

IQ Test Results

Mensa Norway IQ Scores (Average of last 7 tests)

Category	Mensa Norway
o4-mini-high	99
Llama-4	98
Gemini Pro (Vision)	96
GPT-5-Thinking	96
o3-vision	94
Claude-4-Sonnet-Vision	93
deepseek-v3	92
Gemini Thinking	90
Bing	86
Mistral	85
GPT4 Omni (Verbal)	85
Claude-4-Opus-Vision	82
Grok-4-Vision	82
o4-mini-vision	79
GPT-5-Vision	67
GPT4 Omni (Vision)	65
llama-vision	62

This site quizzes 18 Verbal & 12 Vision AIs every week | Last Updated: 04:24PM EDT on August 14, 2025

Reset

Show Offline Test

Show Mensa Norway

Vision Only

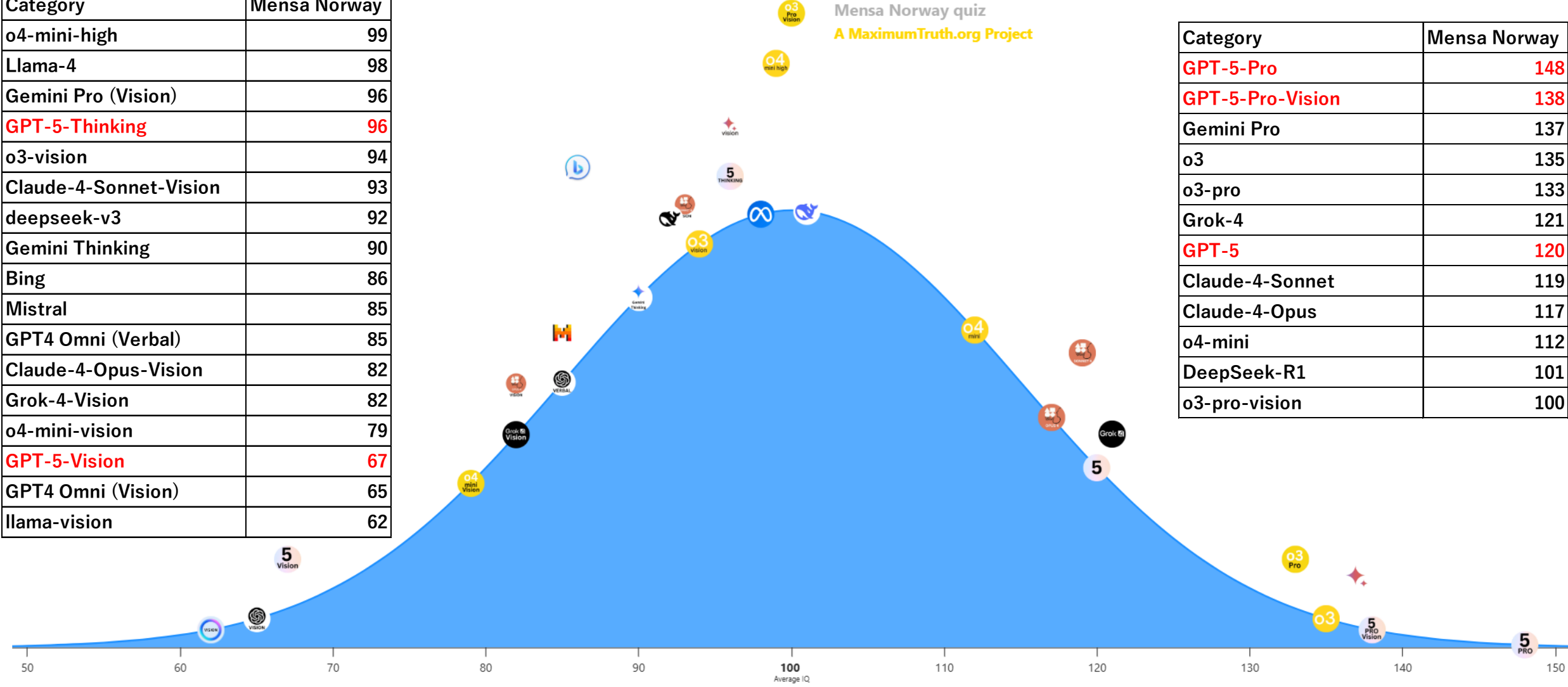
Vision Only

TrackingAI.org

Mensa Norway quiz

A MaximumTruth.org Project

Category	Mensa Norway
GPT-5-Pro	148
GPT-5-Pro-Vision	138
Gemini Pro	137
o3	135
o3-pro	133
Grok-4	121
GPT-5	120
Claude-4-Sonnet	119
Claude-4-Opus	117
o4-mini	112
DeepSeek-R1	101
o3-pro-vision	100



- Llama 4 Maverick (Vision)
- Gemini 2.5 Pro (Vision)
- Gemini 2.5 Flash Thinking.
- Gemini 2.5 Pro
- GPT-4o (Vision)
- GPT-4o
- Mistral
- Claude-4 Opus
- Claude-4 Sonnet
- Claude-4 Opus (Vision)
- Claude-4 Sonnet (Vision)
- DeepSeek V3
- DeepSeek R1
- OpenAI o3 (Vision)
- OpenAI o3
- Llama 4 Maverick
- OpenAI o4 mini (Vision)
- OpenAI o4 mini high
- OpenAI o4 mini
- OpenAI o3 Pro
- OpenAI o3 Pro (Vision)
