09 Oct 2026 05:00:00 ET | 30 pages

Artificial Intelligence

Citi's Inference Ahead - Models Race Ahead While Infrastructure Catches Up

Heath Terry[AC]
Shelby Spencer
Ashley Kim
Janna Withrow

CITI'S TAKE

AI capabilities are advancing, but generating returns requires infrastructure and enterprises to keep pace. A central theme at infra/STRUCTURE Summit was that maintaining DC project timelines increasingly depends on working with local communities rather than simply building around them, making responsible development a key enabler of long-term capacity expansion. At AlphaSummit, AlphaSense said AI performs 98% of platform research, although Lanai CEO Lexi Reese estimated only about 2% of companies can directly tie AI usage to revenue or profit. With the intelligence ceiling rising 9% MoM (Figure 7) and total model releases over a trailing 90-day window increasing 38% MoM for proprietary models versus 7% MoM for open models (Figure 6), we believe the next test will be whether power delivery to DC sites can improve as quickly as the models.

Models. An unreleased model reportedly solved 90 of the top 500 open problems in math (Latent Space, 10/7). The scale of this progress indicates the public frontier continues to understate capability inside labs, elevating safety validation and controlled release as near-term bottlenecks. TasteVal (Figure 2) could form the basis of an RSI tracker after finding that frontier models' research efficiency has doubled every 3 months, signaling potential for faster development cycles. Reflection AI reports that its latest open-weight model, Beam, achieves GLM-5.2-level results with 3-4x less compute (10/5), reinforcing a Western push for efficiency that Google may continue with a frontier-tier model release within the next month (Figure 5). Mistral trained Large 4 on 3.8K Blackwells (10/6), illustrating that advances in efficiency are occurring alongside continued investment in larger training clusters.

Agents. Hark Pro (10/7) and Underdog (TechCrunch, 10/6) further mark a shift towards always-on assistants, though the next adoption test is becoming whether agents earn permission to act instead of whether users try another assistant. In South Korea, authorities are investigating breaches across several banks after signs that AI was used (Reuters, 10/5). The incidents could accelerate public concern and regional regulations around AI, making South Korea a case study for controls that may spread as agent autonomy rises. Vercel's closed-model share doubled from 23% to 46% over the past 2 weeks, narrowing open models' lead to 54% (Figure 9). Proprietary models may attract unrouted traffic while the intelligence gap persists, but open models' cost and control advantages could draw routed workloads.

Capacity. Local support will continue to shape where capital for DC buildouts is deployed over the next month, with Oklo at infra/STRUCTURE Summit citing restrictions across 400+ communities in 43 states and Amazon committing $1B over 5 years (10/6) to DC communities. Enterprises are buying 5-40MW blocks in sub-year cycles, according to First Citizens, despite turnkey shell costs rising from $8-10M per MW in 2018-20 to $12-14M today, with power shells as high as ~$5M depending on land costs. Meanwhile, GPU demand continues to strengthen, with H100, B200, and B300 rental prices rising 7.6%, 2.7%, and 17.5% WoW (Figure 20). Next week, OCP and SEMICON West should clarify system efficiency, while IMN may reveal how far lenders are willing to fund projects without IG customers (Figure 3).

See Appendix A-1 for Analyst Certification, Important Disclosures and Research Analyst Affiliations



Figure 1. A Summary of Signals

Metric	Unit	Latest Date	Period	Calculation	Period-Over-Period Δ	Latest Period	2nd Latest Period	3rd Latest Period	4th Latest Period	5th Latest Period	6th Latest Period	Trendline
Proprietary and Open Models (Artificial Analysis)												
Highest Intelligence for Proprietary Models	Points	10/8/26	Week	Point-in-Time	— 0.0%	58	58	58	53	53	48	
Highest Intelligence for Open Models	Points	10/8/26	Week	Point-in-Time	— 0.0%	46	46	46	45	44	43	
Gap in Proprietary and Open Intelligence	Points	10/8/26	Week	Point-in-Time	— 0.0%	12	12	12	8	9	5	
Highest Intelligence Models from the Top-20 Providers (Artificial Analysis)												
Median Intelligence	Points	10/8/26	Week	Point-in-Time	▲ +17.6%	40	34	38	33	32	32	
Median Latency	Sec	10/8/26	Week	Point-in-Time	▼ -9.6%	2.74	3.03	2.68	2.78	2.60	2.65	
Median Speed	Tok/Sec	10/8/26	Week	Point-in-Time	▲ +10.4%	106	96	104	109	90	86	
Release Cadence (Artificial Analysis, LLMTimeline, Model Provider Websites)												
Proprietary Models Released in Last Week	Models	10/6/26	Week	Weekly Total	▼ -70.0%	3	10	22	0	12	6	
Open Models Released in Last Week	Models	10/6/26	Week	Weekly Total	— 0.0%	0	0	2	4	6	2	
Proprietary Models Released in Last 90 Days	Models	10/6/26	Week	Trailing 90d	▲ +2.2%	95	93	90	68	69	58	
Open Models Released in Last 90 Days	Models	10/6/26	Week	Trailing 90d	▼ -2.2%	44	45	46	44	41	38	
Index Pricing (Bloomberg)												
A100 (SDA100RT)	$/Hour	10/7/26	Week	Weekly Avg	▼ -1.9%	1.54	1.57	1.57	1.59	1.61	1.61	
H100 (DESKH100)	$/Hour	10/7/26	Week	Weekly Avg	▲ +7.6%	2.99	2.78	2.69	2.65	2.65	2.57	
H200 (DESKH200)	$/Hour	10/7/26	Week	Weekly Avg	▼ -0.9%	3.39	3.42	3.42	3.37	3.36	3.32	
B200 (DESKB200)	$/Hour	10/7/26	Week	Weekly Avg	▲ +2.7%	5.37	5.23	5.06	4.95	4.90	4.90	
B300 (DESKB300)	$/Hour	10/7/26	Week	Weekly Avg	▲ +17.5%	6.78	5.77	5.76	6.57	6.26	5.77	
Avg Inference Token Expenditure (SDLLMTK)	$/1M Tok	10/6/26	Week	Weekly Avg	▼ -1.0%	0.97	0.98	0.98	0.96	0.97	1.06	
Model Pricing for Blended Tokens (Artificial Analysis)												
US+EU Avg Price	$/1M Tok	10/7/26	Week	Point-in-Time	▲ +3.5%	1.78	1.72	1.72	2.30	2.30	1.80	
China Avg Price	$/1M Tok	10/7/26	Week	Point-in-Time	— 0.0%	0.83	0.83	0.83	0.83	0.73	0.91	
Frontier Avg Price	$/1M Tok	10/7/26	Week	Point-in-Time	▲ +2.9%	1.40	1.36	1.36	1.71	1.67	1.44	
Proprietary Avg Price	$/1M Tok	10/7/26	Week	Point-in-Time	▼ -2.8%	2.09	2.15	2.15	2.89	3.17	2.23	
Open Avg Price	$/1M Tok	10/7/26	Week	Point-in-Time	▼ -10.3%	0.61	0.68	0.68	0.68	0.67	0.75	
Cost per Intelligence Index Task (Artificial Analysis)												
Proprietary Avg Cost	$/Task	10/8/26	Week	Point-in-Time	▼ -4.2%	3.38	3.53	2.00	2.06	1.46	0.42	
Open Avg Cost	$/Task	10/8/26	Week	Point-in-Time	— 0.0%	0.80	0.80	0.80	1.24	0.85	0.35	
Model Cost Composition per Intelligence Index Task (Artificial Analysis)												
Input Share of Task Cost	%	10/8/26	Week	Point-in-Time	▲ +0.3	4.0	3.7	6.2	6.3	2.7	1.8	
Cache Hit Share of Task Cost	%	10/8/26	Week	Point-in-Time	▼ -3.1	35.7	38.8	36.9	36.2	45.4	35.0	
Cache Write Share of Task Cost	%	10/8/26	Week	Point-in-Time	▲ +1.1	22.5	21.4	23.9	23.9	26.1	31.1	
Reasoning Share of Task Cost	%	10/8/26	Week	Point-in-Time	▲ +0.7	22.5	21.8	17.2	17.2	16.3	22.5	
Answer Share of Task Cost	%	10/8/26	Week	Point-in-Time	▲ +1.0	15.3	14.3	15.7	16.3	9.6	9.7	
Avg Output Tokens per Task	Tokens	10/8/26	Week	Point-in-Time	▼ -1.7%	57,067	58,035	37,667	37,269	43,690	26,029	
Developer Usage (PyPI)												
Multi-Provider SDK Downloads	%	10/7/26	Week	WoW Δ in 7d Avg	▼ -3.3	+4.0	+7.3	+4.4	+2.4	+5.6	-32.5	
Gemini SDK Downloads	%	10/7/26	Week	WoW Δ in 7d Avg	▼ -1.9	+3.0	+4.9	-0.5	+0.9	+8.2	-47.2	
Mistral SDK Downloads	%	10/7/26	Week	WoW Δ in 7d Avg	▲ +9.5	-9.1	-18.6	+17.0	-12.0	+27.6	-42.7	
App Usage (Sensor Tower)												
Gemini App WAU	%	10/5/26	Week	WoW Δ	▼ -2.0	+1.6	+3.6	+0.6	-0.2	+2.7	+3.4	
Qwen App WAU	%	10/5/26	Week	WoW Δ	▲ +4.2	+1.5	-2.7	-2.4	+4.8	-1.1	+6.8	
Kimi App WAU	%	10/5/26	Week	WoW Δ	▲ +2.6	+0.4	-2.2	-0.9	-3.0	-1.8	-3.5	
Web Usage (Similarweb)												
Gemini Web Visits	%	9/30/26	Week	WoW Δ in 7d Avg	▲ +2.4	+1.1	-1.3	+0.3	+1.9	-1.0	+3.1	
Qwen Web Visits	%	9/30/26	Week	WoW Δ in 7d Avg	▲ +1.4	-0.9	-2.3	-0.5	-1.9	+2.8	+2.4	
Kimi Web Visits	%	9/30/26	Week	WoW Δ in 7d Avg	▼ -5.7	-15.6	-9.9	-10.1	-10.5	-12.7	-27.9	
Layoffs (WSJ)												
All Layoffs Announced	People	10/7/26	Month	MTD	▼ -91.9%	652	8,089	1,761	17,420	24,709	26,160	
AI-Driven Layoffs Announced	People	10/7/26	Month	MTD	▲ +21.5%	570	469	321	3,220	21,640	15,360	

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Artificial Analysis updated its Intelligence Index tasks on Sep. 7, 2026; cost-per-task comparisons should be made separately for the periods before and after the update.

