

First Place

Second Place

Third Place

Default

Compact View

Model 224 / 224	Overall	Hard Prompts	Coding	Math	Creative Writing	Instruction Following	Longer Query	Multi-Turn
gpt-5	1	1	1	1	1	1	1	1
gemini-2.5-pro	2	2	2	1	1	2	1	1
o3-2025-04-16	2	2	2	2	5	4	6	5
chatgpt-4o-latest-2...	3	3	2	10	2	3	2	1
gpt-4.5-preview-202...	4	4	3	5	2	3	2	1
grok-4-0709	5	5	7	1	2	3	5	5
qwen3-235b-a22b-ins...	5	2	1	1	2	2	2	3
claude-opus-4-20250...	6	3	2	5	2	2	1	5
glm-4.5	6	3	3	5	5	3	6	6
kimi-k2-0711-preview	6	4	3	8	8	11	8	5
deepseek-r1-0528	7	4	3	6	6	11	8	8
claude-opus-4-20250...	8	6	6	6	2	4	1	5
gemini-2.5-flash	10	13	19	5	3	8	6	11
gpt-4.1-2025-04-14	10	7	8	25	5	6	4	5
grok-3-preview-02-24	10	7	12	16	5	6	5	9
qwen3-235b-a22b-thi...	11	7	5	3	5	5	8	8
claude-sonnet-4-202...	14	7	3	5	6	4	2	5
qwen3-235b-a22b-no-...	14	13	8	6	16	18	14	8
o1-2024-12-17	15	13	16	6	12	10	10	20
o4-mini-2025-04-16	15	16	14	5	23	20	25	17
deepseek-r1	16	13	12	6	12	11	16	8

IQ Test Results

Mensa Norway IQ Scores (Average of last 7 tests)

Category	Mensa Norway
o4-mini-high	99
Gemini Pro (Vision)	97
o3-vision	96
Claude-4-Sonnet-Vision	93
deepseek-v3	92
Gemini Thinking	89
GPT4 Omni (Verbal)	88
Bing	86
Mistral	85
Grok-4-Vision	82
o4-mini-vision	82
Claude-4-Opus-Vision	82
GPT-5-Thinking	79
GPT4 Omni (Vision)	65
GPT-5-Vision	64
llama-vision	62