- Grok-4 (Vision)
- Grok-4
- OpenAI GPT 5 Pro (Vision)
- OpenAI GPT 5 Thinking
- OpenAI GPT 5
- OpenAI GPT 5 Pro
- OpenAI GPT 5 (Vision)

Arena Overview

Scroll to the right to see full stats of each model ➔

(14 Aug '25)

First Place Second Place Third Place

Default

Compact View

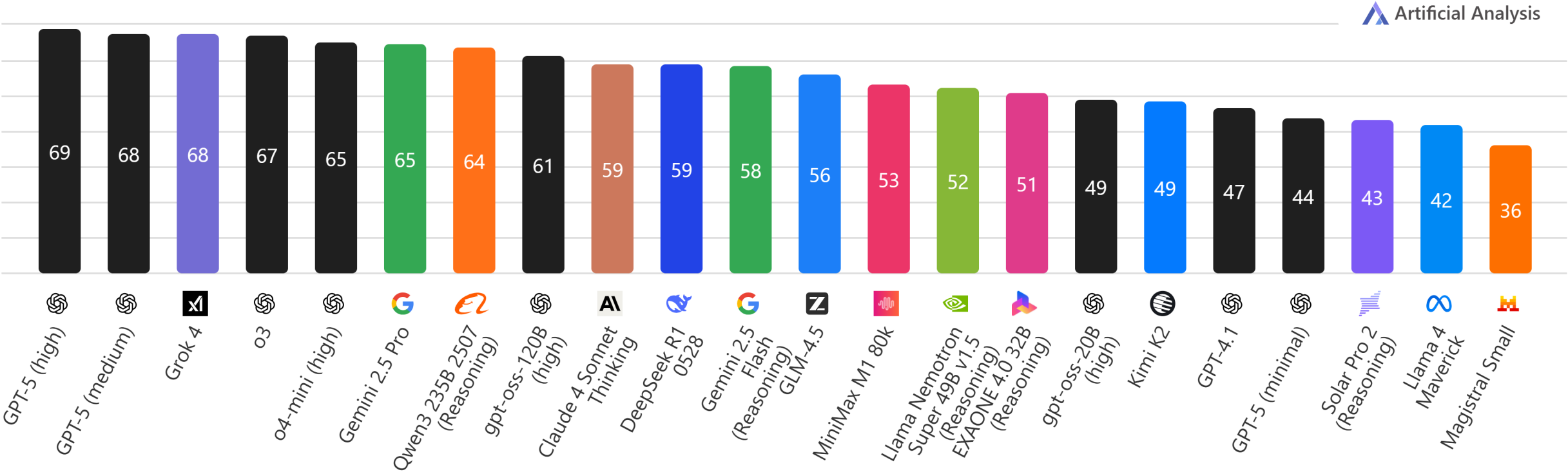
Model	226 / 226	Overall	Hard Prompts	Coding	Math	Creative Writing	Instruction Following	Longer Query	Multi-Turn
gpt-5-high		1	1	1	1	1	1	1	1
claude-opus-4-1-202...		2	1	1	3	1	1	1	1
gemini-2.5-pro		2	2	3	1	1	2	1	1
o3-2025-04-16		2	3	3	2	6	6	7	6
chatgpt-4o-latest-2...		4	4	3	11	2	4	2	1
gpt-4.5-preview-202...		4	5	4	6	2	3	2	1
grok-4-0709		5	6	8	1	2	4	5	6
qwen3-235b-a22b-ins...		5	2	1	1	4	2	1	5
claude-opus-4-20250...		7	4	2	5	2	2	1	6
kimi-k2-0711-preview		7	6	4	6	9	12	10	6
deepseek-r1-0528		8	6	6	6	8	12	10	11
glm-4.5		8	4	4	5	6	4	6	8
claude-opus-4-20250...		9	7	7	6	2	5	2	6
gemini-2.5-flash		11	14	20	5	4	9	7	13
grok-3-preview-02-24		11	8	13	18	6	8	6	11
gpt-4.1-2025-04-14		12	8	9	29	6	9	5	6
qwen3-235b-a22b-thi...		12	8	4	4	6	6	7	9
claude-sonnet-4-202...		15	8	3	6	8	5	3	6
o1-2024-12-17		16	14	17	6	13	11	12	20
o4-mini-2025-04-16		16	17	15	5	24	21	25	17
qwen3-235b-a22b-no-...		16	15	10	8	17	21	15	9
deepseek-r1		17	14	13	6	13	12	17	9

