18 Sep 2026 15:34:26 ET | 29 pages

Artificial Intelligence

Citi's Inference Ahead - Where Intelligence Compounds

Heath Terry[AC]
Shelby Spencer
Ashley Kim
Janna Withrow

CITI'S TAKE

Early evidence that RSI may scale, including Dream-RSI from Google (9/14, Figure 2), points to a new source of compute demand. Lower average realized cost-to-serve (down 54% from its May 28 peak, 43% since July 1, and 2% WoW) may provide another tailwind to internal and external usage growth (Figure 19). The pacing debate is shifting from release timing to compute allocation. Meta delayed Muse for several months (X, 9/15) to prioritize safety, while Alexandr Wang stated most of Meta's compute is dedicated to serving people rather than racing on RSI (X, 9/15). Refocusing compute capacity from developing performance to safety, auditability, and alignment is unlikely to impact projections of $1T in 2026 and $1.6T in 2027 for global CapEx related to AI demand. Jensen Huang supported product-level pauses but believes new AI regulation is unnecessary (Salesforce, 9/15), reflecting a broader tension between pacing frontier development and defending competitive positioning.

RSI. Recursive self-improvement will likely originate in the infrastructure that models use to develop their successors. Z.ai's Infra Agent (9/17) moved GLM-5.3-Flash to production in under 2 weeks, tripled throughput across 100K+ accelerators, and helped process 62T+ tokens in 6 days. Though Infra Agent is still far from RSI, autonomous optimization should accelerate model development while increasing demand for inference and governance infrastructure. Z.ai delayed releasing model weights by 2 weeks due to GLM-5.3's advanced cybersecurity capabilities (Hugging Face, 8/28), while the AI Infra Summit further indicated that allocating more compute to safety and alignment may increase overall compute requirements rather than reduce them. On the deployment side, an Unsloth quantization of GLM-5.3 (8/29) retains ~81% accuracy while being 83% smaller, reducing on-prem memory and accelerator requirements. Lower deployment barriers could expand on-prem AI adoption, improve control over sensitive data, and lay the groundwork for RSI at scale, extending the MoM adoption gains Ramp recorded in 6 of 7 sectors (Figure 12).

Cadence. Only 1 proprietary and 3 open models launched this week, down from 5 and 6 WoW and from 3 and 6 four weeks ago, respectively (Figure 7). Trailing-90-day totals reached 68 proprietary and 40 open models, with proprietary releases up 31% MoM following the early-September surge and open releases flat (Figure 1). Across the broader model timeline, ~69% of releases since August 1 established a new intelligence high for their provider (Figure 8). Unconfirmed speculation of a model release from a frontier provider near month-end (CellCog, 9/18) could test the extent to which a capability gain can redirect premium usage despite the growing adoption of model routing.

Harnesses. Harness design is becoming a core driver of realized intelligence. On the interactive ARC-AGI-3 benchmark, adding a provider adapter harness to the same model raised performance 37 points to 99.95% from 62.71% and reduced cost 28% to $18.8K from $26.1K (Figure 5). Similarly, Salesforce found that evolving its agent harness increased average success across 7 enterprise-agent tasks from 29.2% to 78.0% (9/8) without any change to the underlying model. A growing share of realized intelligence gains may come from harnesses, reinforcing the application layer as a key source of differentiation for providers and enterprises.

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	9/16/26	Week	Point-in-Time	— 0.0%	53	53	66	63	63	63	
Highest Intelligence for Open Models	Points	9/16/26	Week	Point-in-Time	— 0.0%	44	44	60	60	60	60	
Gap in Proprietary and Open Intelligence	Points	9/16/26	Week	Point-in-Time	— 0.0%	9	9	6	3	3	3	
Highest Intelligence Models from the Top-20 Providers (Artificial Analysis)												
Median Intelligence	Points	9/16/26	Week	Point-in-Time	— 0.0%	33	33	44	44	44	44	
Median Latency	Sec	9/16/26	Week	Point-in-Time	▲ +3.0%	2.75	2.67	2.62	2.70	2.64	2.44	
Median Speed	Tok/Sec	9/16/26	Week	Point-in-Time	▲ +3.8%	108	104	84	74	76	74	
Release Cadence (Artificial Analysis, LLMTimeline, Model Provider Websites)												
Proprietary Models Released in Last Week	Models	9/16/26	Week	Weekly Total	▼ -80.0%	1	5	12	2	3	6	
Open Models Released in Last Week	Models	9/16/26	Week	Weekly Total	▼ -50.0%	3	6	0	6	6	8	
Proprietary Models Released in Last 90 Days	Models	9/16/26	Week	Trailing 90d	— 0.0%	68	68	64	53	52	53	
Open Models Released in Last 90 Days	Models	9/16/26	Week	Trailing 90d	— 0.0%	40	40	36	42	40	35	
Index Pricing (Bloomberg)												
A100 (SDA100RT)	$/Hour	9/16/26	Week	7d Avg	▼ -1.3%	1.58	1.60	1.61	1.61	1.63	1.65	
H100/H200 (CIHOPUS)	$/Hour	9/16/26	Week	7d Avg	— 0.0%	2.97	2.97	2.96	2.89	2.88	2.91	
(G)B200/(G)B300 (CIBLKWUS)	$/Hour	9/16/26	Week	7d Avg	— 0.0%	5.30	5.30	5.27	5.27	5.31	5.27	
Avg Inference Token Expenditure (SDLLMTK)	$/1M Tok	9/16/26	Week	7d Avg	▼ -2.1%	0.95	0.97	0.98	1.02	1.06	1.04	
Model Pricing for Blended Tokens (Artificial Analysis)												
US+EU Avg Price	$/1M Tok	9/16/26	Week	Point-in-Time	— 0.0%	2.30	2.30	1.80	1.42	1.56	1.63	
China Avg Price	$/1M Tok	9/16/26	Week	Point-in-Time	▲ +13.7%	0.83	0.73	0.91	0.91	0.91	0.82	
Frontier Avg Price	$/1M Tok	9/16/26	Week	Point-in-Time	▲ +2.4%	1.71	1.67	1.44	1.22	1.30	1.31	
Proprietary Avg Price	$/1M Tok	9/16/26	Week	Point-in-Time	▼ -8.8%	2.89	3.17	2.23	1.75	1.94	2.02	
Open Avg Price	$/1M Tok	9/16/26	Week	Point-in-Time	▲ +1.5%	0.68	0.67	0.75	0.75	0.75	0.68	
Cost per Intelligence Index Task (Artificial Analysis)												
Proprietary Avg Cost	$/Task	9/16/26	Week	Point-in-Time	▲ +71.7%	2.06	1.20	0.42	0.38	0.36	0.54	
Open Avg Cost	$/Task	9/16/26	Week	Point-in-Time	▲ +53.1%	1.24	0.81	0.35	0.35	0.38	0.27	
Model Cost Composition per Intelligence Index Task (Artificial Analysis)												
Input Share of Task Cost	%	9/16/26	Week	Point-in-Time	▲ +4.3	6.3	2.0	1.8	1.8	1.7	1.9	
Cache Hit Share of Task Cost	%	9/16/26	Week	Point-in-Time	▼ -11.5	36.2	47.7	35.0	34.7	32.0	31.4	
Cache Write Share of Task Cost	%	9/16/26	Week	Point-in-Time	▼ -3.1	23.9	27.0	31.1	30.5	34.6	33.1	
Reasoning Share of Task Cost	%	9/16/26	Week	Point-in-Time	▲ +2.4	17.2	14.8	22.5	22.7	21.8	22.9	
Answer Share of Task Cost	%	9/16/26	Week	Point-in-Time	▲ +7.8	16.3	8.5	9.7	10.3	9.9	10.8	
Avg Output Tokens per Task	Tokens	9/16/26	Week	Point-in-Time	▼ -12.8%	37,269	42,716	26,029	25,350	24,863	24,908	
Developer Usage (PyPI)												
Multi-Provider SDK Downloads	%	9/9/26	Week	WoW Δ in 7d Avg	▼ -0.9	+4.1	+5.0	-37.4	-9.6	+0.7	+1.3	
Gemini SDK Downloads	%	9/9/26	Week	WoW Δ in 7d Avg	▼ -6.6	+1.6	+8.2	-52.6	-12.8	+0.5	+3.7	
Mistral SDK Downloads	%	9/9/26	Week	WoW Δ in 7d Avg	▼ -38.7	-11.8	+26.9	-48.9	-32.2	-22.1	-1.4	
App Usage (Sensor Tower)												
Gemini App WAU	%	9/9/26	Week	WoW Δ	▼ -1.9	+0.7	+2.6	+3.4	+2.0	+3.3	+1.5	
Qwen App WAU	%	9/9/26	Week	WoW Δ	▲ +0.5	-0.5	-1.0	+6.8	-4.4	+8.3	-1.0	
Kimi App WAU	%	9/9/26	Week	WoW Δ	— 0.0	-1.8	-1.8	-3.5	-2.8	-3.6	-6.1	
Web Usage (Similarweb)												
Gemini Web Visits	%	8/26/26	Week	WoW Δ in 7d Avg	▼ -3.5	-0.1	+3.4	+1.3	-1.6	-2.6	+0.6	
Qwen Web Visits	%	8/26/26	Week	WoW Δ in 7d Avg	▲ +2.2	+4.2	+2.0	+0.3	+4.9	+1.7	-3.7	
Kimi Web Visits	%	8/26/26	Week	WoW Δ in 7d Avg	▲ +18.5	-10.0	-28.5	-10.0	-11.9	-10.8	-1.5	
Layoffs (WSJ)												
All Layoffs Announced	People	9/10/26	Month	MTD	▲ +124.6%	3,955	1,761	17,420	24,709	26,160	12,643	
AI-Driven Layoffs Announced	People	9/1/26	Month	MTD	— +0.3%	322	321	3,220	21,640	15,360	10,400	

