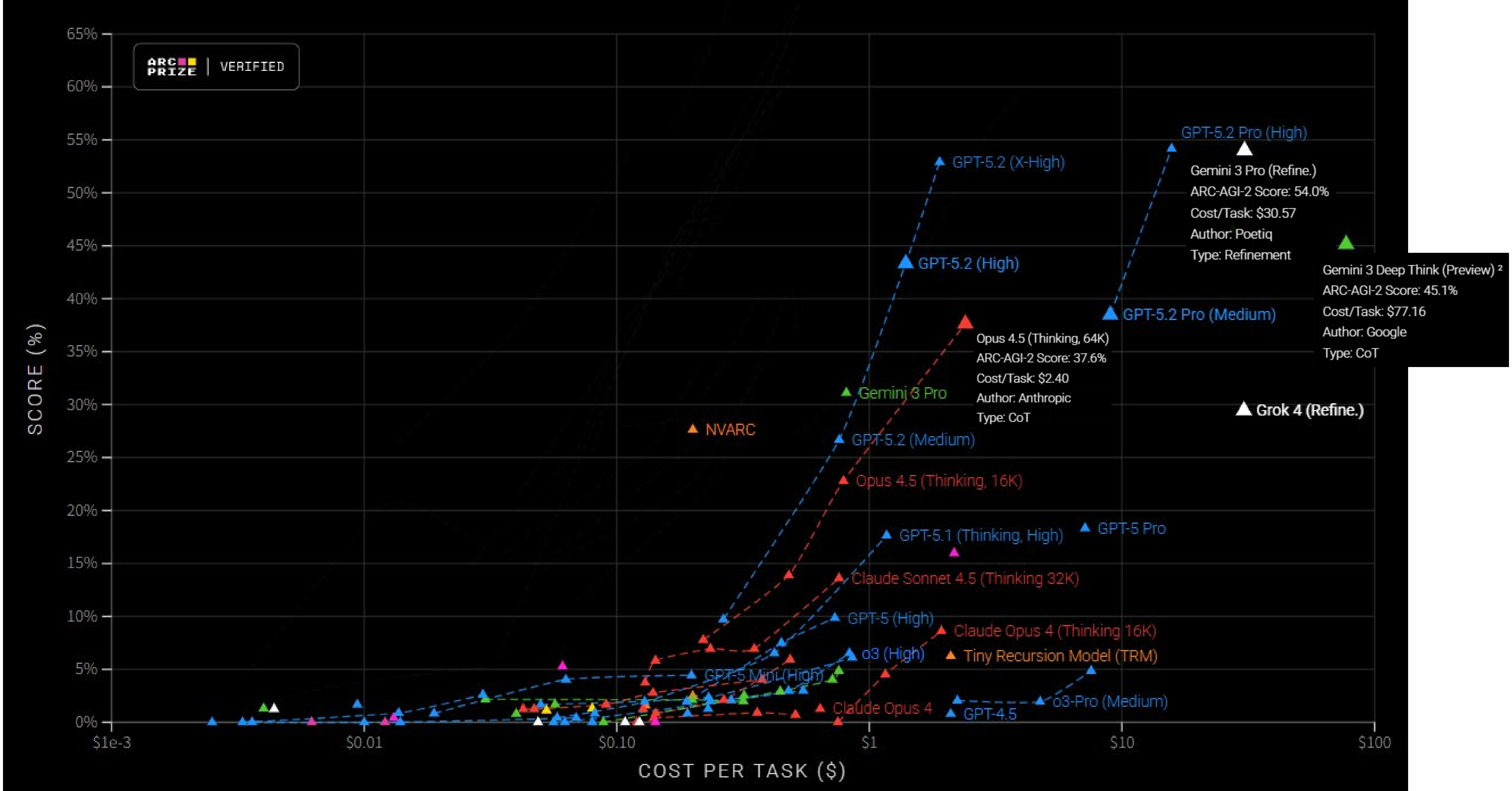