Source: Citi Research, Artificial Analysis, Bloomberg, PyPI Stats, Sensor Tower, Similarweb, Wall Street Journal

Figure 2. Weekend Content

Type	Source	Publication Date	Resource
Article	P-Zero Research	October 6, 2026	Introducing TasteVal
Podcast	Training Data	October 6, 2026	Google's AI Infrastructure Chief, Amin Vahdat, on the Physics & Economics of Frontier AI
Podcast	Lenny's Podcast	October 4, 2026	OpenAI's Head of ChatGPT: We're entering a new era of AI (again) | Tibo Sottiaux
Paper	Ash et al.	October 7, 2026	Demand for Conversational AI
Paper	Nvidia	October 7, 2026	VeriFine: Scaling Verification for Self-Improvement in Embodied Reasoning
Paper	Zhang et al.	October 1, 2026	Recursive Harness Self-Improvement for Frontier Reasoning Data Synthesis
Podcast	The a16z Show	October 3, 2026	Beyond the God Model | Alex Atallah & Amjad Masad
Podcast	The a16z Show	October 6, 2026	Building Defense for the Agentic Era: Kevin Mandia
Paper	Bek et al.	October 5, 2026	Knowledge or Calculator? Decomposing the Skill Premium in Verifiable Financial Agent Workflows
Article	Latent Space	October 6, 2026	Quasi-Riemann-Hypothesis: OpenAI publishes 722 math papers solving 90 of the top 500 open math problems
Podcast	Big Technology Podcast	October 7, 2026	Can AI Keep Growing Exponentially? Let's Ask SemiAnalysis — With Dylan Patel and Jordan Nanos
Paper	Sonthalia et al.	October 6, 2026	Agent in a Bottle: Can LLM Agents Turn Their Capabilities Into Cheap, Scalable Artifacts?
Podcast	The Artificial Intelligence Show	October 1, 2026	#244: How Alibaba.com Is Building the Future of Agent-to-Agent Commerce
Interview	Latent Space	October 2, 2026	Can a Cloud-Native Harness Make Agents Reliable Beyond the Desktop?
Article	Anaconda	October 6, 2026	Introducing the Expanded Anaconda Platform
Article	Redpine	October 6, 2026	Better Answers, Sharper Retrieval: Redpine Science Evaluated
Podcast	Machine Learning Street Talk	October 1, 2026	How a Voice Agent Learns the Rhythm of Conversation — Shawn Wen
Paper	Alibaba Cloud Computing	October 6, 2026	Rethinking Cross-Tokenizer On-Policy Distillation: From Alignment Coverage to Supervision Reliability
Podcast	The TWIML AI Podcast	October 6, 2026	Why Jev Is Changing How We Build With AI with Diogo Almeida

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Citi Research

Figure 3. Events Calendar

Start Date	End Date	Event	Type	Location
October 12, 2026	October 15, 2026	OCP Global Summit	Conference	San Jose, CA
October 13, 2026	October 15, 2026	SEMICON West	Conference	San Francisco, CA
October 15, 2026	October 15, 2026	IMN's Data Centers Private Equity	Conference	San Francisco, CA
October 18, 2026	October 26, 2026	Open Source AI Week	Conference	San Francisco, CA
October 19, 2026	October 21, 2026	Cloudflare Connect	Conference	San Francisco, CA
October 20, 2026	October 21, 2026	PyTorch Conference	Conference	San Jose, CA
October 20, 2026	October 22, 2026	NVIDIA GTC Berlin	Conference	Berlin, Germany
October 24, 2026	October 29, 2026	EMNLP	Conference	Budapest, Hungary
October 25, 2026	October 28, 2026	Oracle AI World	Conference	Las Vegas, NV
November 9, 2026	November 12, 2026	Conference on Robotic Learning	Conference	Austin, TX
November 15, 2026	November 19, 2026	ACM CCS 2026	Conference	The Hague, The Netherlands
November 15, 2026	November 20, 2026	Supercomputing Conference	Conference	Chicago, IL
November 17, 2026	November 20, 2026	Microsoft Ignite	Conference	San Francisco, CA
November 30, 2026	December 3, 2026	NVIDIA GTC Washington, D.C.	Conference	Washington, D.C.
November 30, 2026	December 4, 2026	AWS re:Invent	Conference	Las Vegas, NV
December 3, 2026	December 4, 2026	Adopt AI	Conference	Paris, France
December 6, 2026	December 12, 2026	NeurIPS	Conference	Sydney, Australia
December 7, 2026	December 11, 2026	GITEX	Conference	Dubai, UAE
December 9, 2026	December 10, 2026	The AI Summit New York	Conference	New York, NY
December 12, 2026	December 16, 2026	IEE International Electron Devices Meeting	Conference	San Francisco, CA
December 31, 2026	December 31, 2026	U.S. BIS HBM & SME Export-Control Deadline	Deadline	Washington, D.C.
January 1, 2027	January 1, 2027	New York RAISE Act Takes Effect	Deadline	NY
January 6, 2027	January 9, 2027	CES	Conference	Las Vegas, NV
January 21, 2027	January 23, 2027	AIEI	Conference	Bangalore, India
January 28, 2027	January 28, 2027	Generative AI Summit	Conference	Washington, D.C.
January 30, 2027	February 4, 2027	SPI Photonics West	Conference	San Francisco, CA
February 16, 2027	February 23, 2027	AAAI	Conference	Montréal, Canada
March 15, 2027	March 18, 2027	NVIDIA GTC	Conference	San Jose, CA
April 5, 2027	April 8, 2027	RSAConference	Conference	San Francisco, CA
April 13, 2027	April 15, 2027	Google Cloud Next	Conference	Las Vegas, NV
April 26, 2027	April 30, 2027	ICLR	Conference	San Francisco, CA
June 1, 2027	June 5, 2027	NAACL	Conference	San Francisco, CA
June 7, 2027	June 10, 2027	Snowflake Summit	Conference	San Francisco, CA
June 9, 2027	June 10, 2027	The AI Summit London	Conference	London, UK
June 20, 2027	June 24, 2027	IEEE CVPR	Conference	Seattle, WA
June 21, 2027	June 24, 2027	Databricks Data + AI Summit	Conference	San Francisco, CA