© 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
Paper	Google	September 14, 2026	Dream-RSI: Recursive Self-Improvement through Evolving Worlds
Podcast	Odd Lots	September 14, 2026	OpenAI President Greg Brockman on Doing Business in the Wake of Hugging Face
Article	Nvidia	September 10, 2026	How Full-Stack NIM Optimizations Deliver 2.5x More Users on Nemotron 3 Ultra
Podcast	Dwarkesh Podcast	September 11, 2026	AI researchers debate how close we are to recursive self-improvement
Podcast	Training Data	September 15, 2026	Box's Aaron Levie: On Reinventing Yourself in the AI Age and Enterprise Diffusion
Paper	Cunningham et al.	September 14, 2026	The Economics of Recursive Self-Improvement
Podcast	Latent Space	September 14, 2026	Humanity's Last Invention — Richard Socher of Recursive
Paper	U.S. Census Bureau	September 10, 2026	Graduating into Disruption: Labor Market Outcomes for AI-Exposed College Majors
Paper	Yale, Jump Trading, University of Cambridge, USC	September 11, 2026	Expert Re-Grading Reveals Broken Evaluations and Near-Saturation of Leading Benchmarks
Podcast	Pioneers of AI	September 16, 2026	The ROI on AI is obvious in these industries
Podcast	Machine Learning Street Talk	September 15, 2026	How Physical AI Learns Across Language, Video and Action — Ming-Yu Liu
Podcast	The a12z Show	September 16, 2026	The AI-Native CRM
Podcast	Odd Lots	September 12, 2026	Robert Friedland on the World's Monumental Shortage of Copper
Paper	Theseus Lab	September 10, 2026	The Last AI Built by Humans: Toward Genuine Recursive Self-Improvement
Paper	Wu et al.	September 14, 2026	ModularRSI: Modular and Generalizable Recursive Harness Self-Improvement
Paper	Aether AI	September 14, 2026	RSIAgent: Autonomous Exploration for Recursive Self-improvement in New Environments
Article	TypeSafe AI	September 15, 2026	Introducing System One Model & Jev
Podcast	Build Mode (TechCrunch)	September 10, 2026	How Mbodi is solving robotics' scaling problem, with Xavier Chi
Interview	Justified Posteriors	September 14, 2026	Daron Acemoglu on pro-worker AI and vibes based capital

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

Source: Citi Research

Figure 3. Events Calendar

Start Date	End Date	Event	Type	Location
September 22, 2026	September 24, 2026	HumanX	Conference	Amsterdam, The Netherlands
September 23, 2026	September 23, 2026	TSMC North America OIP Ecosystem Forum	Conference	Santa Clara, CA
September 23, 2026	September 24, 2026	Meta Connect	Conference	Menlo Park, California
September 24, 2026	September 24, 2026	U.S. State Dinner for President Xi Jinping	Dinner	Washington, D.C.
September 29, 2026	September 29, 2026	PJM Reliability Backstop Procurement Deadline	Deadline	Valley Forge, PA
September 29, 2026	September 29, 2026	OpenAI DevDay	Conference	San Francisco, CA
September 29, 2026	October 1, 2026	The AI Conference	Conference	San Francisco, CA
October 5, 2026	October 11, 2026	a16z Tech Week	Conference	San Francisco, CA
October 6, 2026	October 9, 2026	COLM	Conference	San Francisco, CA
October 7, 2026	October 8, 2026	World Summit AI	Conference	Amsterdam, The Netherlands
October 8, 2026	October 8, 2026	Science: A New Golden Age	Conference	Washington, D.C.
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 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

