

Google выпустила обновлённую версию своей мощной модели искусственного интеллекта — Gemini 2.5 Pro. Новая версия исправляет ошибки прошлых обновлений и обещает более «стабильный» опыт использования.

Ранее версия Gemini 2.5 Pro (I/O Edition) и так была ориентирована на улучшение программирования, но теперь модель показывает ещё лучше результаты — 82,2% в тесте Aider Polyglot, опережая конкурентов вроде OpenAI и Anthropic.

Однако за пределами задач по коду модель ранее работала хуже. Google учла критику и улучшила креативность, оформление ответов (например, заголовки, списки и выделение текста), и в целом сделала модель более понятной и полезной для пользователей.

Benchmark		Gemini 2.5 Pro Preview 06-05 Thinking	OpenAI o3	OpenAI o4-mini	Claude Opus 4	Grok 3 Beta	DeepSeek R1
Input price	\$/1M tokens no caching	\$1.25 \$2.50 > 200k tokens	\$10.00	\$1.10	\$15.00	\$3.00	\$0.55
Output price	\$/1M tokens	\$10.00 \$15.00 > 200k tokens	\$40.00	\$4.40	\$75.00	\$15.00	\$2.19
Reasoning & knowledge Humanity's Last Exam (no tools)		21.6%	20.3%	14.3%	10.7%	—	14.0%*
Science GPQA diamond	single attempt multiple attempts	86.4%	83.3%	81.4%	79.6% 83.3%	80.2% 84.6%	81.0% —
Mathematics AIME 2025	single attempt multiple attempts	88.0%	88.9%	92.7%	75.5% 90.0%	77.3% 93.3%	87.5% —
Code generation LiveCodeBench (UI: 1/1/2025-5/1/2025)	single attempt	69.0%	72.0%	75.8%	51.1%	—	70.5%
Code editing Aider Polyglot		82.2% diff-fenced	79.6% diff	72.0% diff	72.0% diff	53.3% diff	71.6%
Agentic coding SWE-bench Verified	single attempt multiple attempts	59.6% 67.2%	69.1% —	68.1% —	72.5% 79.4%	— —	— 57.6%
Factuality SimpleQA		54.0%	48.6%	19.3%	—	43.6%	27.8%
Factuality FACTS Grounding		87.8%	69.6%	62.1%	77.7%	74.8%	—
Visual reasoning MMMU	single attempt multiple attempts	82.0% —	82.9% —	81.6% —	76.5% —	76.0% 78.0%	no MM support no MM support

Google

Также появилась новая функция для разработчиков — возможность настраивать «бюджет мышления», то есть управлять тем, сколько ресурсов тратит модель на ответ.

Новая версия уже доступна в Vertex AI и AI Studio, а также в приложении Gemini (пока в режиме предварительного просмотра). Полный релиз ожидается в ближайшее время.