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Citi Research

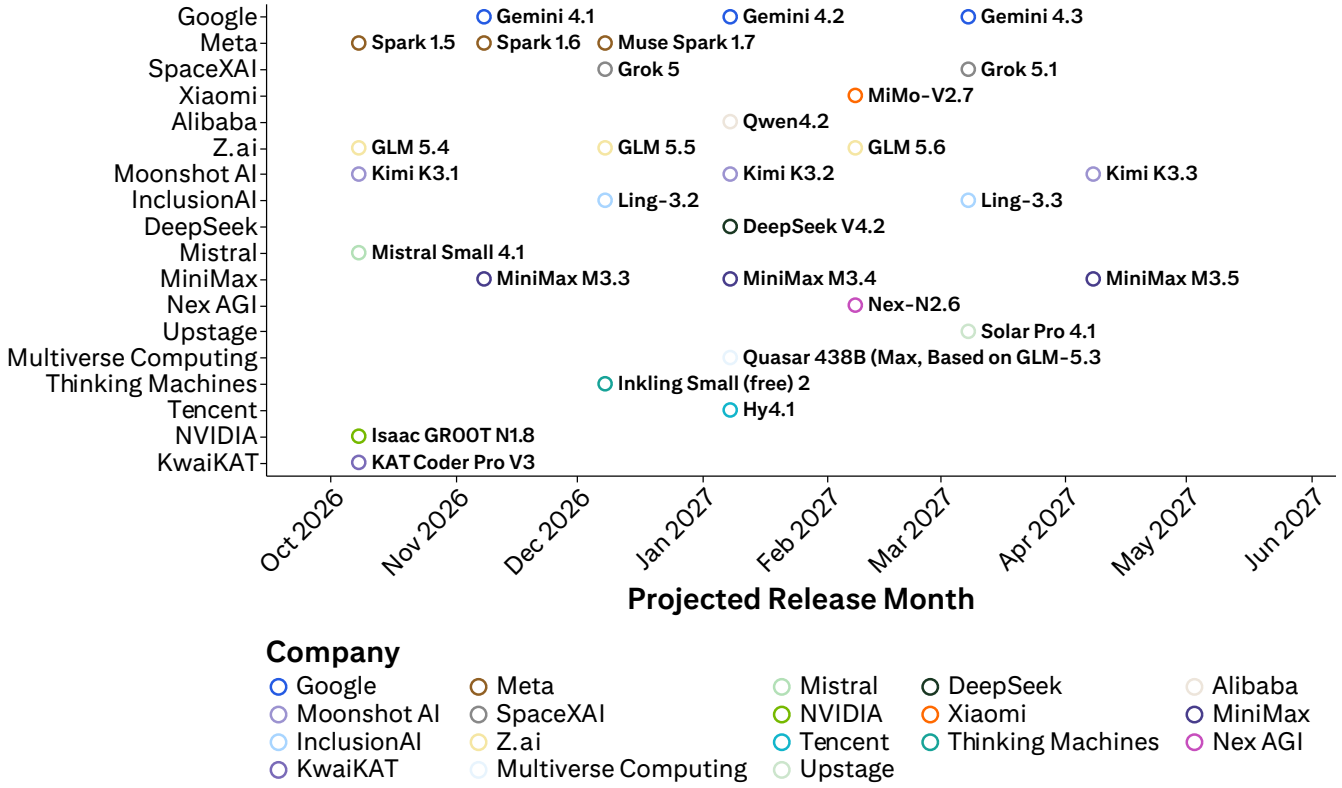
Figure 4. Model Metrics

MoM Δ in Provider Rank	Model Provider	Leading Model	Artificial Analysis Intelligence Index	Speed (Output Tokens/ Second)	Blended Price (USD/ 1M Tokens)	Context Window (1K Tokens)	Release Date	License
▲5	**Google**	Gemini 4 Argon (high)	53	239	$1.47	1000	Sep. 30, 2026	Proprietary
▼1	**OpenAI**	GPT-6 Astra (max)	53	48	$7.70	1000	Sep. 3, 2026	Proprietary
▼1	**Meta**	Muse Spark 1.3 (max)	48	209	$0.78	1000	Sep. 2, 2026	Proprietary
▲1	**SpaceXAI**	Grok 4.7 (xhigh)	46	73	$1.35	500	Sep. 21, 2026	Proprietary
▲11	**Xiaomi**	MiMo-V2.6-Pro	46	38	$0.17	1000	Sep. 22, 2026	Open
▲1	**Alibaba**	Qwen3.8 Max (0902)	45	37	$1.18	984	Sep. 16, 2026	Proprietary
▼4	**Z.ai**	GLM-5.3 (max)	45	79	$0.90	1000	Aug. 18, 2026	Open
▼4	**Moonshot AI**	Kimi K3 (max)	44	42	$2.31	1050	Jul. 16, 2026	Open
▲8	**InclusionAI**	Ling 3.1 Flash	41	211	$0.19	1000	Oct. 8, 2026	Proprietary
▼2	**DeepSeek**	DeepSeek V4.1 Flash (max)	39	216	$0.19	1000	Sep. 14, 2026	Open
NEW	**Mistral**	Mistral Large 4 Preview	38	116	$0.79	524	Oct. 8, 2026	Proprietary
▼2	**MiniMax**	MiniMax-M3	29	94	$0.22	1000	Jun. 1, 2026	Open
▼2	**Nex AGI**	Nex-N2-Pro	28	116	$0.53	262	Jun. 2, 2026	Open
▼2	**Upstage**	Solar Pro 4	28	111	$0.22	512	Aug. 6, 2026	Proprietary
▼2	**Multiverse Computing**	Quasar 438B (max)	27	126	$0.72	1000	Aug. 10, 2026	Proprietary
▼1	**Thinking Machines**	Inkling Small	26	205	$0.22	1000	Jul. 30, 2026	Open
▼3	**Tencent**	Hy3	25	85	$0.10	256	Jul. 6, 2026	Open
—	**NVIDIA**	Nemotron 3 Ultra	23	155	$0.48	262	Jun. 1, 2026	Open
—	**KwaiKAT**	KAT-Coder-Pro V2	22	102	$0.22	256	Mar. 27, 2026	Proprietary

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Citi Research, Artificial Analysis, Model Provider Websites

Figure 5. Projected Model Releases



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Citi Research, OpenRouter, Artificial Analysis, LLMTimeline, Model Provider Websites

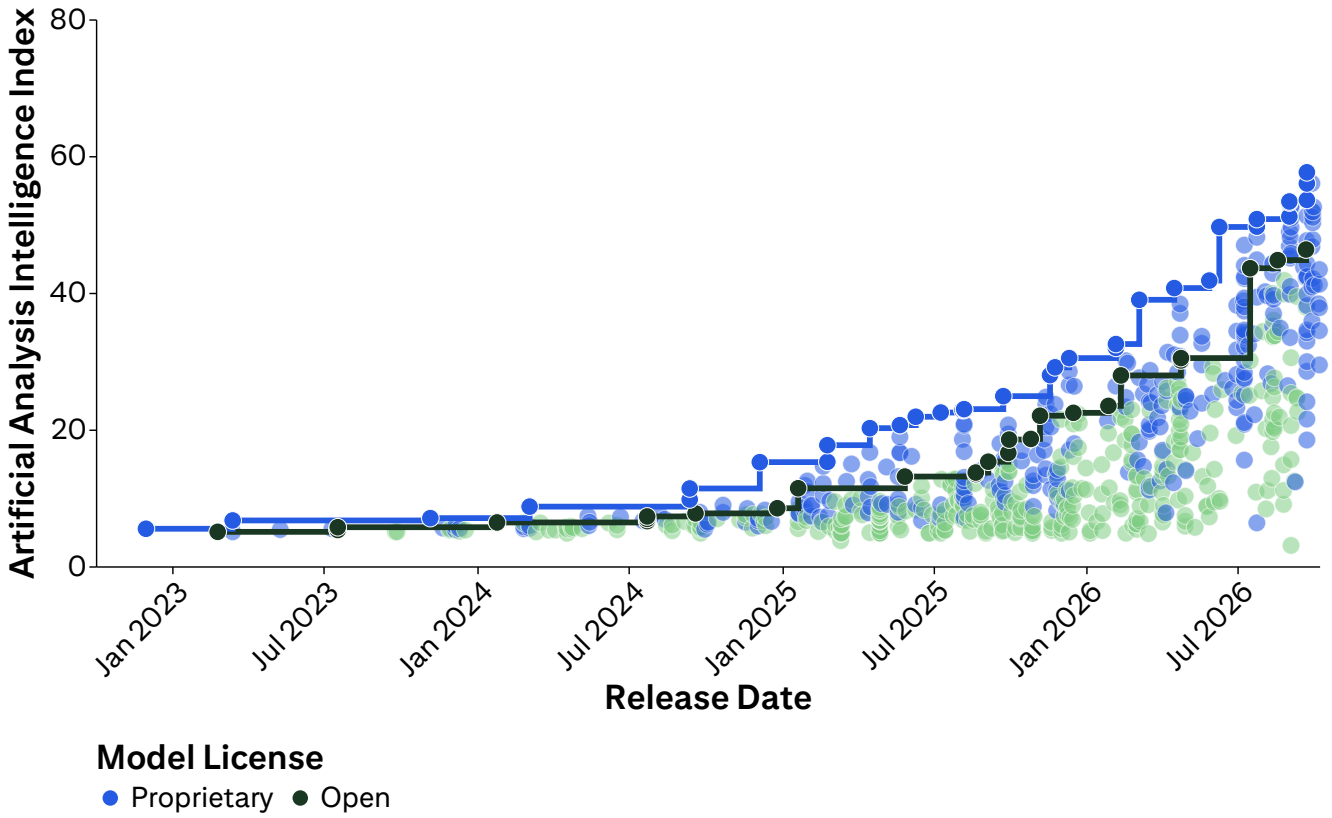
Figure 6. Historical Model Releases



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

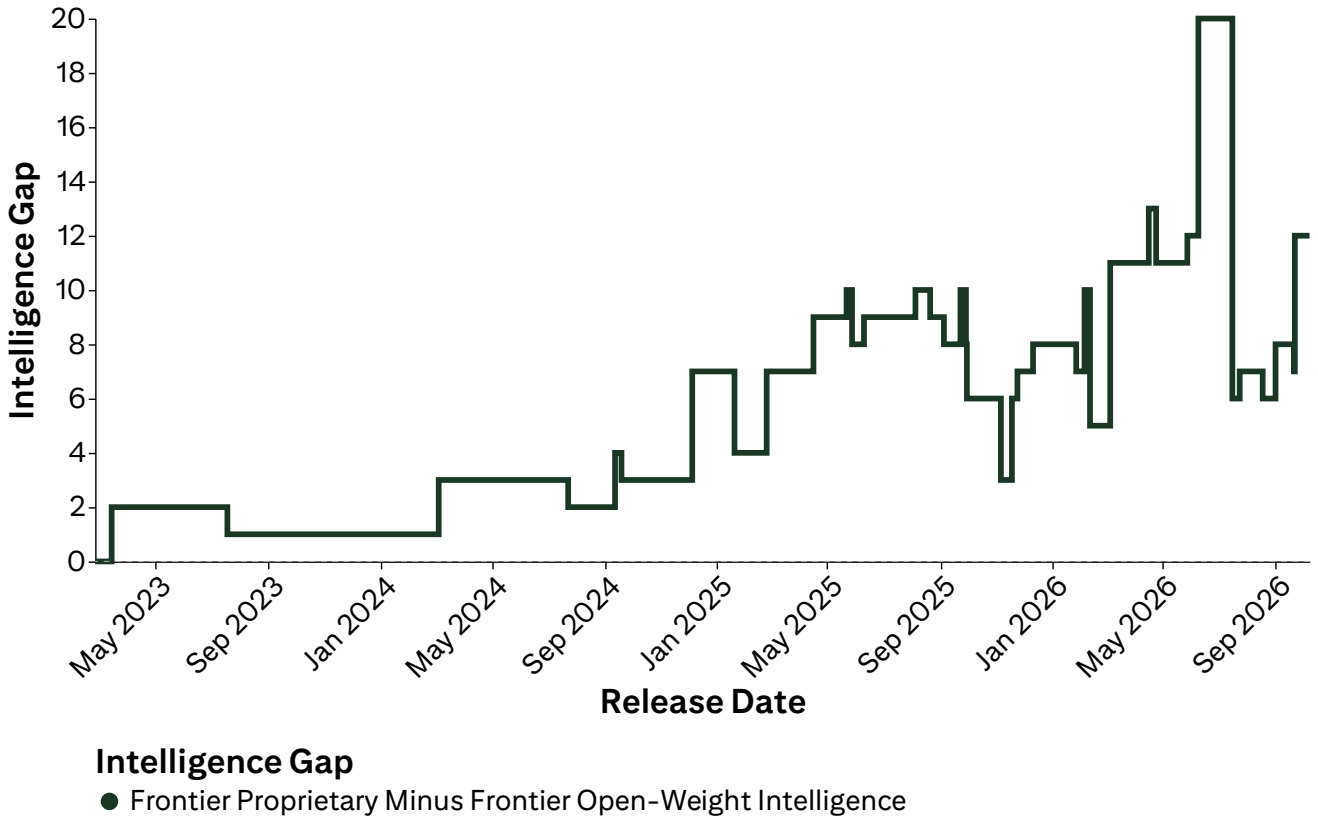
Source: Citi Research, OpenRouter, Artificial Analysis, LLMTimeline, Model Provider Websites

Figure 7. Frontier Releases for Proprietary and Open Models



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Citi Research, Artificial Analysis