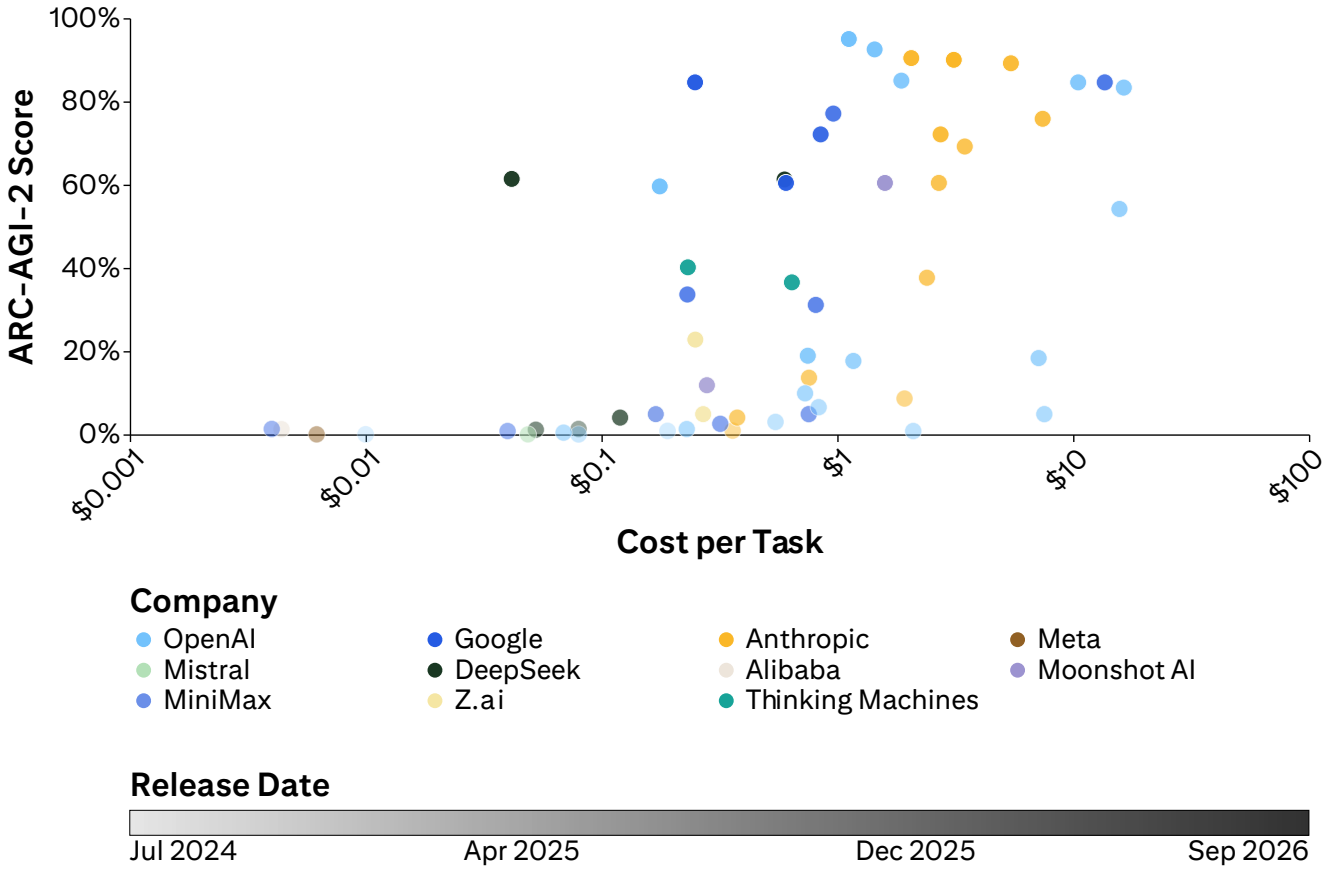
Figure 4. Model Metrics

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
—	**Anthropic**	Claude Fable 5.1 (max with fallback)	53	68	$7.17	1000	Sep. 1, 2026	Proprietary
—	**OpenAI**	GPT-6 Astra (max)	53	53	$7.70	1000	Sep. 3, 2026	Proprietary
▲4	**Meta**	Muse Spark 1.3 (max)	48	229	$0.78	1000	Sep. 2, 2026	Proprietary
▲2	**Alibaba**	Qwen3.8 Max (0902)	45	40	$1.18	984	Sep. 16, 2026	Proprietary
—	**Z.ai**	GLM-5.3 (max)	45	67	$0.90	1000	Aug. 18, 2026	Open
▼2	**Moonshot AI**	Kimi K3 (max)	44	36	$2.31	1050	Jul. 16, 2026	Open
▼4	**SpaceXAI**	Grok 4.6 (high)	44	62	$1.35	500	Aug. 12, 2026	Proprietary
—	**Google**	Gemini 3.8 Flash (high)	41	341	$0.57	1000	Sep. 2, 2026	Proprietary
—	**DeepSeek**	DeepSeek V4.1 Flash (max)	40	216	$0.19	1000	Sep. 14, 2026	Open
▲7	**Sapiens AI**	Agnes 3.0 Flash	36	242	$0.03	1000	Sep. 14, 2026	Proprietary
▼1	**MiniMax**	MiniMax-M3	30	130	$0.22	1000	Jun. 1, 2026	Open
—	**Nex AGI**	Nex-N2-Pro	28	116	$0.53	262	Jun. 2, 2026	Open
▲2	**Upstage**	Solar Pro 4	28	71	$0.22	512	Aug. 6, 2026	Proprietary
▲7	**Multiverse Computing**	Quasar 438B (max)	27	147	$0.72	1000	Aug. 10, 2026	Proprietary
▼2	**Tencent**	Hy3	26	86	$0.11	256	Jul. 6, 2026	Open
▼2	**Thinking Machines**	Inkling Small	26	170	$0.22	1000	Jul. 30, 2026	Open
▼6	**Xiaomi**	MiMo-V2.5-Pro	26	52	$0.17	1000	Apr. 22, 2026	Open
▼2	**InclusionAI**	Ling-3.0-flash-VL	25	144	$0.00	262	Sep. 14, 2026	Open
▼2	**NVIDIA**	Nemotron 3 Ultra	23	173	$0.50	262	Jun. 1, 2026	Open
▼2	**KwaiKAT**	KAT-Coder-Pro V2	22	102	$0.22	256	Mar. 27, 2026	Proprietary

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

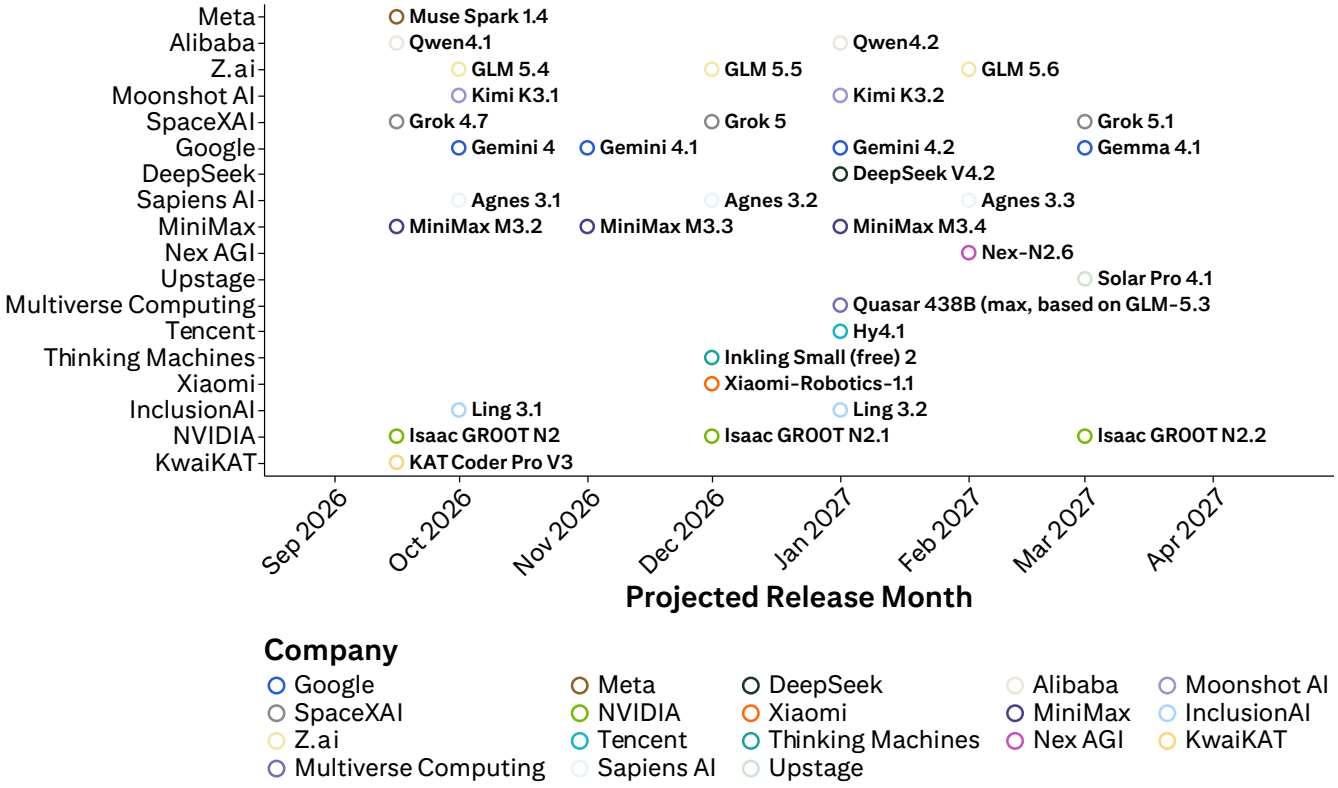
Source: Citi Research, Artificial Analysis, Model Provider Websites