TrackingAI.org

Mensa Norway quiz

A MaximumTruth.org Project

Reset

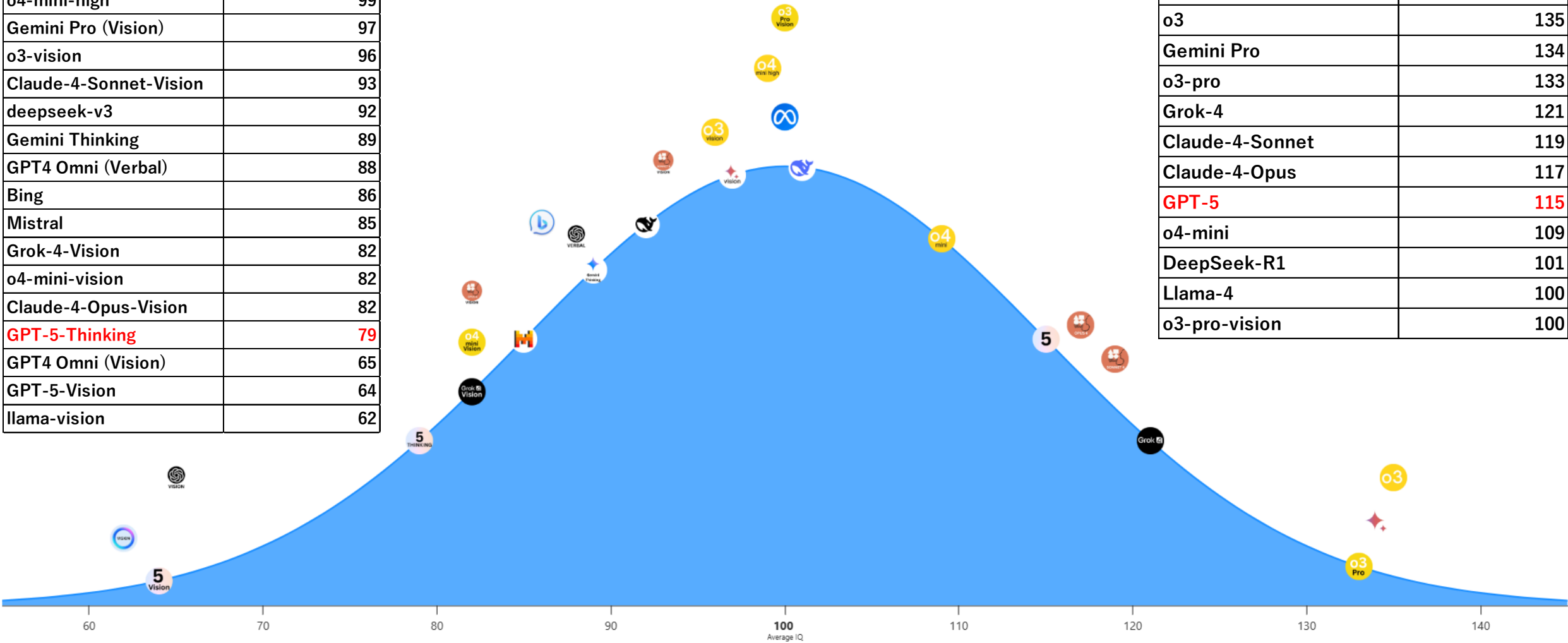
Show Offline Test

Show Mensa Norway

Vision Only

Vision Only

Category	Mensa Norway
o3	135
Gemini Pro	134
o3-pro	133
Grok-4	121
Claude-4-Sonnet	119
Claude-4-Opus	117
GPT-5	115
o4-mini	109
DeepSeek-R1	101
Llama-4	100
o3-pro-vision	100



Mistral



Llama 4 Maverick (Vision)



Gemini 2.5 Pro (Vision)



Gemini 2.5 Flash Thinking.



Gemini 2.5 Pro



GPT-4o (Vision)



GPT-4o



Claude-4 Opus



Claude-4 Opus (Vision)



Claude-4 Sonnet (Vision)



Claude-4 Sonnet (Vision)



Claude-4 Opus (Vision)



Claude-4 Opus (Vision)



Claude-4 Opus (Vision)



Claude-4 Opus (Vision)



Llama 4 Maverick



OpenAI o3 Pro



OpenAI o4 mini high



OpenAI o4 mini (Vision)



OpenAI o3 Pro (Vision)



OpenAI o3 (Vision)



OpenAI o3



OpenAI GPT 5 Thinking



OpenAI GPT 5 (Vision)



OpenAI o4 mini



OpenAI o4 mini (Vision)



OpenAI o3 Pro (Vision)