Artificial Analysis社の「長い文脈での考える力」のベンチマークで首位

Artificial Analysis Intelligence Index

Artificial Analysis Intelligence Index (14 Aug '25)

Artificial Analysis Intelligence Index v2.2 incorporates 8 evaluations: MMLU-Pro, GPQA Diamond, Humanity's Last Exam, LiveCodeBench, SciCode, AIME, IFBench, AA-LCR



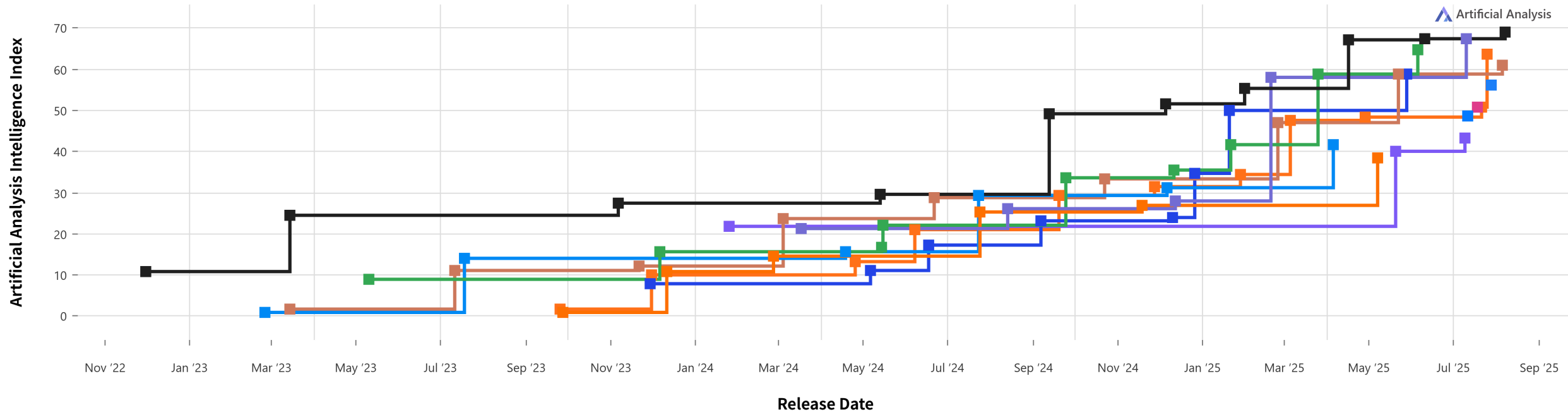
「AIインテリジェンス全般」でも29日ぶりにx.AIから首位を奪還

Frontier Language Model Intelligence, Over Time (14 Aug '25)