Figure 5. Highest ARC Score per Company by Date



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

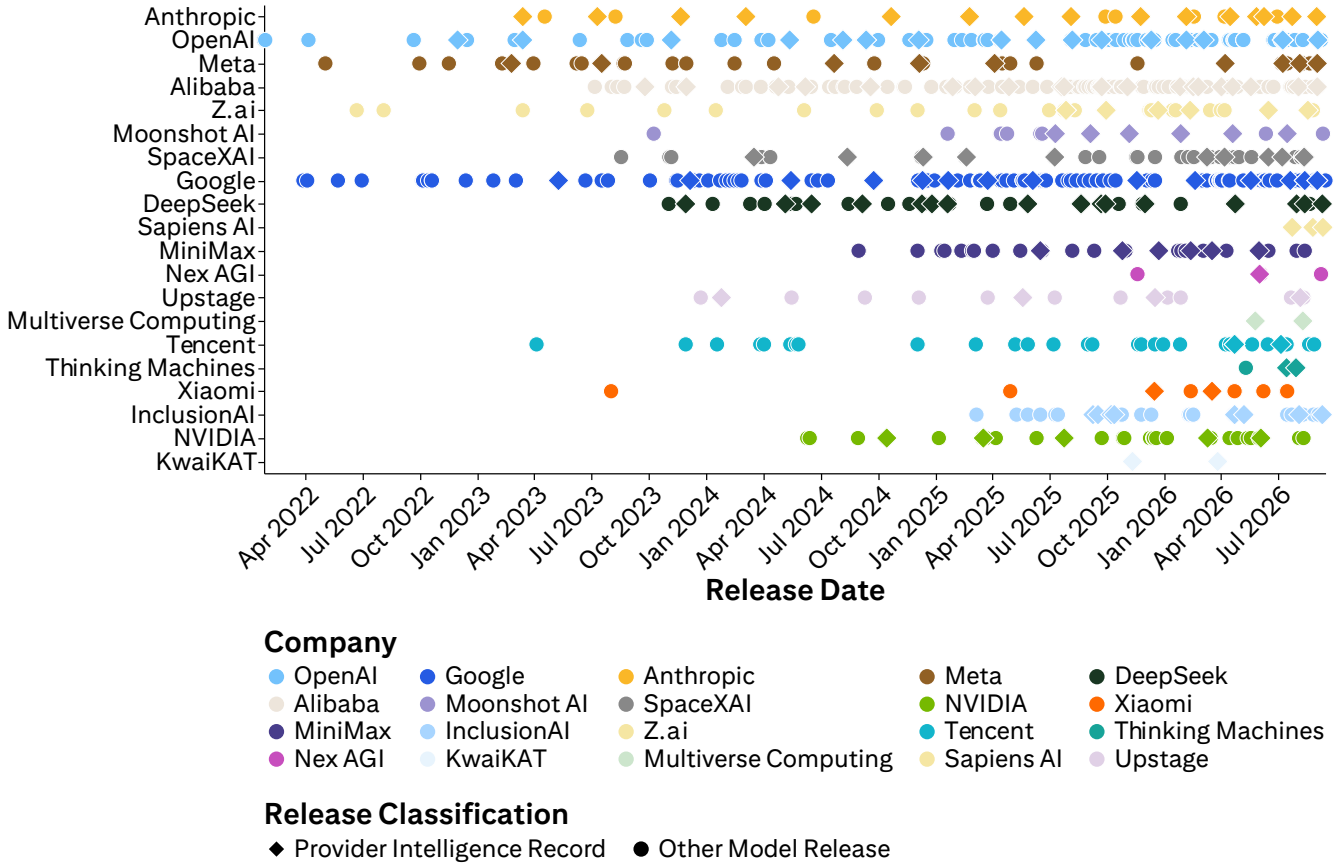
Figure 6. Projected Model Releases



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

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

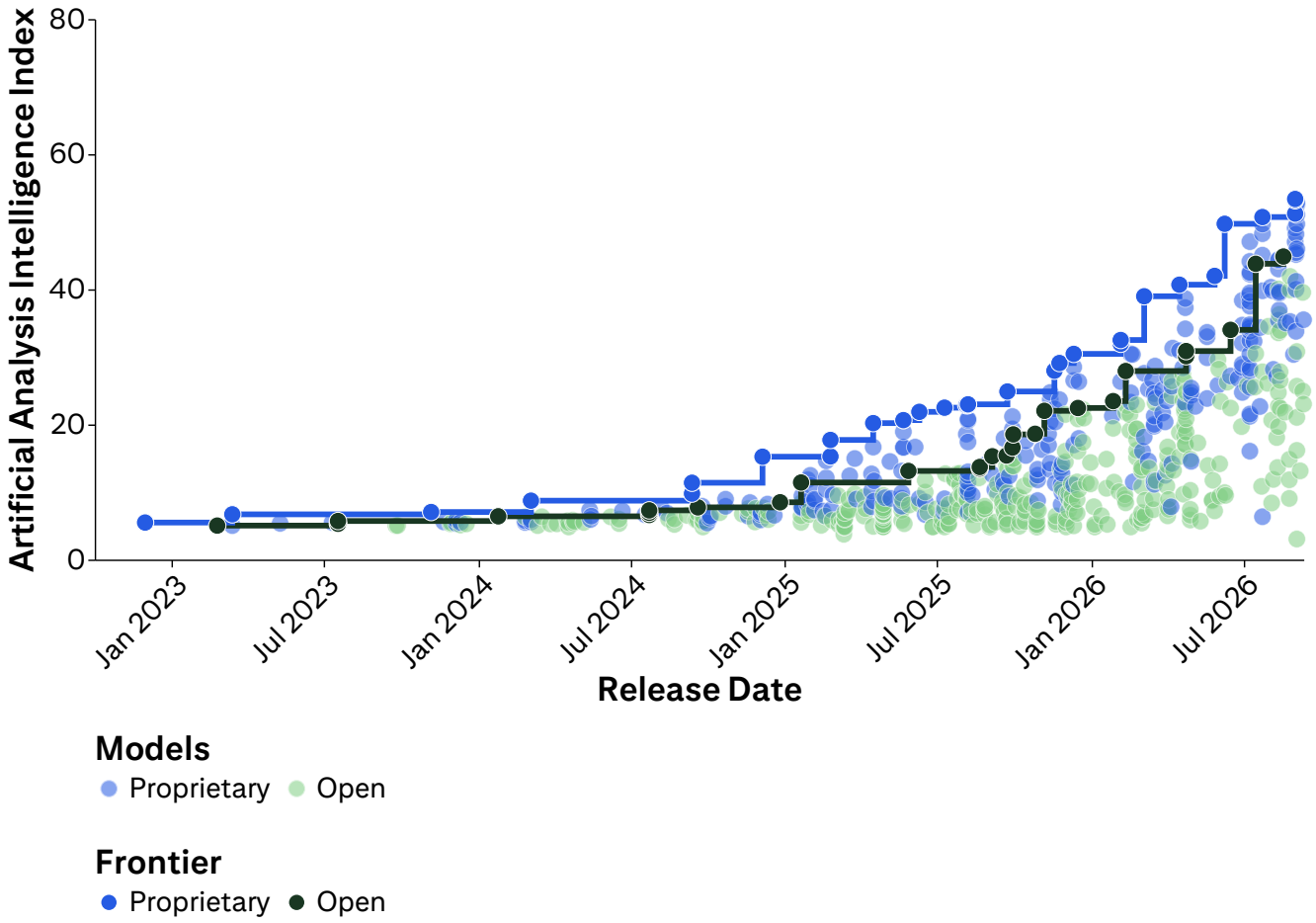
Figure 7. Historical Model Releases