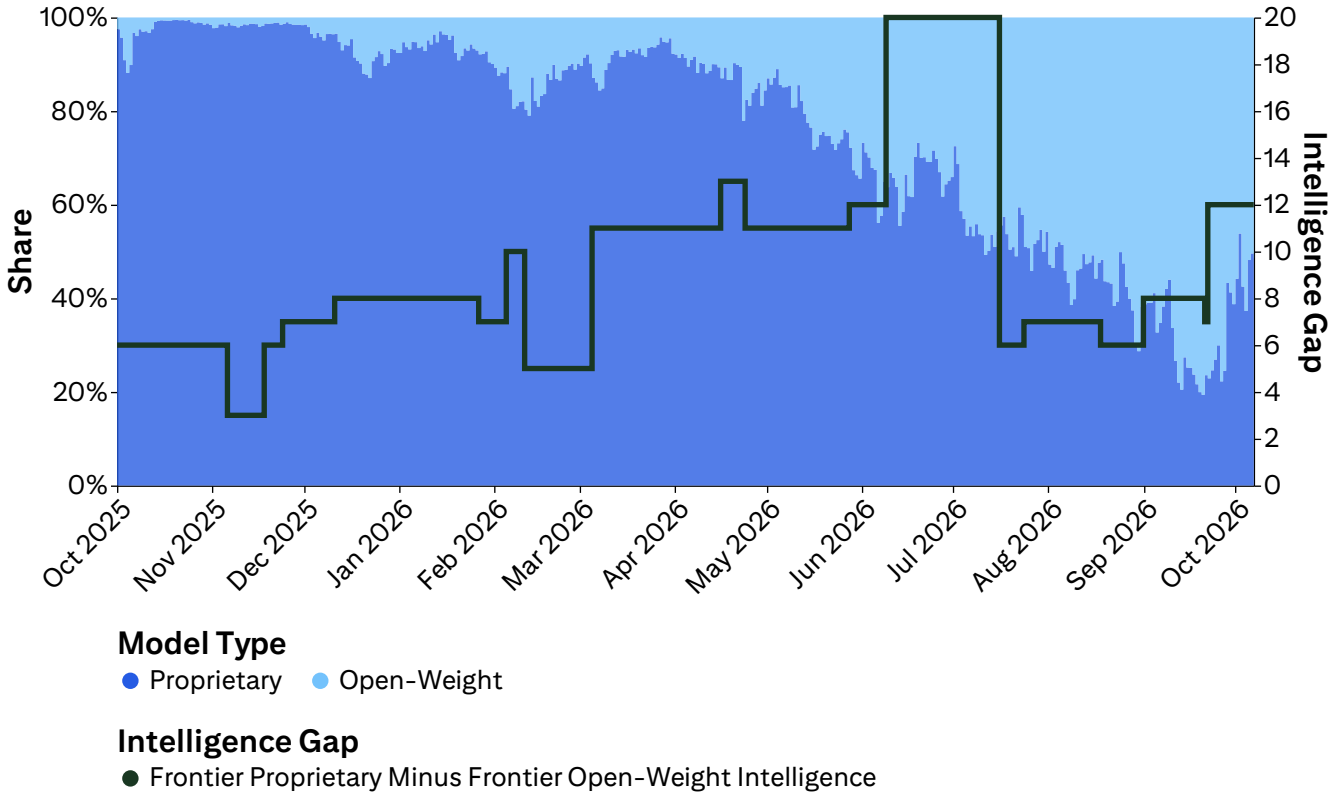
Figure 8. Gap in Proprietary and Open Intelligence



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Citi Research, Artificial Analysis

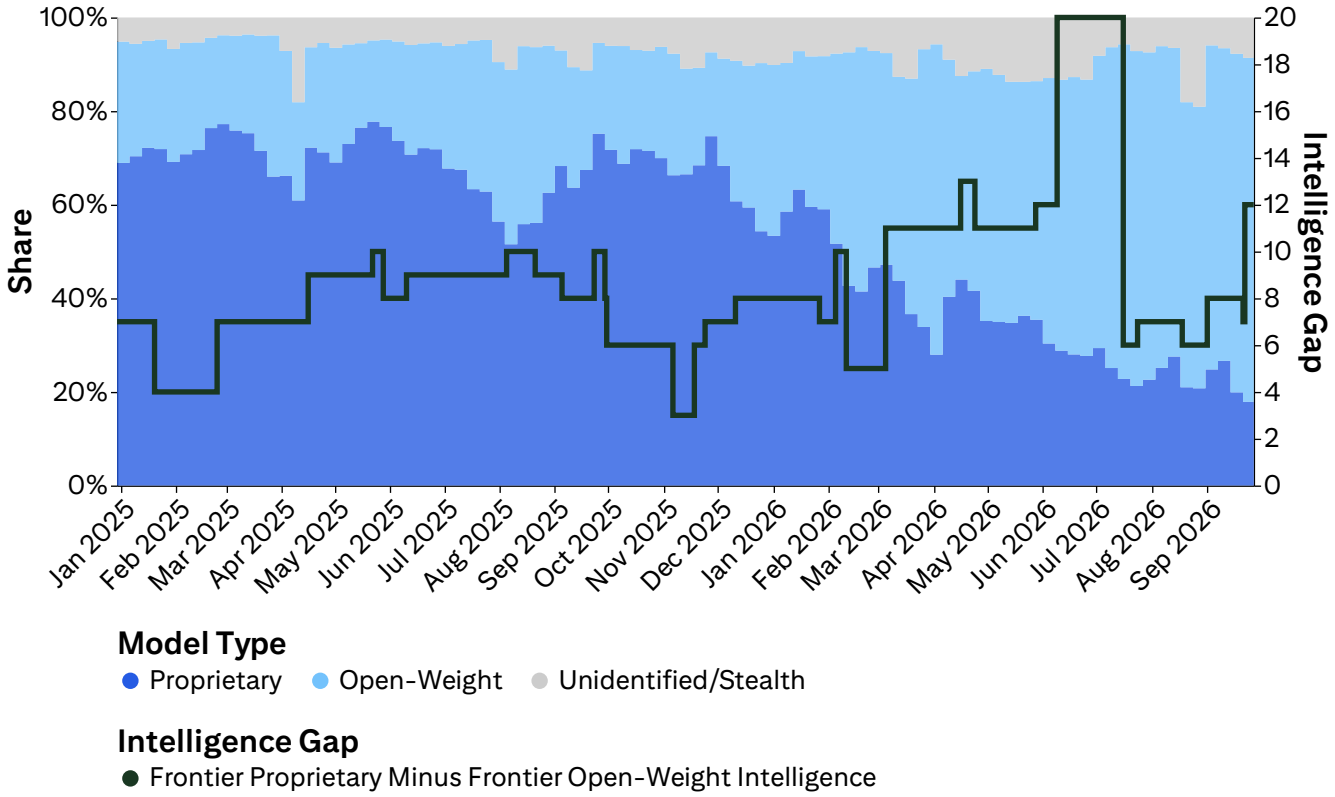
Figure 9. Vercel Token Volume Shares



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

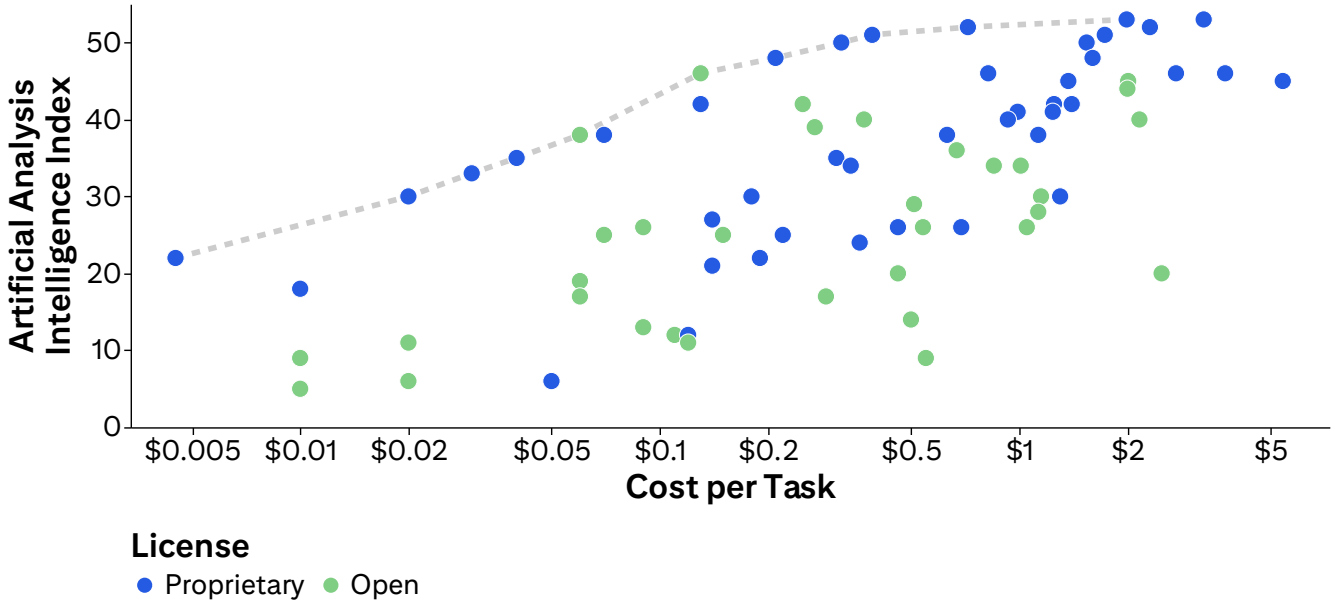
Source: Citi Research, Vercel

Figure 10. OpenRouter Token Volume Shares



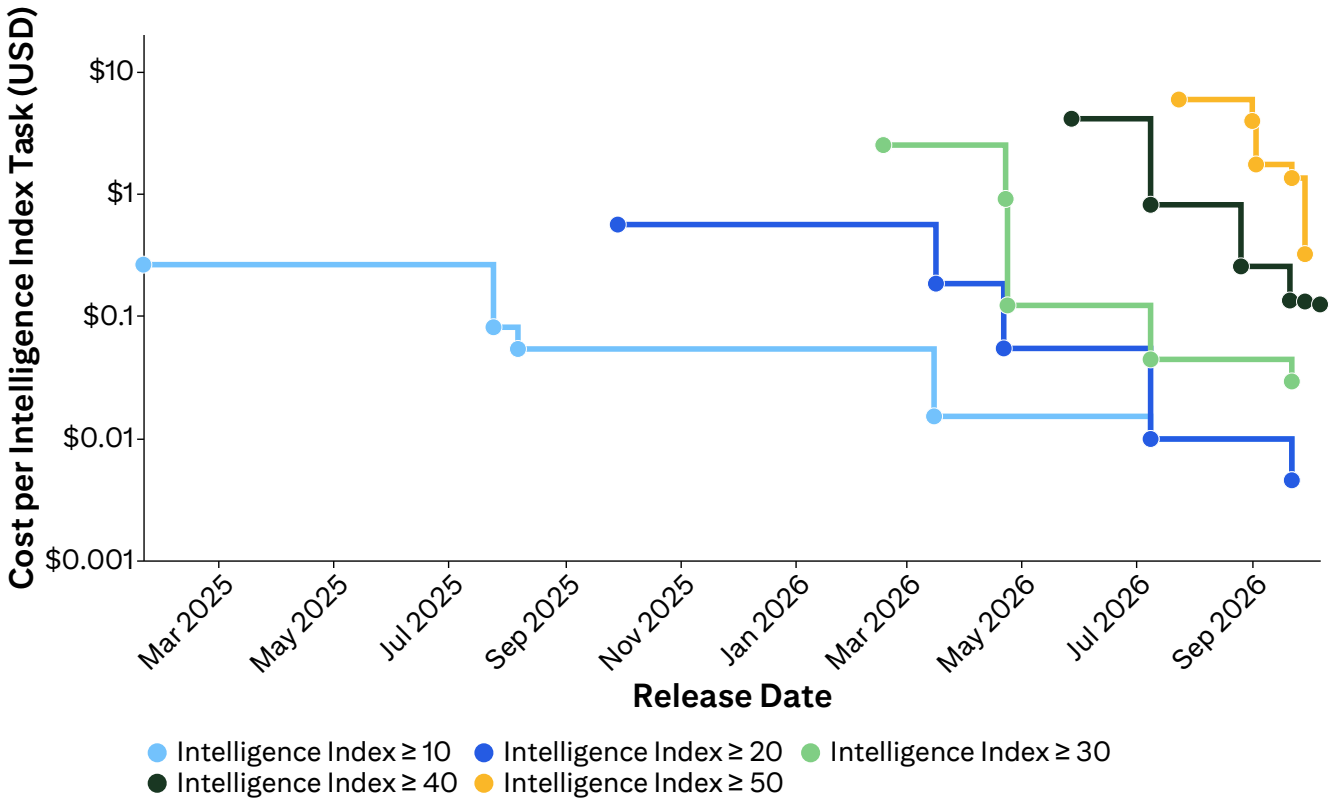
© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Citi Research, OpenRouter