Frontier Language Model Intelligence, Over Time

Artificial Analysis Intelligence Index v2.2 incorporates 8 evaluations: MMLU-Pro, GPQA Diamond, Humanity's Last Exam, LiveCodeBench, SciCode, AIME, IFBench, AA-LCR

Alibaba Anthropic DeepSeek Google LG AI Research Meta Mistral Moonshot AI OpenAI Upstage xAI Z AI



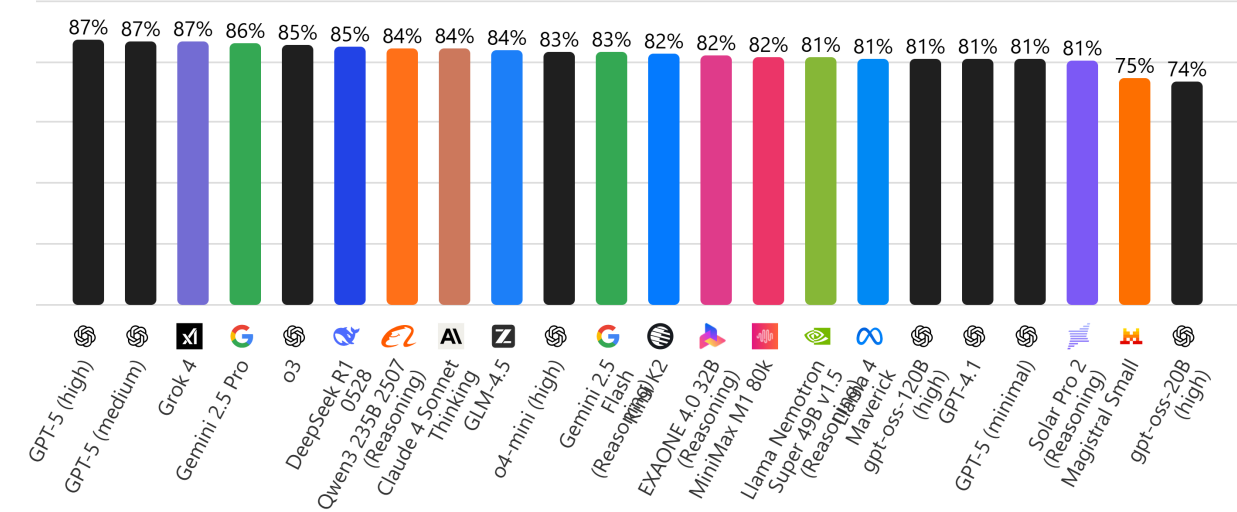
Intelligence Evaluations

Intelligence evaluations measured independently by Artificial Analysis; Higher is better