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

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

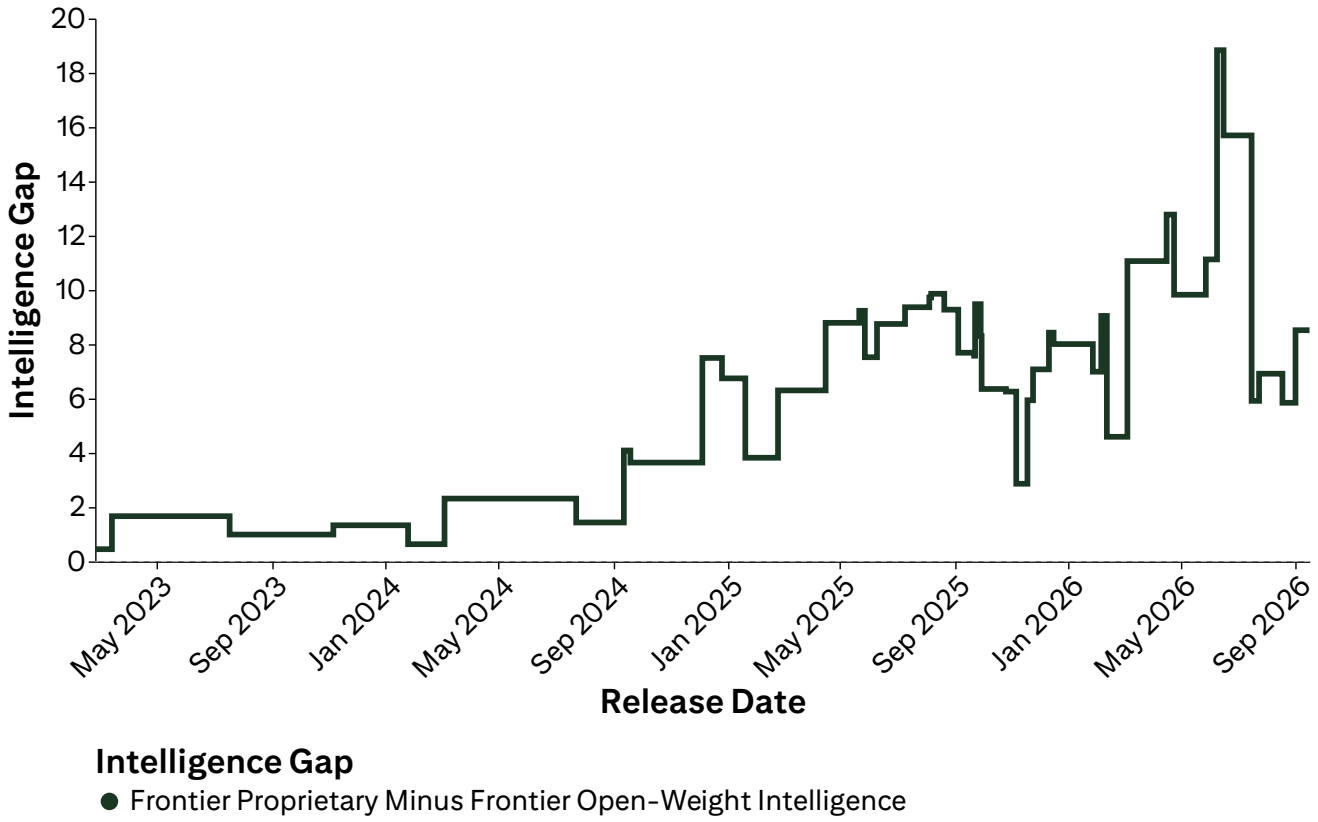
Figure 8. Frontier Releases for Proprietary and Open Models



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

Source: Citi Research, Artificial Analysis

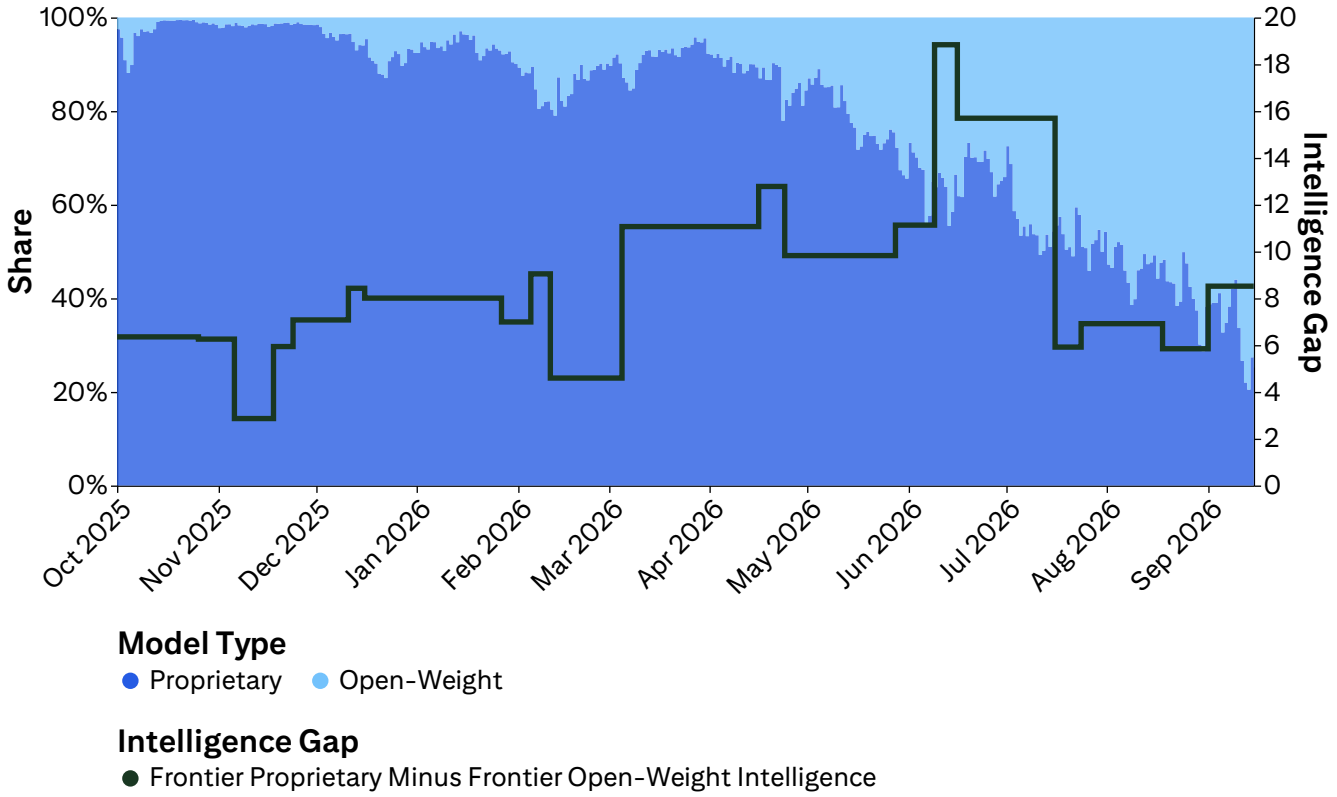
Figure 9. Gap in Proprietary and Open Intelligence