Figure 11. Intelligence v. Cost per AA Task



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Citi Research, Artificial Analysis

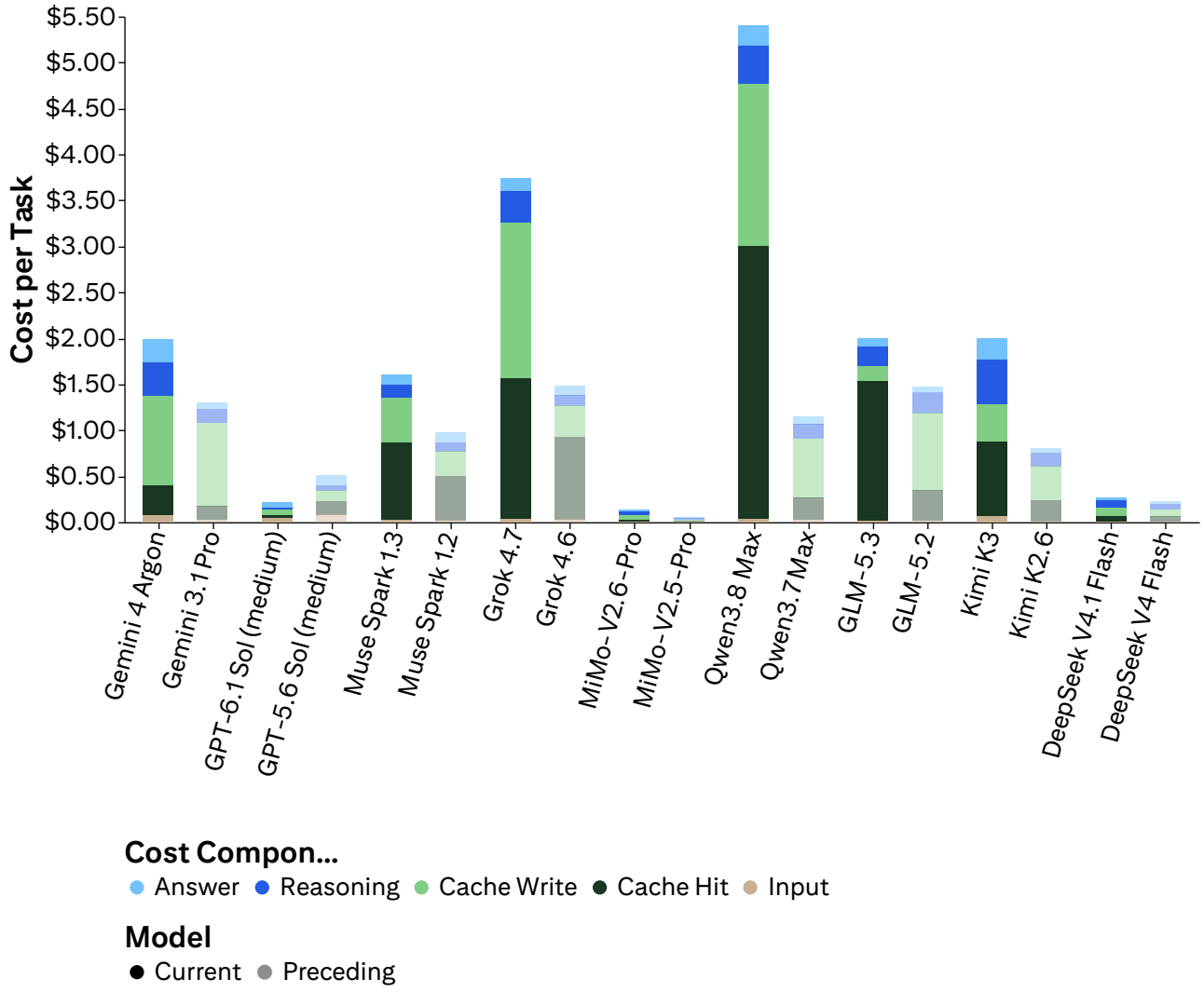
Figure 12. Intelligence Over Time v. Blended Price



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Citi Research, Artificial Analysis

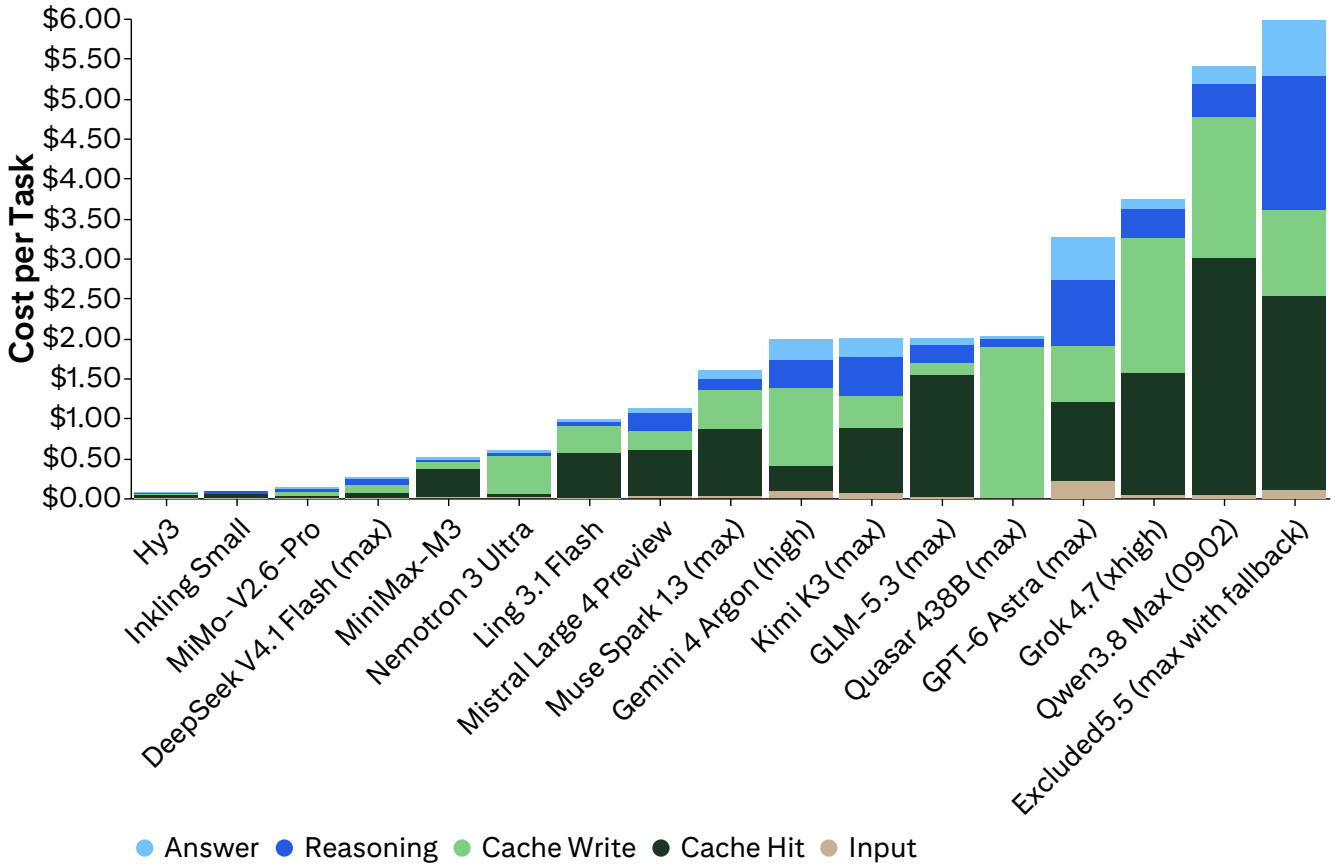
Figure 13. Cost per AA Task for Current Versus Prior Models