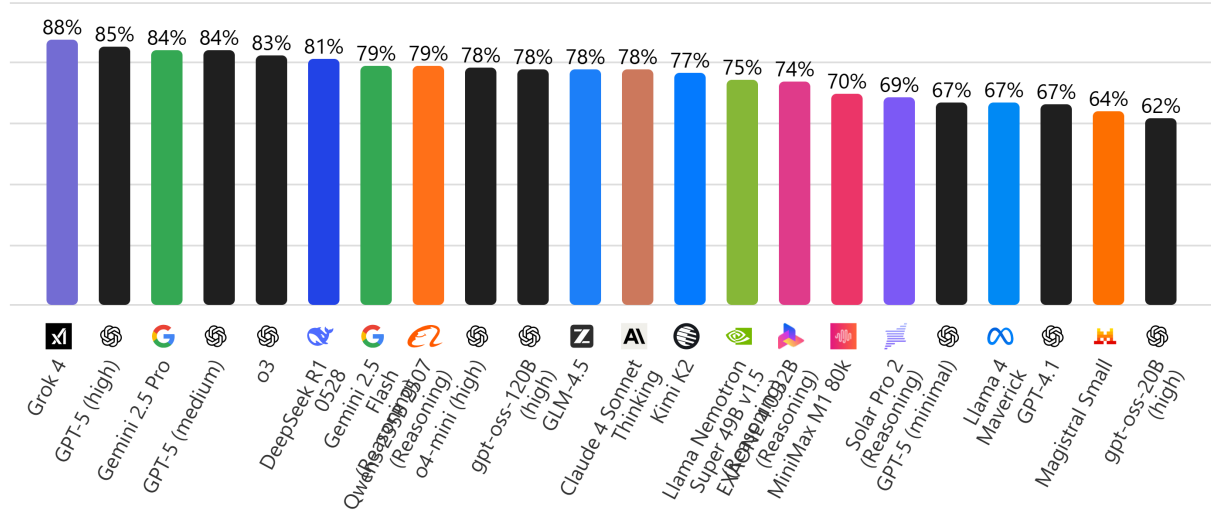
Results claimed by AI Lab (not yet independently verified)

(14 Aug '25)

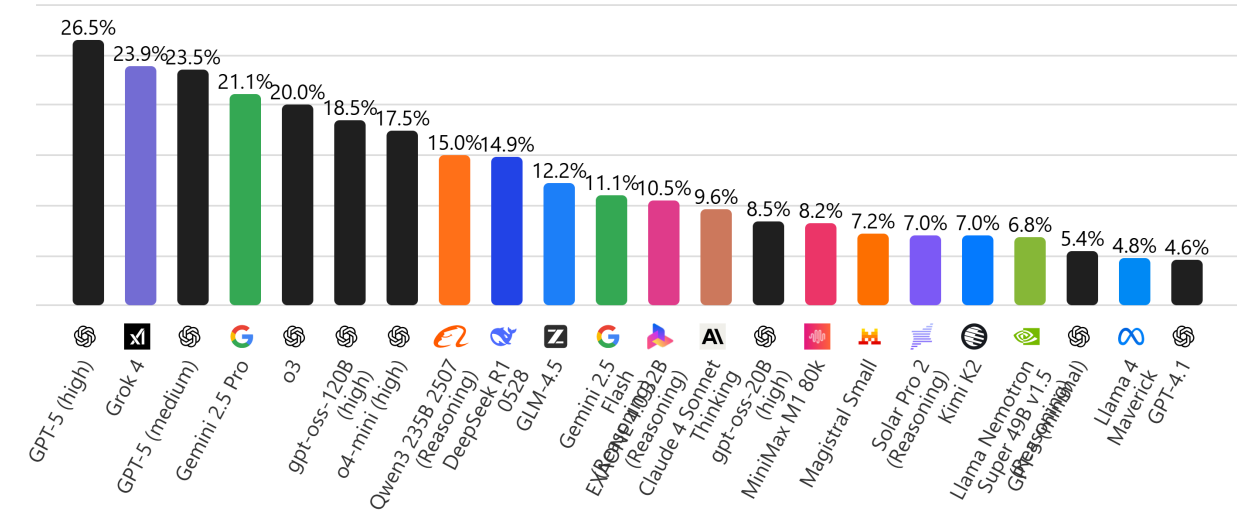
MMLU-Pro (Reasoning & Knowledge)