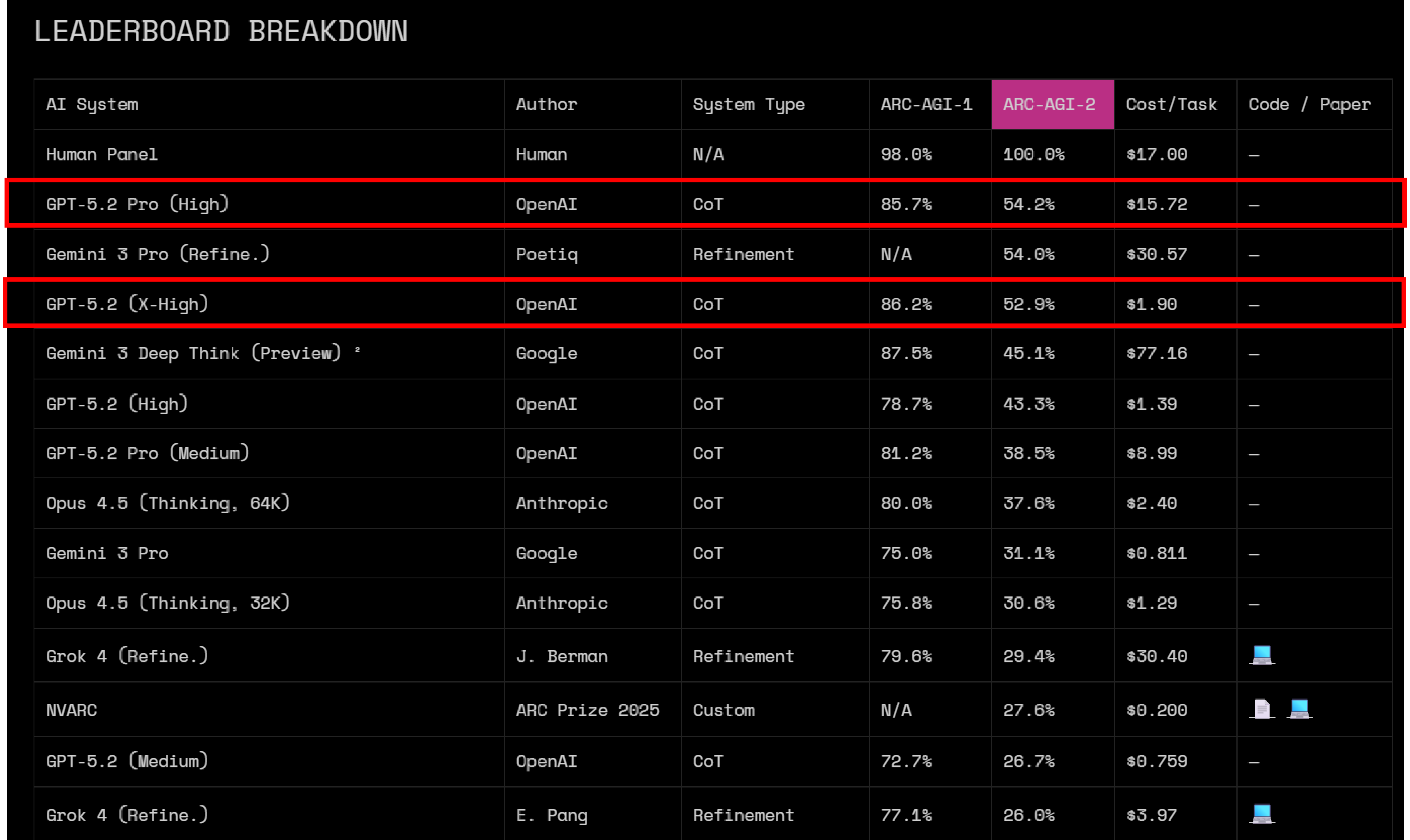
ARC-AGI-2 LEADERBOARD



