

Gemini

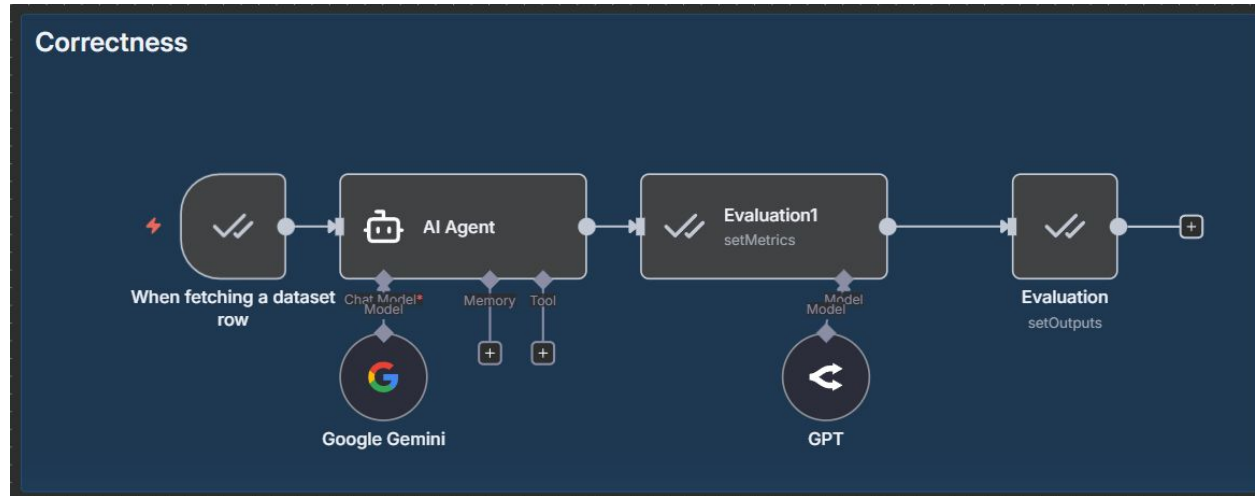
x



n8n

Today's Experiment

1. Connecting to a Gemini 3 Pro agent in n8n
2. Image understanding
3. Context window reasoning
4. Building n8n workflows



Gemini 3 Pro Release Overview

- **Release Date:** November 18, 2025 (Preview stage as of now)
- **Availability:** The preview is live in the Gemini app, Search, and through Google AI developer platforms. Wider features will launch soon.
- **What is Gemini 3?** Gemini 3 Pro is Google's newest and smartest AI model. It's designed to understand and create content from text, images, audio, and code, making it the most advanced AI from Google so far.
- **Value:** It aims to handle tougher questions, longer and more complex tasks, and work smoothly across all types of media, not just text.



Gemini Model Family Overview

Model	Context Window	Pricing (per 1M tokens)	What It's For / Description
Gemini 3 Pro Preview	1M in, 64k out	\$2 in / \$12 out (<200k) \$4 in / \$18 out (>200k)	Most advanced, top reasoning, true multimodal
Gemini 2.5 Pro	1M in, 65k out	\$1.25 in / \$10 out (<200k) \$2.50 in / \$15 out (>200k)	Best for complex coding, reasoning, docs
Gemini 2.5 Flash	1M in, 65k out	\$0.30 in / \$2.50 out	Hybrid reasoning, cheap, large-scale tasks
Gemini 2.5 Flash-Lite	1M in, 65k out	\$0.10 in / \$0.40 out	Smallest, most cost-effective, fast



Gemini 3 Power

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%

Gemini 3 Power

Benchmark	Description		Gemini 3 Pro	Gemini 2.5 Pro	Claude Sonnet 4.5	GPT-5.1
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%
τ2-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) 1M (pointwise)	77.0% 26.3%	58.0% 16.4%	47.1% not supported	61.6% not supported

For details on our evaluation methodology please see deepmind.google/models/evals-methodology/gemini-3-pro