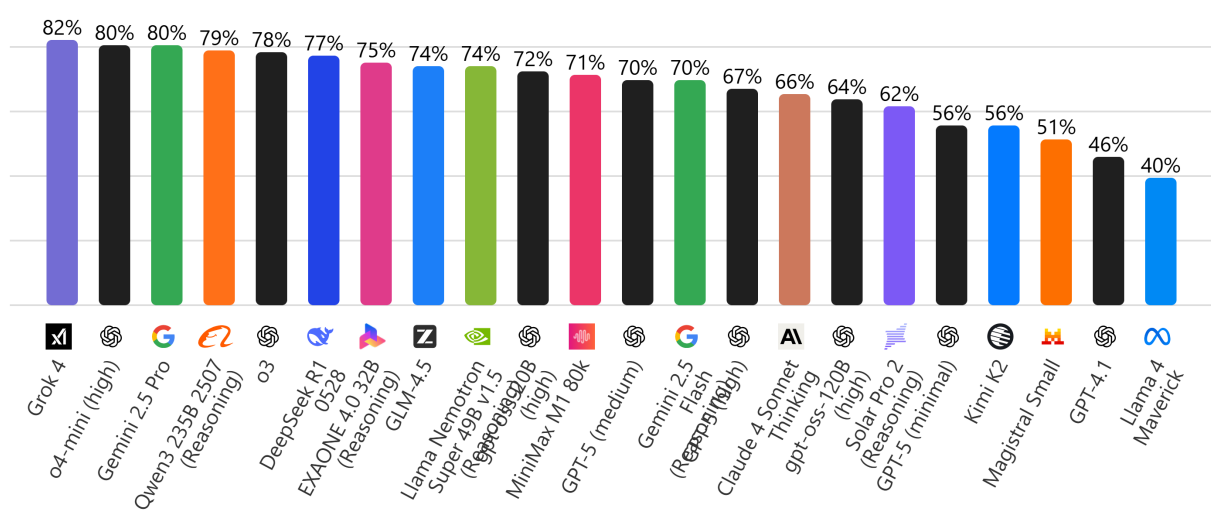
GPQA Diamond (Scientific Reasoning)



Humanity's Last Exam (Reasoning & Knowledge)

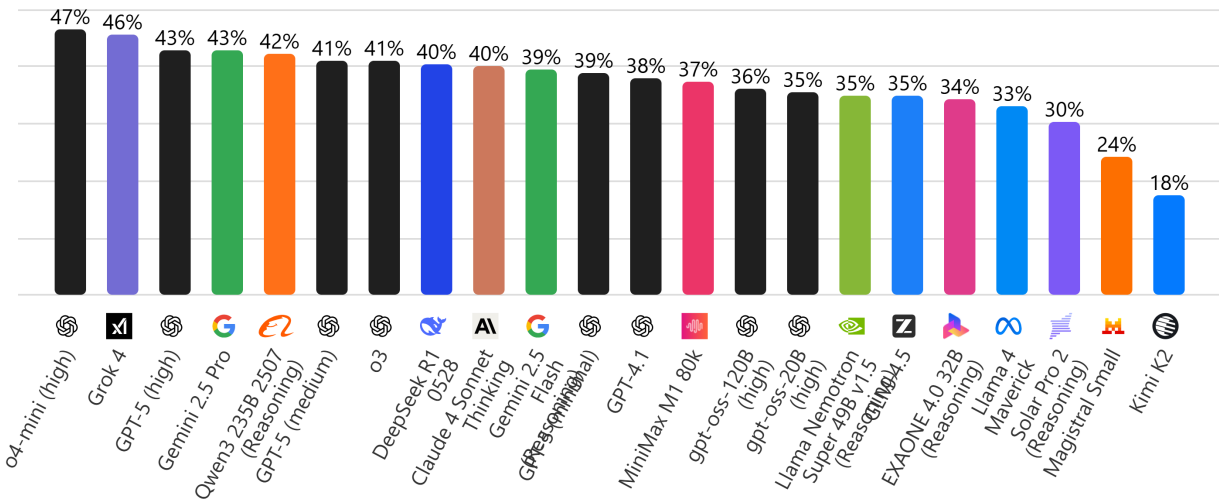


LiveCodeBench (Coding)