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

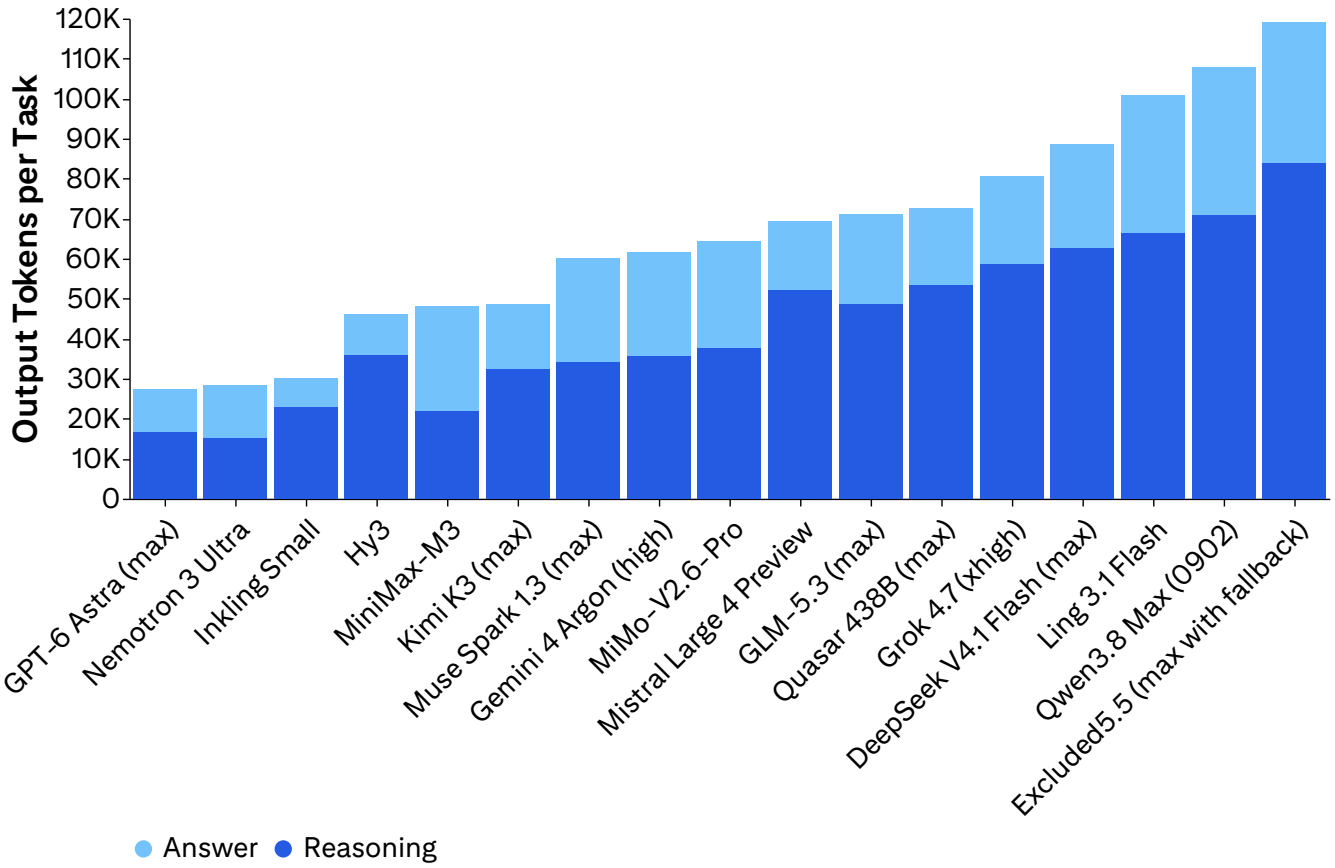
Source: Citi Research, Artificial Analysis

Figure 14. Cost Breakdown per AA Task



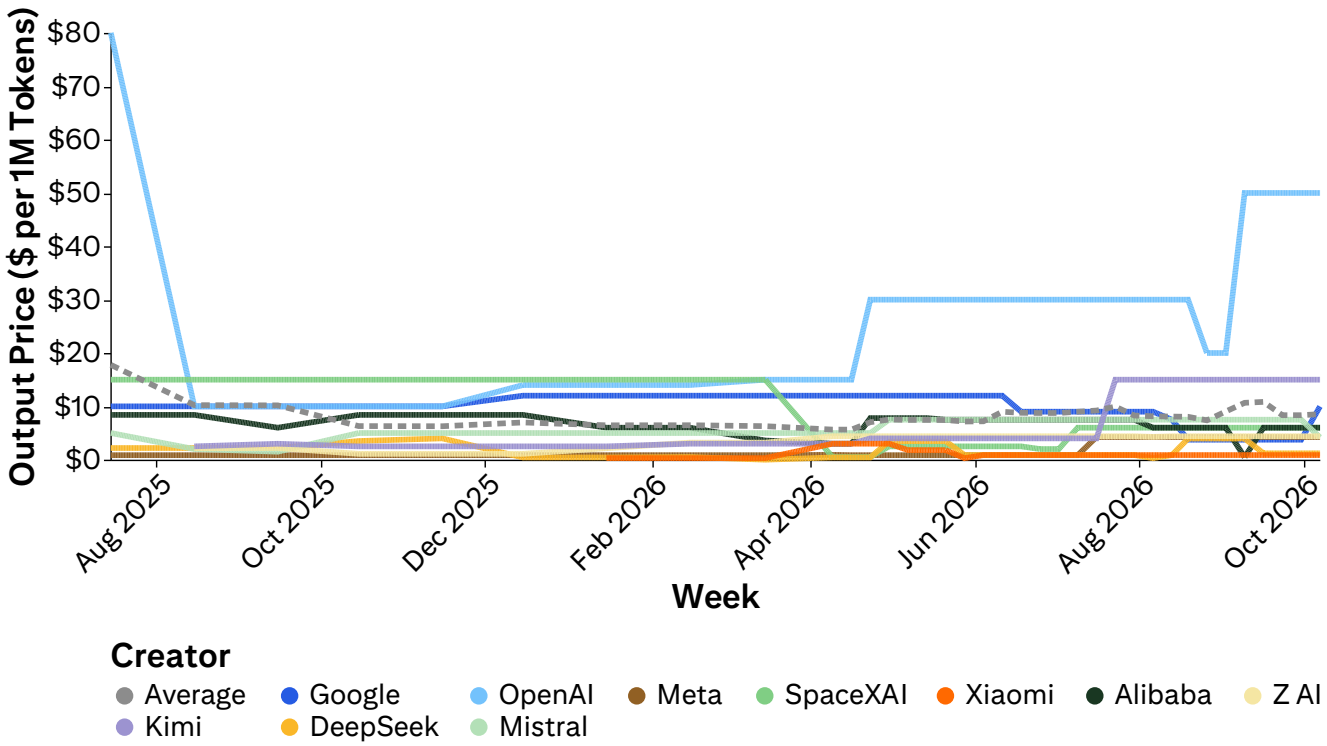
© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Citi Research, Artificial Analysis

Figure 15. Output Tokens per AA Task



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Citi Research, Artificial Analysis

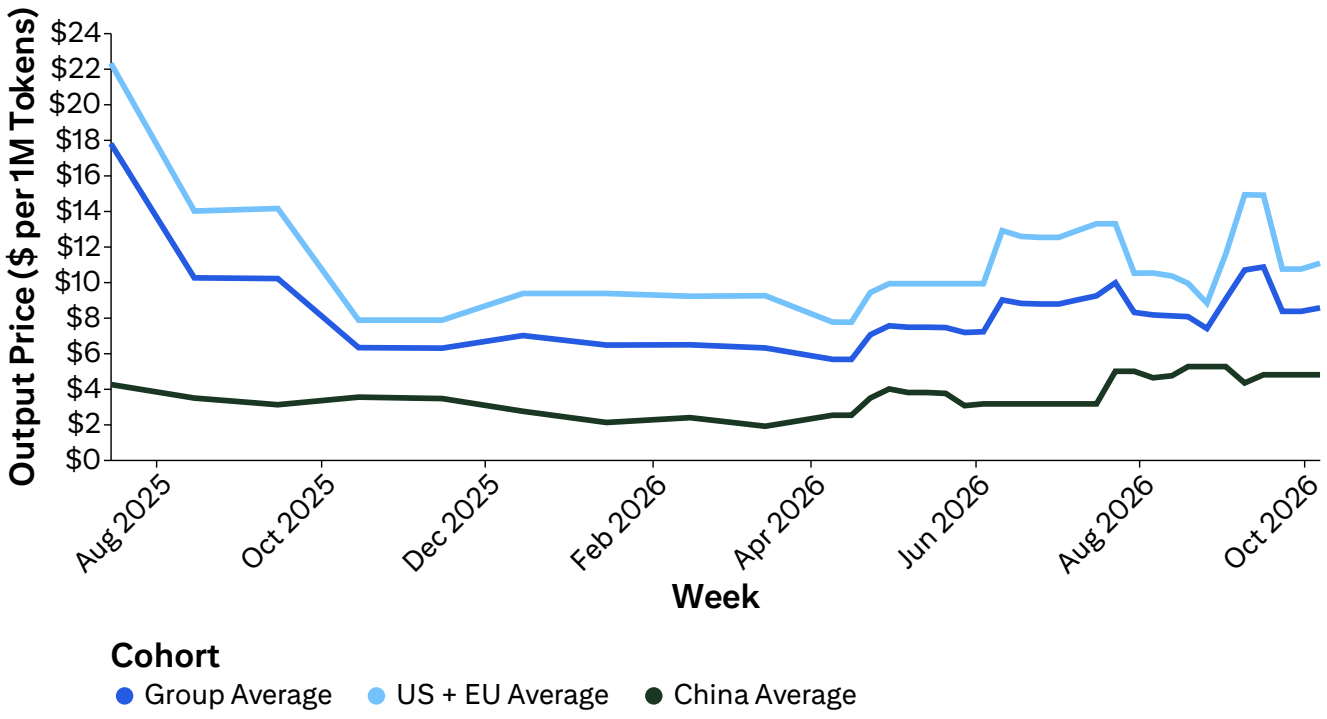
Figure 16. Output Token Pricing by Creator



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

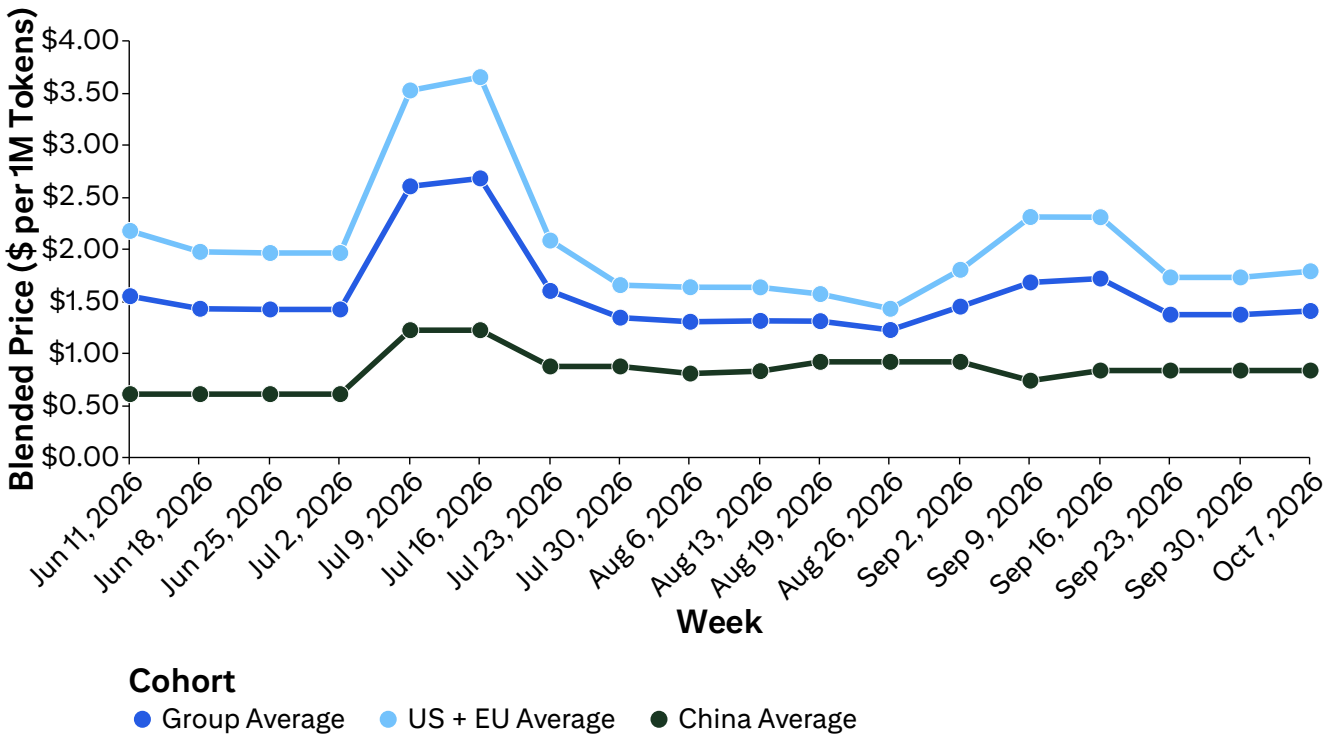
Source: Citi Research, Artificial Analysis