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

Source: Citi Research, Artificial Analysis

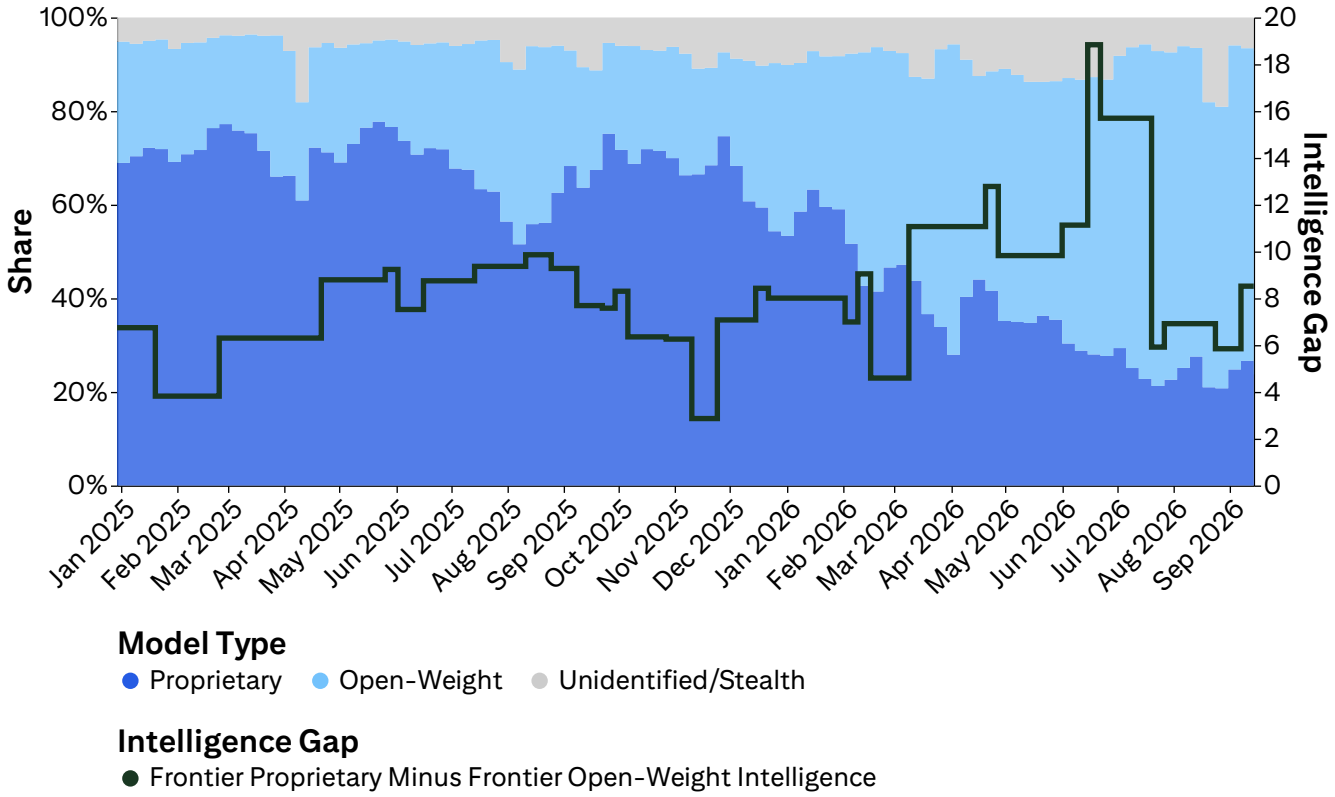
Figure 10. Vercel Token Volume Shares



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

Source: Citi Research, Vercel

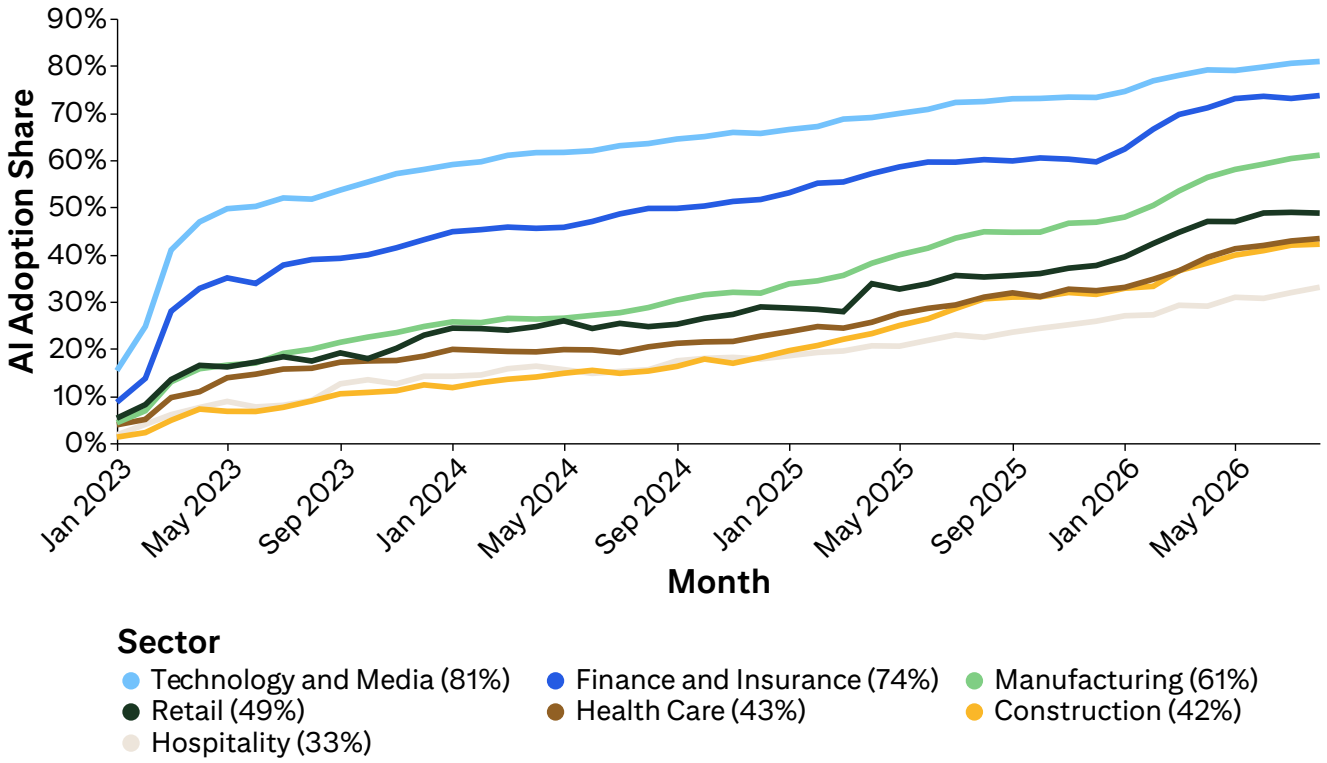
Figure 11. OpenRouter Token Volume Shares