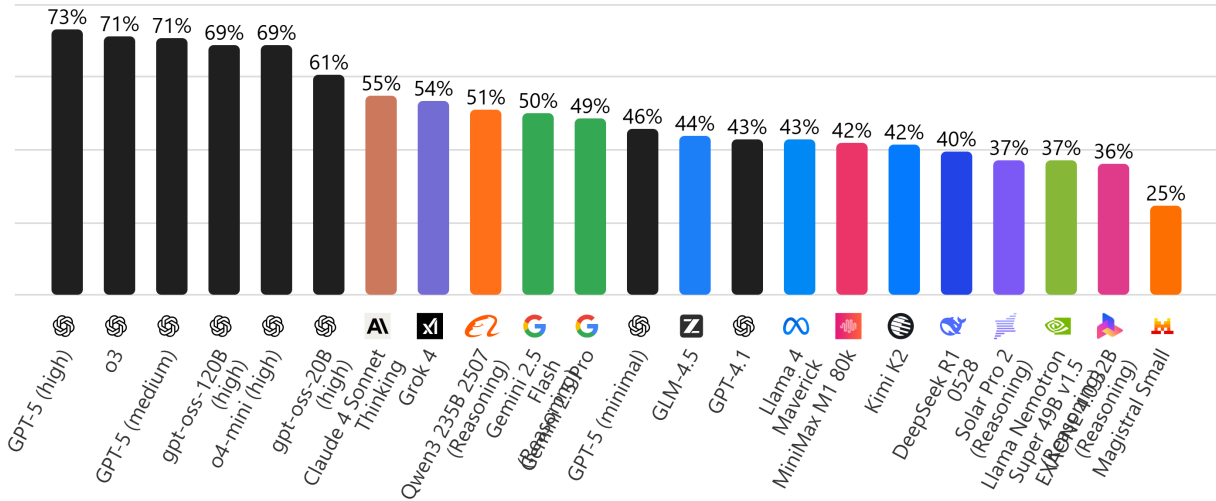


(14 Aug '25)

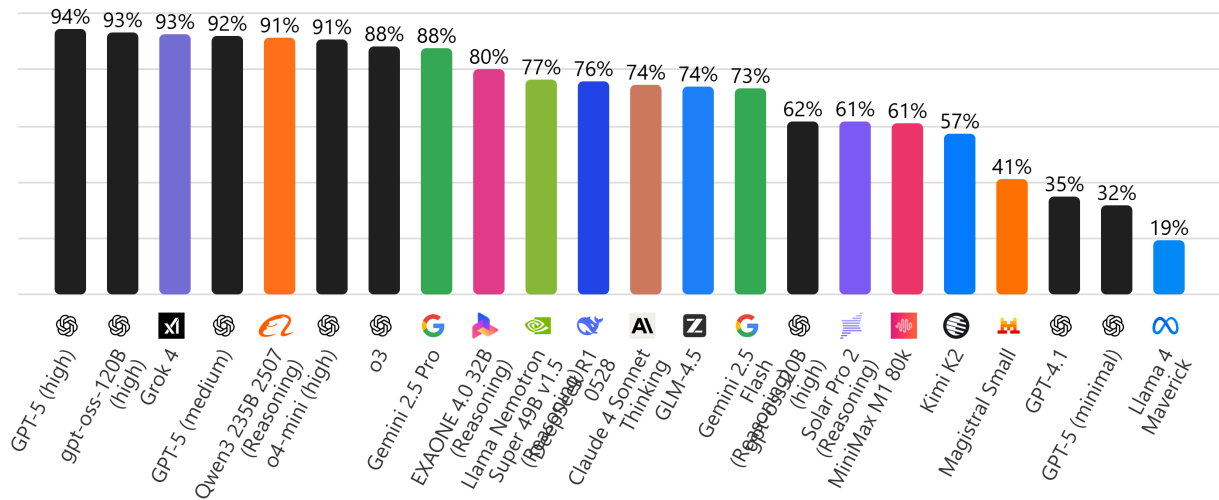
SciCode (Coding)



IFBench (Instruction Following)



AIME 2025 (Competition Math)



AA-LCR (Long Context Reasoning)

