Model Evaluation – Approach, Methodology & Results

Gemini 3 Deep Think

Approach: Gemini 3 Deep Think was evaluated across a range of benchmarks, including reasoning, math, code, physics, quantum mechanics, and chemistry.

Methodology: All Gemini scores are pass @1 unless otherwise specified. All results are run with the default sampling settings unless indicated otherwise. To reduce variance, we average over multiple trials for smaller benchmarks.

All the results for non-Gemini models are sourced from providers' self-reported numbers unless mentioned otherwise below. For Claude Opus 4.6, and GPT-5.2 we default to reporting maximum thinking/reasoning settings available, but when reported results are not available we use best available reasoning results. Google DeepMind calculated the following scores using official provider APIs, as self-reported or official leaderboard numbers were unavailable: *International Math Olympiad 2025*, Codeforces, *International Physics Olympiad 2025 (theory), 2, International Chemistry Olympiad 2025 (theory)*. We excluded Claude Opus 4.6 results for International Chemistry and Math Olympiads 2025 due to empty model responses. Additionally, we omitted GPT-5.2 (xhigh) results for CodeForces due to persistent CloudFlare timeouts via the OpenAI API.

Additional Details: Our benchmarks span several capabilities as of February, 2026:

- **Reasoning and Academic Knowledge**
  - *ARC-AGI-2* results are sourced from the [ARC Prize website](#) and are ARC Prize Verified. The set reported is v2, semi-private.
  - *Humanity's Last Exam* results for Gemini 3 Pro Preview are from the ScaleAI [leaderboard](#). Gemini 3 Deep Think results are self-computed. For search and code on results we run the Gemini model using Gemini API with a blocklist implemented to avoid results that could include benchmark results, such as [huggingface.com](#).
  - *MMMU-Pro* scores are averaged across the Standard (10 options) and Vision settings.
- **Math** – *International Math Olympiad 2025* comprises all 6 problems from the official [2025 IMO competition](#). Results were self-computed and averaged across 4 runs. Answer correctness was evaluated with reference to canonical solutions using independent experts.
- **Code** – Codeforces uses all Codeforces contests from April to Oct 2025 (222 problems) where the problems are medium level and above (Elo rating > 1500 per https://clist.by/). Results were self-computed using a single attempt.
- **Physics**
  - *International Physics Olympiad 2025 (theory)* comprises all 3 theory problems from the official [2025 IPhO competition](#). Results were self-computed and averaged across 8 runs. Answer correctness was evaluated with reference to canonical solutions using Gemini as a judge. Gemini 3 Deep Think achieved gold-medal level performance as validated by independent subject-matter experts.

- **Quantum mechanics** – *CMT-Benchmark (Condensed Matter Theory)* results are self-computed for all models. Results are pass@8 for all models. Answer correctness was computed using Gemini as judge.
- **Chemistry**
  - *International Chemistry Olympiad 2025 (theory)* comprises all 9 theory problems from the official [2025 IChO competition](). Results were self-computed and averaged across 8 runs. Answer correctness was evaluated with reference to canonical solutions using Gemini as a judge. Gemini 3 Deep Think achieved gold-medal level performance as validated by independent subject-matter experts.

Results: Gemini 3 Deep Think results as of February, 2026 are below:

Benchmark		Gemini 3 Deep Think Feb 2026	Gemini 3 Pro Preview Thinking High	Claude Opus 4.6 Thinking Max	GPT-5.2 Thinking xhigh
ARC-AGI-2 Abstract reasoning puzzles	ARC prize verified	84.6%	31.1%	68.8%	52.9%
Humanity's Last Exam Academic reasoning (full set, text + MM)	No tools / Search + code execution	48.4% / 53.4%	37.5% / 45.8%	40.0% / 53.1%	34.5% / 45.5%
MMMU-Pro Multimodal understanding and reasoning	No tools	81.5%	81.0%	73.9%	79.5%
International Math Olympiad 2025 Mathematics		81.5%	14.3%	-	71.4%
Codeforces Coding and algorithms	No tools, Elo	3455	2512	2352	-
International Physics Olympiad 2025 (theory) Physics		87.7%	76.3%	71.6%	70.5%
CMT-Benchmark Condensed matter theory		50.5%	39.5%	17.1%	41.0%
International Chemistry Olympiad 2025, (theory) Chemistry		82.8%	69.6%	-	72.0%