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

Source: Citi Research, OpenRouter

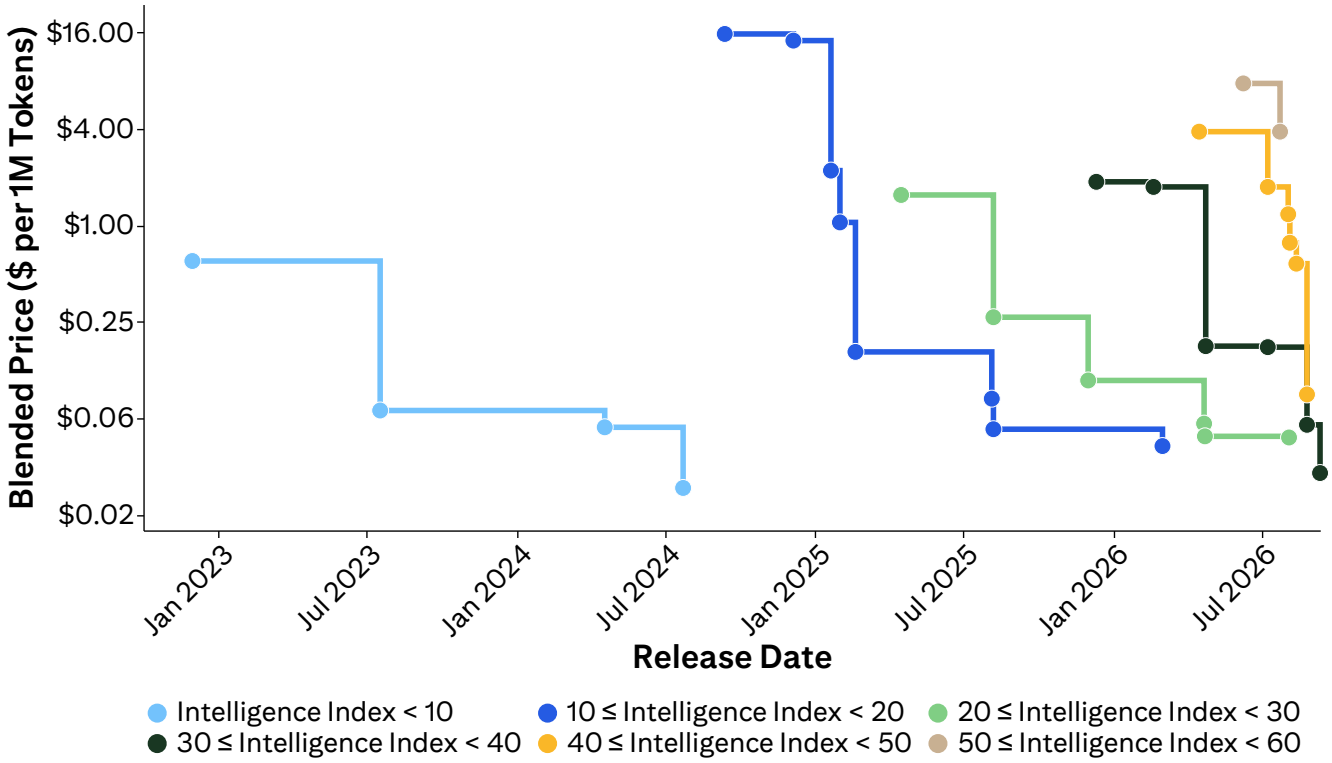
Figure 12. Sector Adoption Rates



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

Source: Citi Research, Ramp

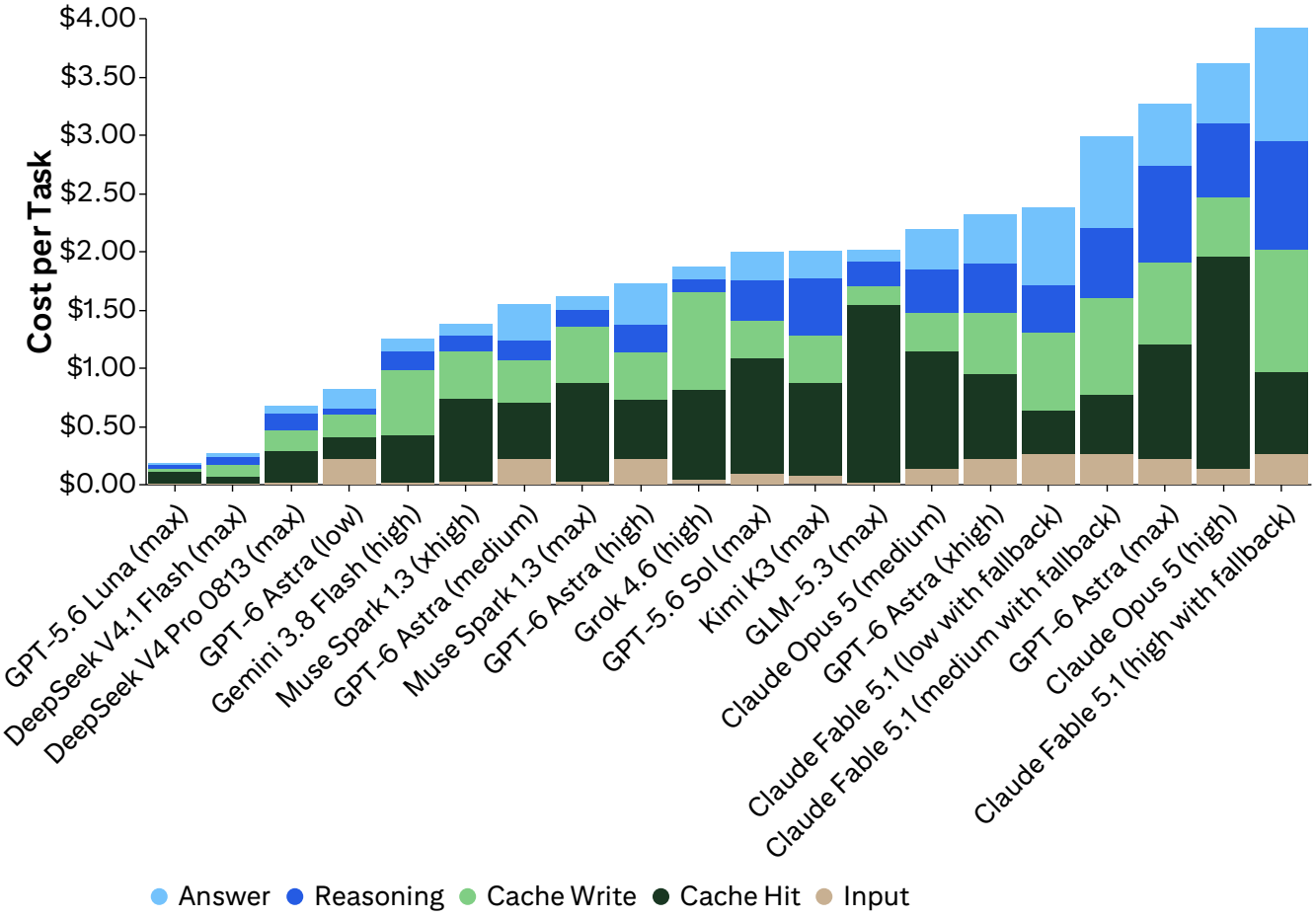
Figure 13. Intelligence Over Time v. Blended Price