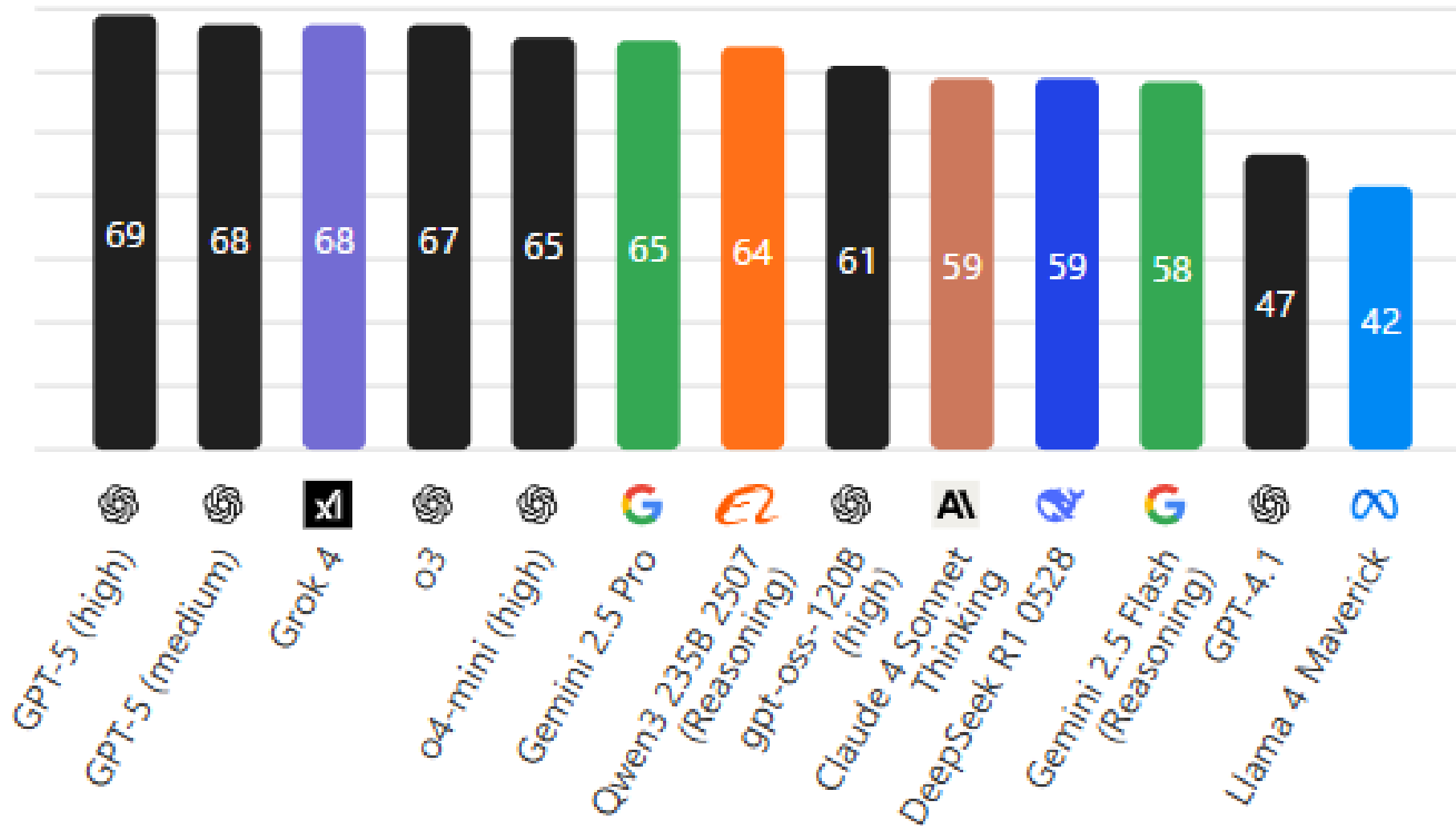
OpenAI o3 (Vision)



OpenAI o3

INTELLIGENCE

Artificial Analysis Intelligence Index; Higher is better



LEADERBOARD BREAKDOWN

AI System	Organization	System Type	ARC-AGI-1	ARC-AGI-2	Cost/Task	Code / Paper
Human Panel	Human	N/A	98.0%	100.0%	\$17.00	—
Grok 4 (Thinking)	xAI	CoT	66.7%	16.0%	\$2.17	
GPT-5 (High)	OpenAI	CoT	65.7%	9.9%	\$0.730	—
Claude Opus 4 (Thinking 16K)	Anthropic	CoT	35.7%	8.6%	\$1.93	
GPT-5 (Medium)	OpenAI	CoT	56.2%	7.5%	\$0.449	—
o3 (High)	OpenAI	CoT	60.8%	6.5%	\$0.834	—
o4-mini (High)	OpenAI	CoT	58.7%	6.1%	\$0.856	—
Claude Sonnet 4 (Thinking 16K)	Anthropic	CoT	40.0%	5.9%	\$0.486	
o3-Pro (High)	OpenAI	CoT + Synthesis	59.3%	4.9%	\$7.55	
Gemini 2.5 Pro (Thinking 32K)	Google	CoT	37.0%	4.9%	\$0.757	
Claude Opus 4 (Thinking 8K)	Anthropic	CoT	30.7%	4.5%	\$1.16	
GPT-5 Mini (High)	OpenAI	CoT	54.3%	4.4%	\$0.198	—
Gemini 2.5 Pro (Thinking 16K)	Google	CoT	41.0%	4.0%	\$0.715	

<https://arcprize.org/leaderboard>