

Google launches Gemini 3: surpasses GPT-5.1 on many important AI tests

Gemini 3 officially launched with stronger inference capabilities, leading many AI benchmarks and surpassing GPT-5.1 in key assessments. Deep Think mode is also greatly improved and ready for both consumers and businesses.

Google has just officially announced Gemini 3 – the AI model that has been awaited for a long time. After a series of 'teasers', Google DeepMind has officially unveiled the new generation AI platform, possessing the ability to reason and process multimedia that is described as the most outstanding ever.

As announced, Gemini 3 is currently leading the LMArena leaderboard with a new record Elo score of 1501. The model also scored 37.5% in Humanity's Last Exam, 91.9% in GPQA Diamond, and 23.4% in MathArena Apex, all of which are top performance benchmarks. Multimodality capabilities have also been significantly improved: Gemini 3 Pro scored 81% in MMMU-Pro and 87.6% in Video-MMMU. In the SimpleQA Verified accuracy test, the model scored 72.1%, topping the leaderboard.

Google also shared a full comparison chart between Gemini 3 Pro, GPT-5.1, and Claude Sonnet 4.5. In parallel, the company introduced Gemini 3 Deep Think mode, a version optimized for complex inference tasks. Gemini 3 Deep Think scored 41% on Humanity's Last Exam, 93.8% on GPQA Diamond, and 45.1% on ARC-AGI-2 (with code execution support, ARC Prize Verified version). Despite the huge performance increase, the entire Gemini 3 family maintains a context window of up to 1 million tokens.

Benchmark	Description		Gemini 3 Pro	Gemini 2.5 Pro	Claude Sonnet 4.5	GPT-5.1
Humanity's Last Exam	Academic reasoning	No tools	37.5%	21.6%	13.7%	26.5%
		With search and code execution	45.8%	—	—	—
ARC-AGI-2	Visual reasoning puzzles	ARC Prize Verified	31.1%	4.9%	13.6%	17.6%
GPQA Diamond	Scientific knowledge	No tools	91.9%	86.4%	83.4%	88.1%
AIME 2025	Mathematics	No tools	95.0%	88.0%	87.0%	94.0%
		With code execution	100%	—	100%	—
MathArena Apex	Challenging Math Contest problems		23.4%	0.5%	1.6%	1.0%
MMMU-Pro	Multimodal understanding and reasoning		81.0%	68.0%	68.0%	76.0%
ScreenSpot-Pro	Screen understanding		72.7%	11.4%	36.2%	3.5%
CharXiv Reasoning	Information synthesis from complex charts		81.4%	69.6%	68.5%	69.5%
OmniDocBench 1.5	OCR	Overall Edit Distance, lower is better	0.115	0.145	0.145	0.147
Video-MMMU	Knowledge acquisition from videos		87.6%	83.6%	77.8%	80.4%
LiveCodeBench Pro	Competitive coding problems from Codeforces, ICPC, and IOI	Elo Rating, higher is better	2,439	1,775	1,418	2,243
Terminal-Bench 2.0	Agentic terminal coding	Terminus-2 agent	54.2%	32.6%	42.8%	47.6%
SWE-Bench Verified	Agentic coding	Single attempt	76.2%	59.6%	77.2%	76.3%
t2-bench	Agentic tool use		85.4%	54.9%	84.7%	80.2%
Vending-Bench 2	Long-horizon agentic tasks	Net worth (mean), higher is better	\$5,478.16	\$573.64	\$3,838.74	\$1,473.43
FACTS Benchmark Suite	Held out internal grounding, parametric, MM, and search retrieval benchmarks		70.5%	63.4%	50.4%	50.8%
SimpleQA Verified	Parametric knowledge		72.1%	54.5%	29.3%	34.9%
MMMLU	Multilingual Q&A		91.8%	89.5%	89.1%	91.0%
Global PIQA	Commonsense reasoning across 100 Languages and Cultures		93.4%	91.5%	90.1%	90.9%
MRCR v2 (8-needle)	Long context performance	128k (average)	77.0%	58.0%	47.1%	61.6%
		1M (pointwise)	26.3%	16.4%	not supported	not supported

Unlike previous generations that took months to reach users, this time Google chose to launch the product faster and more widely. AI Mode in Google Search used Gemini 3 to bring a content-generated search interface, including a more intuitive layout, interactive simulations, and dynamic tools that appear immediately according to user queries.

On the SWE-bench Verified test for programming ability, Gemini 3 Pro scored 76.2% – slightly lower than GPT-5.1 and Sonnet 4.5. However, Google says Gemini 3 is available to developers through Google AI Studio, Vertex AI, Gemini CLI, Cursor, GitHub, JetBrains, Manus, Replit, and the company's new agentic development platform, Google Antigravity.

For general users, Gemini 3 is available in the Gemini app. Google AI Pro and Ultra subscribers can experience Gemini 3 in AI Mode on Google Search. For businesses, the model is available through Vertex AI and Gemini Enterprise. Gemini 3 Deep Think will be available to Google AI Ultra customers in the coming weeks.

You finished reading the article "**Google launches Gemini 3: surpasses GPT-5.1 on many important AI tests**" edited by the [TipsMake](#) team. We hope this article has provided you with many useful tech tips and tricks. You can search for similar articles on tips and guides. Thank you for reading and for following us regularly.