Figure 17. Output Token Pricing



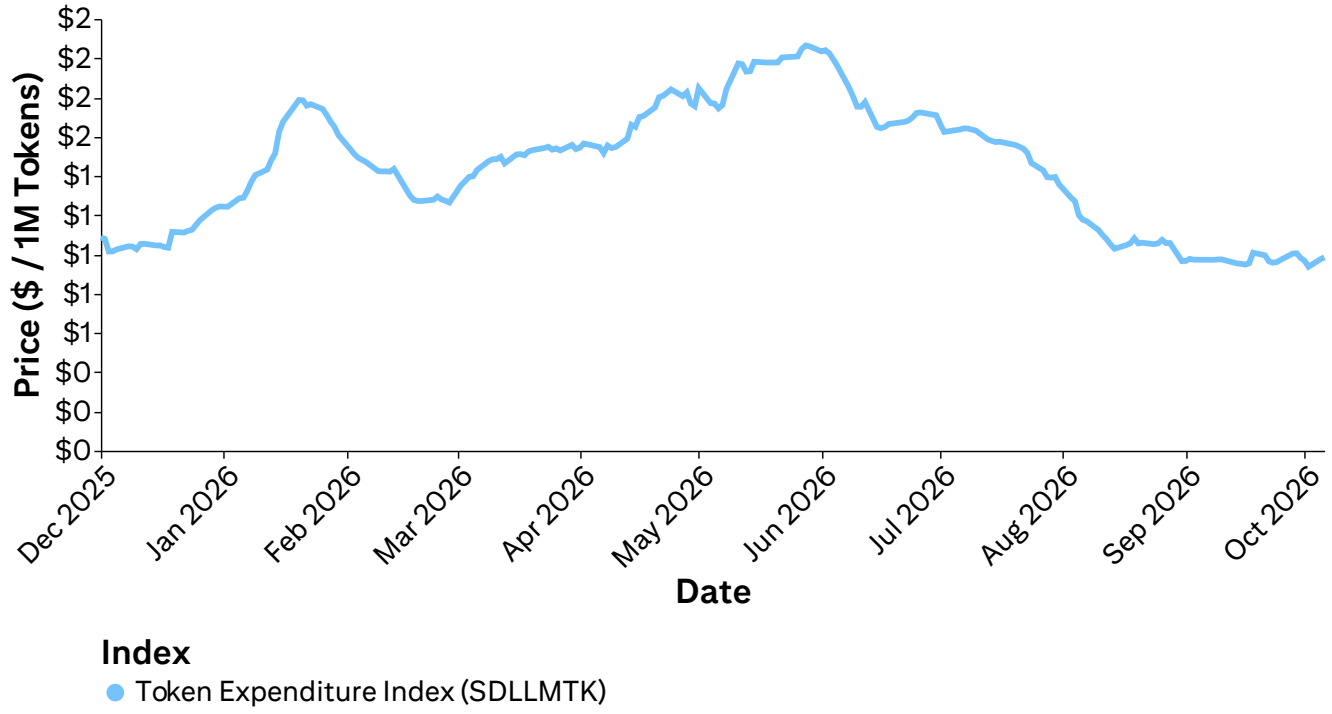
© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Citi Research, Artificial Analysis

Figure 18. Blended Token Pricing



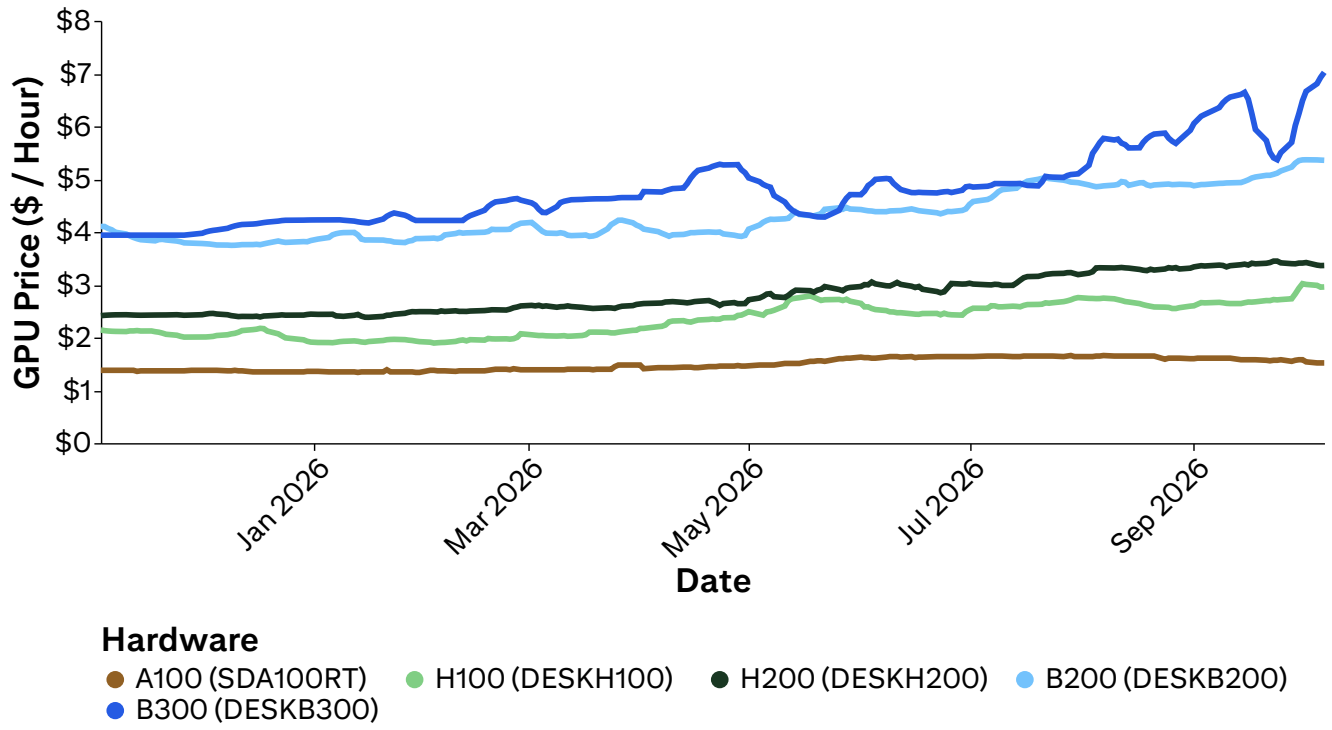
© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Citi Research

