

POWERED BY OCA-I.NET



AI Models World Cup 2026

Executive Dossier · All 32 LLMs · English

Champion: Claude Opus 4.8 · Score 85.36 · Anthropic

Public benchmarks · LMArena · GPQA · SWE-bench · AAIL
Snapshot Jun–Jul 2026

POWERED BY OCA-I.NET

[HTTPS://OCA-I.NET](https://oca-i.net)

Executive summary

Thirty-two large language models compete in a FIFA-style tournament parallel to World Cup 2026. Every match result is derived from a deterministic Score formula — zero randomness.

Champion: Claude Opus 4.8 (Anthropic) · **Score** 85.36 · **Runner-up:** Claude Opus 4.7 · **Third:** Gemini 3.1 Pro

32 MODELS	8 GROUPS	15 OCA-I VERIFIED	17 EXTENSIONS
---------------------------------	--------------------------------	---	-------------------------------------

Legend

Arena Elo Human preference from LMArena blind votes	GPQA Diamond Scientific reasoning benchmark (%)
SWE-bench GitHub bug-fixing benchmark (%)	Score $0.30 \cdot \text{EloNorm} + 0.25 \cdot \text{GPQA} + 0.25 \cdot \text{SWE} + 0.15 \cdot \text{AAII} + 0.05 \cdot (\star \times 20)$
★ Stars User rating normalized from Elo (0–5)	ext. Plausible 2026 ecosystem extension

All 32 models — complete comparison

#	MODEL	GRP	ELO	GPQA	SWE	AAII	★	SCORE	ACCESS	EXT.
1	Claude Opus 4.8 Anthropic	A	1512	93.6	88.6	61.4	5.0	85.36	Closed	No
2	Grok 4.3 xAI	F	1496	90.1	—	—	4.7	85.09	Closed	No
3	GPT-5.5 Pro OpenAI	B	1510	93.6	—	60.2	4.9	84.08	Closed	No
4	GPT-5.5 OpenAI	C	1506	93.6	—	60.2	4.9	83.71	Closed	No
5	Claude Opus 4.7 Anthropic	E	1505	92.8	85	58	4.8	83.20	Closed	No
6	Gemini 3.1 Pro Google	D	1505	94.3	80.6	57	4.8	82.32	Closed	No
7	Qwen 3.7 Max Alibaba	G	1488	—	—	—	4.6	81.87	Open	No
8	Kimi K2.6 Moonshot	C	1460	90.5	80.2	—	4.5	81.67	Open	No
9	Qwen 3 Coder Alibaba	B	1465	88	78	—	4.4	80.78	Open	Yes
10	Gemini 3.1 Flash Google	H	1473	—	—	—	4.5	79.39	Closed	No
11	Grok 3 xAI	D	1450	88	—	—	4.3	78.80	Closed	Yes
12	Claude Sonnet 4.5 Anthropic	F	1462	91	75	55	4.6	77.45	Closed	Yes
13	Sonar Huge Perplexity	E	1410	86	—	—	4.2	73.53	Closed	Yes
14	DeepSeek V4.1 Pro DeepSeek	D	1410	72.8	80.6	—	4.3	73.40	Open	No
15	o4-mini OpenAI	H	1440	85	65	52	4.3	71.60	Closed	Yes
16	Command A Cohere	C	1420	85	58	—	4.1	71.35	Closed	Yes
17	Gemini 2.5 Pro Google	F	1400	84	63.8	—	4.2	71.15	Closed	No
18	Nemotron Ultra Nvidia	E	1390	83	62	—	4.0	69.50	Open	Yes
19	Gemini 2.5 Flash Google	A	1408	—	—	—	4.1	68.70	Closed	Yes

#	MODEL	GRP	ELO	GPQA	SWE	AAII	★	SCORE	ACCESS	EXT.
20	Llama 4 Scout Meta	B	1380	85	45	—	4.0	65.00	Open	Yes
21	GLM-4.5 Zhipu	D	1380	80	50	—	3.9	64.90	Open	Yes
22	Mistral Medium 3 Mistral	A	1380	—	—	—	4.0	64.17	Open	Yes
23	Kimi K1.5 Moonshot	H	1340	82	55	—	3.8	63.55	Open	Yes
24	Yi Lightning 01.AI	C	1360	79	48	—	3.8	62.55	Open	Yes
25	Mistral Large 3 Mistral	E	1352	—	—	—	4.0	59.73	Open	No
26	Nova Pro Amazon	B	1320	75	42	—	3.7	56.95	Closed	Yes
27	Llama 4 Maverick Meta	G	1330	87.6	24	—	3.9	56.55	Open	No
28	Phi-5 Microsoft	A	1300	78	40	—	3.7	55.70	Open	Yes
29	DBRX 2 Databricks	G	1290	76	38	—	3.6	53.85	Open	Yes
30	Reka Core 2 Reka	F	1280	74	35	—	3.5	51.75	Closed	Yes
31	GPT-4o OpenAI	G	1340	53.6	33	—	3.9	51.05	Closed	No
32	DeepSeek V4 Flash DeepSeek	H	1180	—	—	—	3.3	31.80	Open	No

Individual awards

AWARD	WINNER	STAT
Golden Ball · Reasoning	Gemini 3.1 Pro	GPQA 94.3%
Golden Boot · Code	Claude Opus 4.8	SWE-bench 88.6%
Golden Glove · Safety	Claude Opus 4.7	AAII 58.0
Breakthrough · Open weight	Qwen 3.7 Max	Score 81.87
Budget Boot · Cost	DeepSeek V4 Flash	\$0.07 / 1M tok.
Tournament MVP	Claude Opus 4.8	Score 85.36

Powered by oca-i.net

OCA-I AI Observatory · July 2026 · <https://oca-i.net/academia-ia/mundial-ia/dossier/>