Gemini 3 Pro (Refine.)
ARC-AGI-2 Score: 54.0%
Cost/Task: $30.57
Author: Poetiq
Type: Refinement

Gemini 3 Deep Think (Preview) ²
ARC-AGI-2 Score: 45.1%
Cost/Task: $77.16
Author: Google
Type: CoT

Opus 4.5 (Thinking, 64K)
ARC-AGI-2 Score: 37.6%
Cost/Task: $2.40
Author: Anthropic
Type: CoT

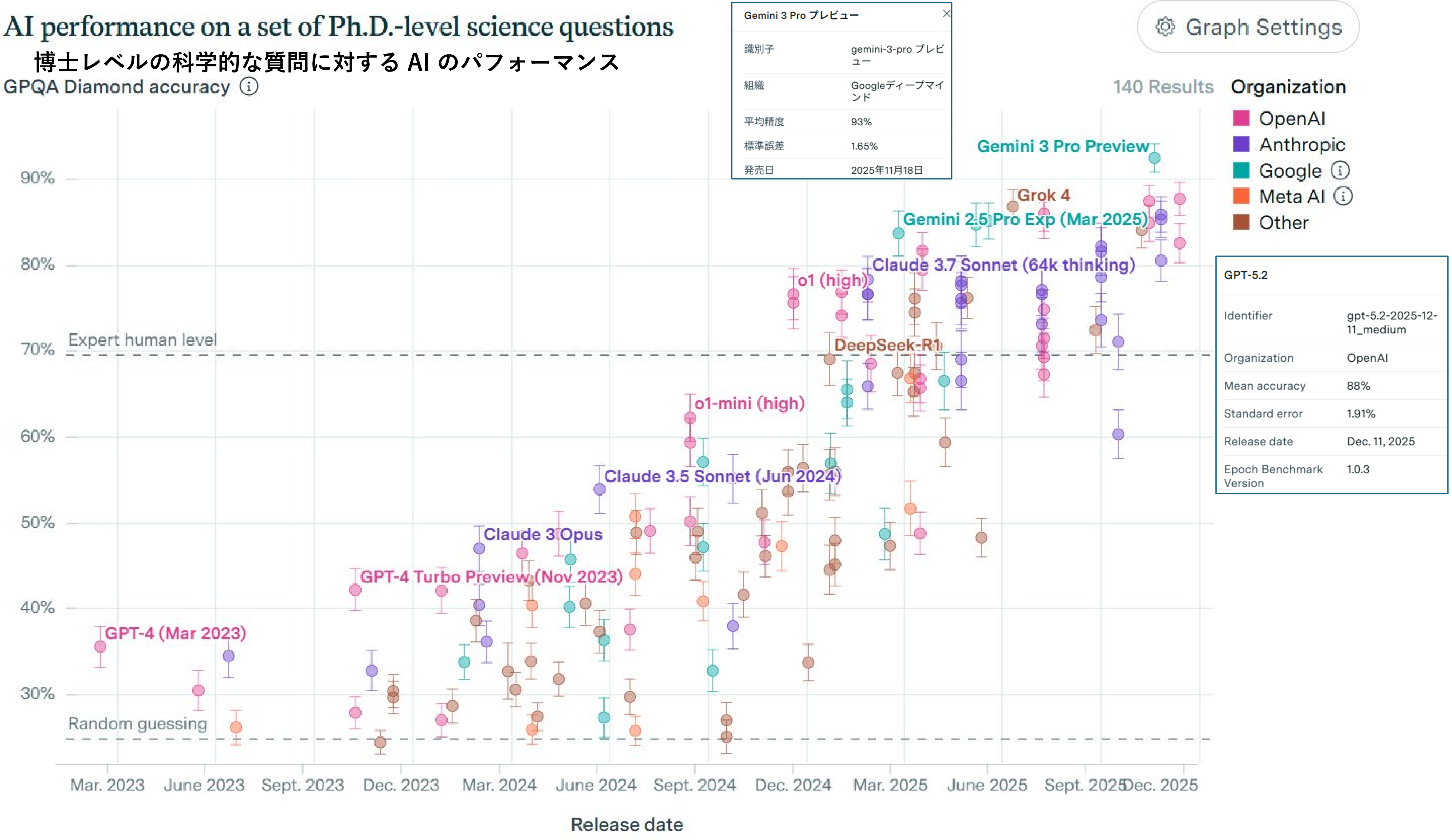


LEADERBOARD BREAKDOWN

AI System	Author	System Type	ARC-AGI-1	ARC-AGI-2	Cost/Task	Code / Paper
Human Panel	Human	N/A	98.0%	100.0%	$17.00	—
GPT-5.2 Pro (High)	OpenAI	CoT	85.7%	54.2%	$15.72	—
Gemini 3 Pro (Refine.)	Poetiq	Refinement	N/A	54.0%	$30.57	—
GPT-5.2 (X-High)	OpenAI	CoT	86.2%	52.9%	$1.90	—
Gemini 3 Deep Think (Preview) [2]	Google	CoT	87.5%	45.1%	$77.16	—
GPT-5.2 (High)	OpenAI	CoT	78.7%	43.3%	$1.39	—
GPT-5.2 Pro (Medium)	OpenAI	CoT	81.2%	38.5%	$8.99	—
Opus 4.5 (Thinking, 64K)	Anthropic	CoT	80.0%	37.6%	$2.40	—
Gemini 3 Pro	Google	CoT	75.0%	31.1%	$0.811	—
Opus 4.5 (Thinking, 32K)	Anthropic	CoT	75.8%	30.6%	$1.29	—
Grok 4 (Refine.)	J. Berman	Refinement	79.6%	29.4%	$30.40	💻
NVARC	ARC Prize 2025	Custom	N/A	27.6%	$0.200	📄 💻
GPT-5.2 (Medium)	OpenAI	CoT	72.7%	26.7%	$0.759	—
Grok 4 (Refine.)	E. Pang	Refinement	77.1%	26.0%	$3.97	💻

AI performance on a set of Ph.D.-level science questions

博士レベルの科学的な質問に対する AI のパフォーマンス



Gemini 3 Pro プレビュー

識別子	gemini-3-pro プレビュー
組織	Googleディープマインド
平均精度	93%
標準誤差	1.65%
発売日	2025年11月18日

GPT-5.2

Identifier	gpt-5.2-2025-12-11_medium
Organization	OpenAI
Mean accuracy	88%
Standard error	1.91%
Release date	Dec. 11, 2025
Epoch Benchmark Version	1.0.3