Figure 19. Silicon Data Token Expenditure Index



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Citi Research, Bloomberg

Figure 20. Rental Pricing for GPUs



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Citi Research, Bloomberg

Figure 21. Recent Notes from Our Team

Category	Date	Title
AI Industry Model	September 3, 2026	Artificial Intelligence: AI Industry Model Update
Deep Dives	September 7, 2026	Space Exploration Technologies (SPCX.O): Potential for $900+; Initiating at Buy
Deep Dives	January 21, 2026	Artificial Intelligence: OpenAI - Architecting the Abstraction Layer for Everything
Deep Dives	August 25, 2026	Artificial Intelligence: AI - The Information Era's Apex Technology
ROI on AI	October 1, 2026	Artificial Intelligence: ROI on AI - The Future of Open Is More Control than Cost as Frontier Labs Pull Ahead
ROI on AI	August 4, 2026	Artificial Intelligence: ROI on AI - Consensus Is Still Underestimating AI Revenues & Returns
ROI on AI	April 30, 2026	Artificial Intelligence: ROI on AI - Backlog Acceleration Drives Continued Confidence in Investment Returns
ROI on AI	February 6, 2026	Artificial Intelligence: ROI on AI - Growth Acceleration & Returns Drive Capex Revisions
Citi's Inference Ahead	October 2, 2026	Artificial Intelligence: Citi's Inference Ahead - Building the Rails for Scaled Intelligence
Citi's Inference Ahead	September 26, 2026	Artificial Intelligence: Citi's Inference Ahead - The Efficiency Dividend
Citi's Inference Ahead	September 18, 2026	Artificial Intelligence: Citi's Inference Ahead - Where Intelligence Compounds
Citi's Inference Ahead	September 14, 2026	Artificial Intelligence: Citi's Inference Ahead - Reasserting the Western Frontier
Citi's Inference Ahead	September 4, 2026	Artificial Intelligence: Citi's Inference Ahead - The Efficiency Frontier
Citi's Inference Ahead	August 31, 2026	Artificial Intelligence: Citi's Inference Ahead - The Intelligence Flywheel
Citi's Inference Ahead	August 21, 2026	Artificial Intelligence: Citi's Inference Ahead - Investing in Open Models
Citi's Inference Ahead	August 14, 2026	Artificial Intelligence: Citi's Inference Ahead - Holding Back the Frontier
Citi's Inference Ahead	August 7, 2026	Artificial Intelligence: Citi's Inference Ahead - From Buildout to Monetization
Citi's Inference Ahead	July 31, 2026	Artificial Intelligence: Citi's Inference Ahead - Pricing, Financing, Routing, and Securing
Citi's Inference Ahead	July 24, 2026	Artificial Intelligence: Citi's Inference Ahead - Constraints v. Capabilities
Citi's Inference Ahead	July 19, 2026	Artificial Intelligence: Citi's Inference Ahead - Open Reaches the Frontier
Citi's Inference Ahead	July 10, 2026	Artificial Intelligence: Citi's Inference Ahead - Models, Robots, & Captives
Citi's Inference Ahead	June 26, 2026	Artificial Intelligence: Citi's Inference Ahead - Open-Source Moment
Citi's Inference Ahead	June 19, 2026	Artificial Intelligence: Citi's Inference Ahead - Open v. Closed
Citi's Inference Ahead	June 14, 2026	Artificial Intelligence: Citi's Inference Ahead - Bridging the Gaps
Citi's Inference Ahead	June 10, 2026	Artificial Intelligence: Citi's Inference Ahead - Rationing Comes for Us All
One-Off Pieces	September 14, 2026	Artificial Intelligence: Pacing Calls Attempt to Shape Regulation
At the Edge	June 11, 2026	Artificial Intelligence: At the Edge - Verification, Agents, and the Infrastructure Around Them
Commentary on Conferences and Events	October 2, 2026	US Application Software - Decagon Dialogues Showcases a More Connected Platform with Expanding Uses
Commentary on Conferences and Events	September 10, 2026	Artificial Intelligence and Systems Software: Enterprise AI: Island - Day 3 Takeaways from Citi's Global TMT Conference
Commentary on Conferences and Events	September 10, 2026	Artificial Intelligence: Orbital Compute: Starcloud - Day 3 Takeaways from Citi's Global TMT Conference
Commentary on Conferences and Events	September 10, 2026	Artificial Intelligence: Context for Enterprise AI: Celonis - Day 2 Takeaways from Citi's Global TMT Conference
Commentary on Conferences and Events	September 9, 2026	Artificial Intelligence: AI's Open Models: Poolside - Day 2 Takeaways from Citi's Global TMT Conference
Commentary on Conferences and Events	September 9, 2026	Artificial Intelligence: AI's Evolution: Perplexity - Day 2 Takeaways from Citi's Global TMT Conference
Commentary on Conferences and Events	September 9, 2026	Artificial Intelligence: Powering AI: Commonwealth Fusion Systems - Day 2 Takeaways from Citi's Global TMT Conference
Commentary on Conferences and Events	September 9, 2026	Artificial Intelligence: State of the Market - Day 2 Takeaways from Citi's Global TMT Conference
Commentary on Conferences and Events	September 8, 2026	Artificial Intelligence: Uniphore - Day 1 Takeaways from Citi's Global TMT Conference
Commentary on Conferences and Events	September 8, 2026	Artificial Intelligence: Genspark and Cresta Intelligence - Day 1 Takeaways from Citi's Global TMT Conference
Commentary on Conferences and Events	September 8, 2026	Artificial Intelligence: Implementing AI & Robotics: Wonder - Day 1 Takeaways from Citi's Global TMT Conference
Commentary on Conferences and Events	September 8, 2026	Artificial Intelligence: Enterprise Technology Leadership Keynote - Day 1 Takeaways from Citi's Global TMT Conference
Commentary on Conferences and Events	August 31, 2026	Global Technology & Communications: Citi's 2026 Global TMT Conference Preview - Key Tech Themes to Watch
Commentary on Conferences and Events	July 8, 2026	Artificial Intelligence: AI Insights from Citi's Space & AI Virtual Teach-In
Commentary on Conferences and Events	July 8, 2026	Artificial Intelligence: Bottlenecks Remain for Scaling Physical AI, But Commercialization Progressing
Commentary on Conferences and Events	June 15, 2026	Artificial Intelligence: Takeaways from the Research and Applied AI Summit
Collaborative Pieces	October 6, 2026	North America Internet: The Muse Moment: Revenue Analysis by Vertical & Framework to Evaluate Muse's Impact Across Multiple Sectors
Collaborative Pieces	September 29, 2026	Space Exploration Technologies (SPCX.O): Starship Flight 14 Top Takes
Collaborative Pieces	September 14, 2026	Continual Learning to Fuel HBM/eSSD Demand (AI Training/Inference)
Collaborative Pieces	September 3, 2026	Multi-Asset: Data Center GAINs (Gen AI Names) - Born to Run: Still Early in the Multi-Year AI Investment Cycle
Collaborative Pieces	August 9, 2026	Space Exploration Technologies (SPCX.O): 2Q26 Model Update; Reiterate $200 Price Target
Collaborative Pieces	August 6, 2026	Takeaways on Starlink's Mobile Strategy & Implications for the Category
Collaborative Pieces	August 4, 2026	Space Exploration Technologies (SPCX.O): 2Q26 Strong Start to Multi-Year Catalyst-Rich Arc
Collaborative Pieces	July 30, 2026	US Application Software: 2026 CFO Survey - Measured Optimism with Clear Commitment to AI
Collaborative Pieces	July 19, 2026	EM Equity Strategy Compass: 2H26 Outlook - The Broadening Question
Collaborative Pieces	July 10, 2026	Global Equity Quarterly: Back to the Broadening Debate
Collaborative Pieces	July 1, 2026	North America: Road Ahead 2H26
Collaborative Pieces	July 1, 2026	North America Internet: Meta Compute Unlocks Capacity (and FCF) for Meta's Superintelligence Ambitions

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Citi Research

Figure 22. Sector-Specific Notes on AI

Category	Date	Title
Hardware and Semis	August 4, 2026	ON Semiconductor (ON.O): "Beat & Raise" as AI Data Center Revenue Accelerates
Hardware and Semis	August 3, 2026	US Semiconductors and Semiconductor Equipment: Earnings so Far...Estimates Up, Stocks Down; Buyers of AMD, TXN, & AMAT
Hardware and Semis	July 16, 2026	TSMC (2330.TW): AI demand keeps accelerating; Upside surprise on further capex increase
Data Centers	July 23, 2026	US Engineering & Construction - 2Q26 E&C Preview – E&C Growth Cycle Still Durable
Data Centers	July 17, 2026	Asia Communication Infrastructure: 800VDC Architecture for Next-Generation AI Data Centers
Data Centers	July 2, 2026	US Industrial Technology & Mobility: Takeaways from Construction Economist Call
Internet	July 24, 2026	North America Internet - 2Q26 Preview: RDDT, PINS, SNAP, CVNA, and DASH
Internet	July 22, 2026	Alphabet Inc (GOOGL.O) - 1st Impression: AI Momentum Drives Cloud Accel & Backlog, Search & YT Momentum Continues
Internet	July 1, 2026	North America Internet: Meta Compute Unlocks Capacity (and FCF) for Meta's Superintelligence Ambitions
Software	July 31, 2026	US Internet and Software - 2Q26 Cloud Compare: Revenues Accelerate Amid Rising Capex Tide
Software	July 23, 2026	US Application Software - Expecting Diverging Trends Across 2Q Front Office Software Earnings
Software	July 9, 2026	US Software: Themes into 2Q26 Software Earnings - AI Drives Improvement in Spending; Remain Selective
Industrials	July 27, 2026	ITG Inc (ITG.O) - FTTH + Data Center Tailwinds Driving Growth; Initiate Buy/High Risk
Industrials	July 2, 2026	US Industrial Technology & Mobility: Takeaways from Construction Economist Call
Industrials	June 23, 2026	US Industrial Technology & Mobility - Automate 2026 Takeaways: Automation Momentum Remains Constructive
Utilities	June 16, 2026	Overcoming Gridlock 2.0: Solving Power Bottlenecks to Drive AI & Energy Security
Utilities	June 8, 2026	US Diversified Utilities: Crowd Metrics Moving Up; Taking Closer Look at Consensus Trades
Utilities	April 22, 2026	US Electric Utilities: Citi's AI Summit Takeaways for US Power/Utilities
Economics	August 3, 2026	US Economics - ISM Manufacturing boosted by AI tail-wind
Economics	July 9, 2026	Research @ Citi Episode 83: AI's Mid-Year Review - The Token Bottleneck
Economics	June 25, 2026	Global Economic Outlook & Strategy: The Resilient Global Economy - Cascading Supply Shocks
Corporate Credit	July 29, 2026	The Speculative Grade Beat - Underweight AI – Uncertainties Rising, Credit Not Adequately Compensating
Corporate Credit	July 24, 2026	The High-Grade Beat - Hyperscaler High Tide
Corporate Credit	July 7, 2026	US High-Grade Credit: AI and Credit Markets - The Buildout, The Risks, and How to Price Them
Miscellaneous	July 24, 2026	Emerging Markets Economic Outlook & Strategy - Artificial Intelligence and the EM Growth Divide
Miscellaneous	July 8, 2026	Global Theme Machine: Theme Leadership Shifts as Technology Improves but Healthcare Themes Lead Returns
Miscellaneous	June 10, 2026	Global Theme Machine: AI and Industrial Themes Extend Their Lead
CFO Survey	July 30, 2026	US Application Software: 2026 CFO Survey: Measured Optimism with Clear Commitment to AI
CIO Survey	July 9, 2026	Global Technology & Communications: Citi CIO Survey: IT Budgets Improve As AI Moves Higher In Priority
CIO Survey	April 9, 2026	Global Technology & Communications: Citi CIO Survey: IT Budgets Moderate Post 4Q Budget Flush
CIO Survey	January 12, 2026	Global Technology & Communications: Citi CIO Survey: Global IT Budgets Accelerate on Potential 4Q Budget Flush
CIO Survey	October 14, 2025	Global Technology & Communications: Citi CIO Survey: IT Budgets Move Below Historical Levels with US Weakening
CIO Survey	July 15, 2025	Global Technology & Communications: Citi CIO Survey: IT Budgets Moderate As Cyber Regains Top Priority
CISO Survey	June 9, 2026	North America Systems Software: On the Ground at Gartner Security Summit - Fresh Diligence on AI Security Trends
CISO Survey	January 12, 2026	North America Systems Software: State of Enterprise Cybersecurity Spending: Vol. V
CHRO Survey	July 9, 2025	US Application Software: 2025 CHRO Survey: Slow but Steady - HR Budgets Resilient in the Face of Macro Noise
Automation Survey	June 18, 2026	Global Industrials: 7th Proprietary Automation Survey Signals Encouraging Demand Outlook

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Citi Research

Companies Mentioned:

Advanced Micro Devices (AMD.O; US$620.68; 1; 08 Oct 26; 16:00) | Applied Materials Inc (AMAT.O; US$509.57; 1; 08 Oct 26; 16:00) | ITG Inc (ITG.O; US$7.62; 1H; 08 Oct 26; 16:00) | ON Semiconductor (ON.O; US$79.41; 2; 08 Oct 26; 16:00) | Space Exploration Technologies (SPCX.O; US$160.57; 1; 08 Oct 26; 16:00) | Texas Instruments Inc (TXN.O; US$288.2; 1; 08 Oct 26; 16:00) | TSMC (2330.TW; NT$2550.0; 1; 08 Oct 26; 13:30)