Google's Gemini 3 Shows Visual-Reasoning Edge


Mandeep Singh
Team: Technology
BI Senior Industry Analyst


Robert Biggar
Team: Technology
BI Associate Analyst

Google Gemini 3 to Aid Image, Video Sharing; Cloud Unit Growth

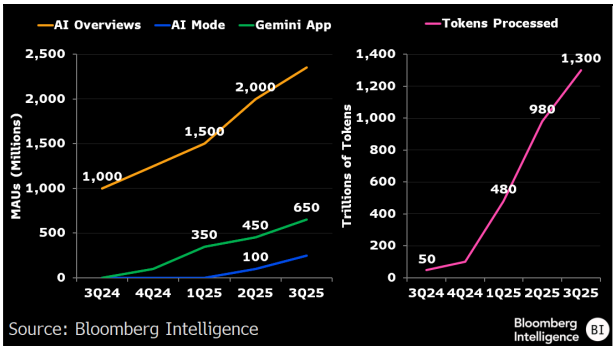
(Bloomberg Intelligence) -- Google's Gemini 3 visual-reasoning lead -- akin to Anthropic's coding-agent edge and OpenAI's in text chatbots -- could aid image- and video-workload adoption, while its widespread rollout suggests confidence in its ability to curb hallucinations. Google's training and inferencing scale with its own TPU chips may help it rent more of its Nvidia allocation for external workloads on Google Cloud, which could see segment-sales growth closer to 40%. (11/20/25)

1. Bigger Upfront Rollout for Gemini 3

Google's Gemini 3 rollout across its Search, Vertex AI, APIs and stand-alone Gemini app show the company's confidence in its ability to reduce hallucinations for large-language-model use across different types of tasks. Gemini 3 is likely optimized to serve traffic for more than a billion users across its family of apps, which we believe to be a likely source of differentiation vs. other frontier LLMs that have a significantly higher cost of revenue to run their LLM workloads.

The cost per token for inferencing at Alphabet's Google gives it an advantage, driven by its lower TPU expense vs. Nvidia's GPU, which could allow Google to serve its Gemini LLM in AI mode without a significant degradation in its search-business gross margin. (11/20/25)

Alphabet AI Products



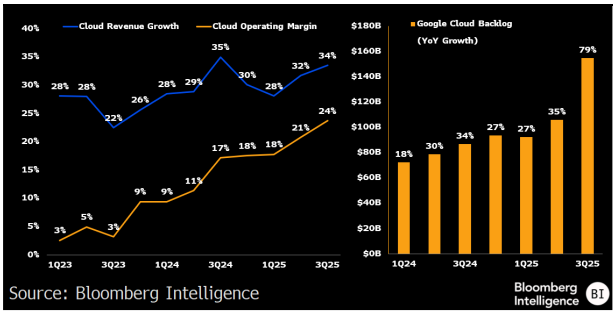
2. TPU Training Shows Commercialization Potential

Gemini 3 Pro was trained using Google's own TPUs, which further lowers the company's reliance on Nvidia GPUs, though TPU is used primarily for external workloads in its Google Cloud segment. With Anthropic looking to spend as much as $50 billion in the buildout of AI infrastructure for its LLM training and inferencing workloads, the commercialization of TPUs could be a source of top-line growth in areas other than the company's Google Cloud segment, where sales growth has improved to the mid-30%, aided by AI-workload contribution.

Google Cloud growth has accelerated to around the mid-30%, which could be further aided by the company's ability to rent its Nvidia GPU allocation for external workloads. Google Cloud margin

has also been expanded to over 20% year to date, aided by its use of TPUs for internal workloads. (11/20/25)

Google Cloud Growth, Margins, Backlog



3. Visual-Reasoning Differentiation vs. Peers

In addition to chatbot search and coding agents, the Gemini 3 model release shows a big leap in visual reasoning, which could be a big driver of the company's model vs. other frontier LLMs. Video- and image- generation capabilities have likely improved in Gemini 3, given the company's big step-up in visual reasoning compared with its 2.5 version. We believe that Google has added more human- preference data, using reinforcement learning, and fine-tuned its Gemini 3 model on proprietary- search index data and YouTube content. (11/20/25)

Gemini 3 Model Benchmarks

Benchmark	Description		Gemini 3 Pro	Gemini 2.5 Pro	Claude Sonnet 4.5	GPT-5.1
Humanity's Last Exam	Academic reasoning	No tools With search and code execution	**37.5%** **45.8%**	21.6% —	13.7% —	26.5% —
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 With code execution	**95.0%** **100%**	88.0% —	87.0% **100%**	94.0% —
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) 1M (pointwise)	**77.0%** **26.3%**	58.0% 16.4%	47.1% not supported	61.6% not supported

Source: Google Blog — Bloomberg Intelligence

4. Pricing, Reasoning vs. OpenAI

Google has a more efficient infrastructure based on a PUE metric, which shows the company requires less utility power per IT load. Its more than 1,300 trillion tokens/month, at least 2x higher than hyperscale peers, reflects its token-cost advantage vs. rivals such as OpenAI, which has a much lower product gross margin. We expect Google to use lower pricing than other LLM and cloud hyperscale companies given its advantage on inferencing costs in using its own TPU chips for compute. (11/20/25)

API Cost Comparison

Company	Model	Input Price ($/1M Tokens)	Output Price ($/1M Tokens)
Anthropic	Claude Sonnet 4.5	$3.00	$15.00
	Claude Opus 4.1	$15.00	$75.00
OpenAI	GPT-5.1	$1.25	$10.00
	o3	$2.00	$8.00
Alphabet	Gemini 3 Pro	$2.00	$12.00
xAI	Grok-4	$3.00	$15.00
Meta	Llama 4 Maverick	$0.20	$0.80

Source: Bloomberg Intelligence

Bloomberg Intelligence

To contact the analyst for this research:

Mandeep Singh at msingh15@bloomberg.net