	<a href="#">Command R</a>	1148	+4/-4	27854	Cohere	CC-BY-NC-4.0
12	<a href="#">Qwen1.5-32B-Chat</a>	1140	+6/-5	10871	Alibaba	Qianwen LICENSE

11:30 AM · Apr 9, 2024 · 414.2K Views

Figure 3: Confidence Intervals on Model Strength (via Bootstrapping)

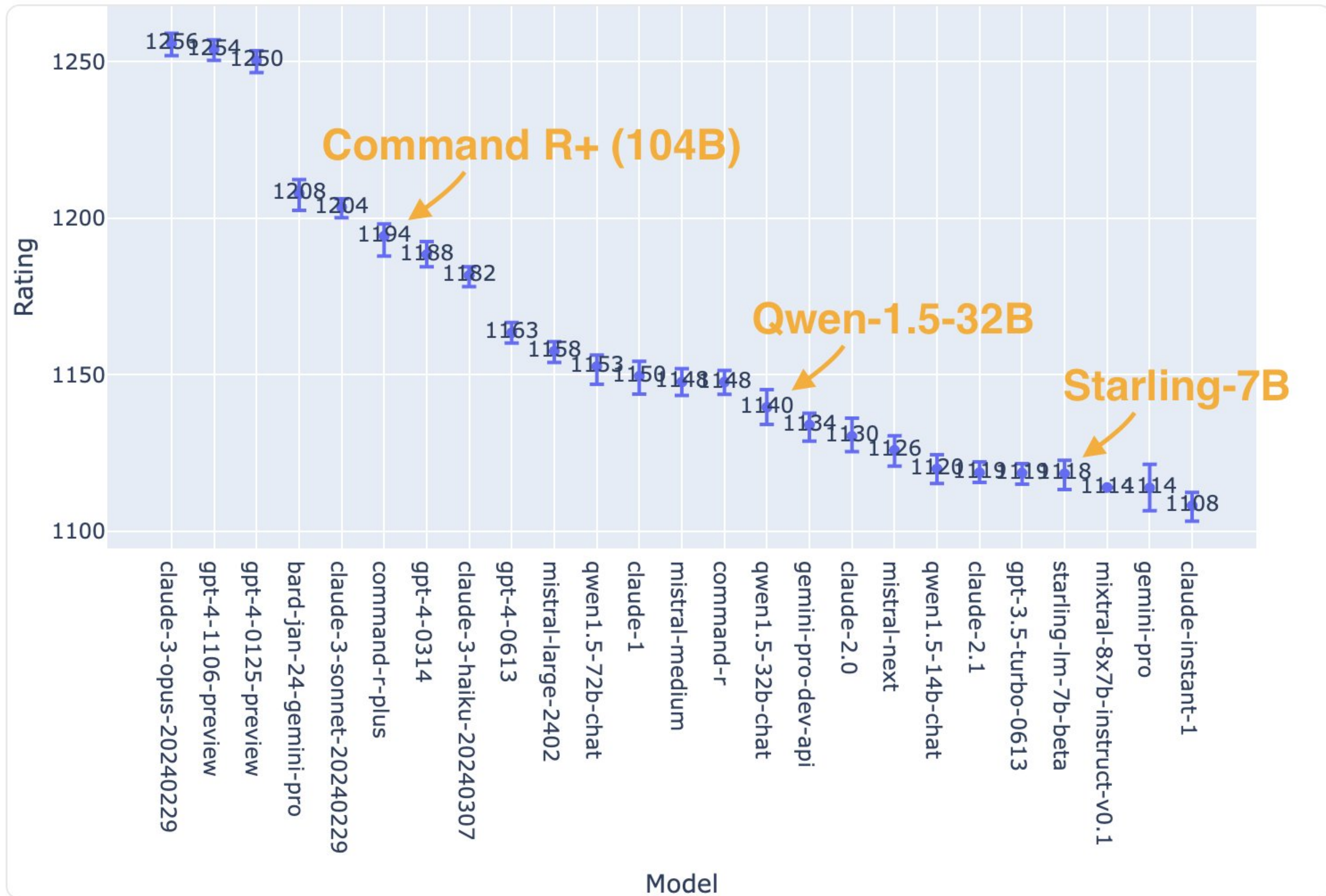


Figure 1: Fraction of Model A Wins for All Non-tied A vs. B Battles