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

Source: Citi Research, Artificial Analysis

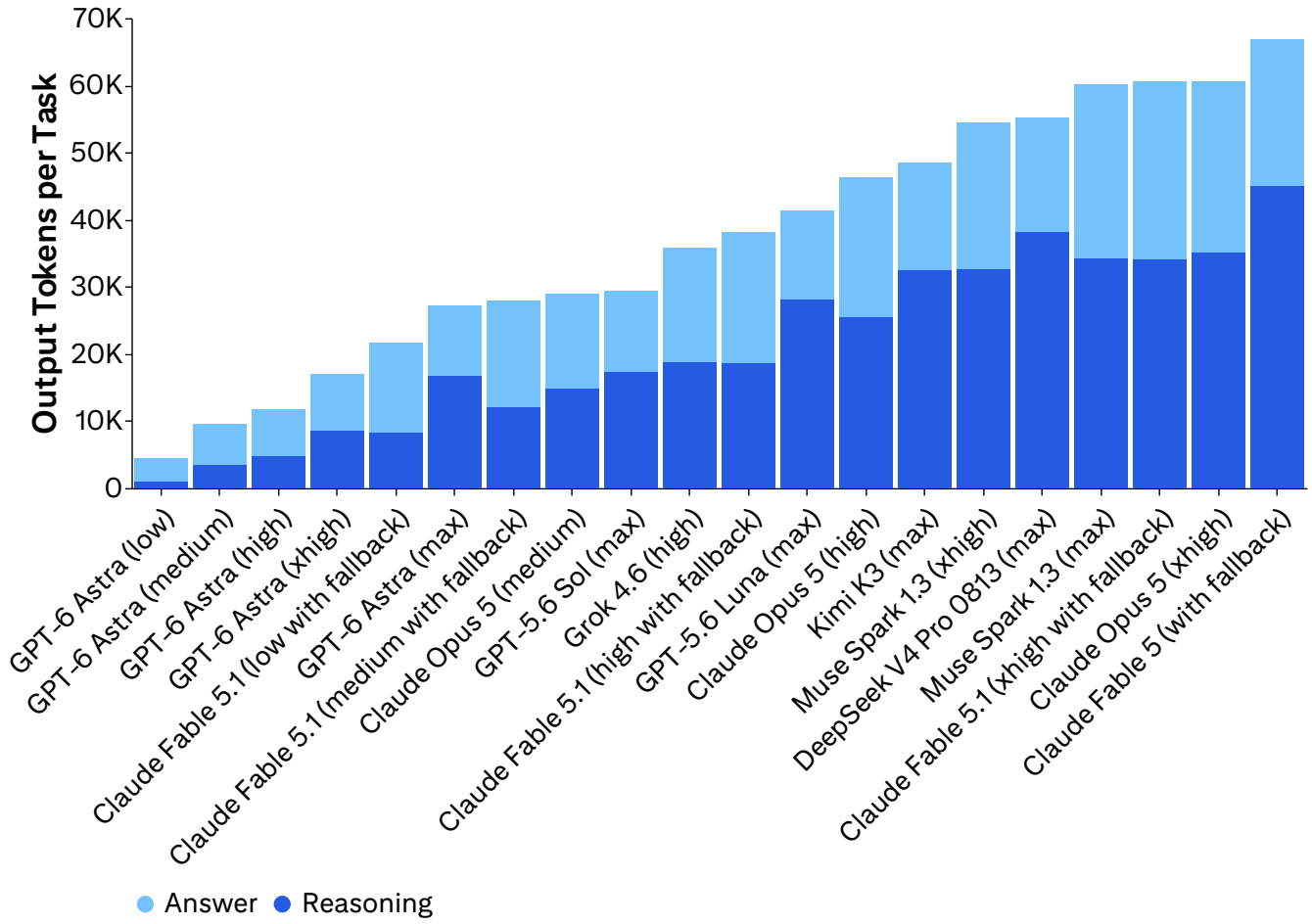
Figure 14. Cost Breakdown per AA Task



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

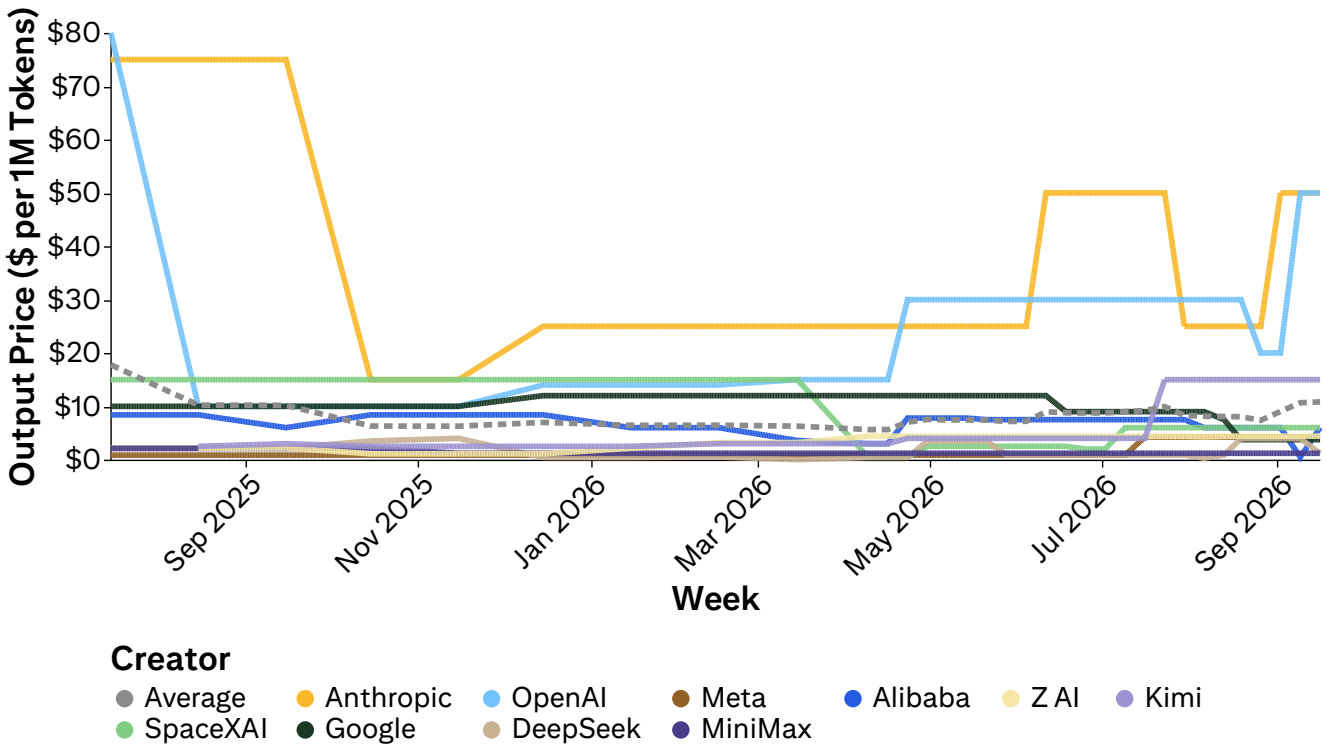
Source: Citi Research, Artificial Analysis

Figure 15. Output Tokens per AA Task



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

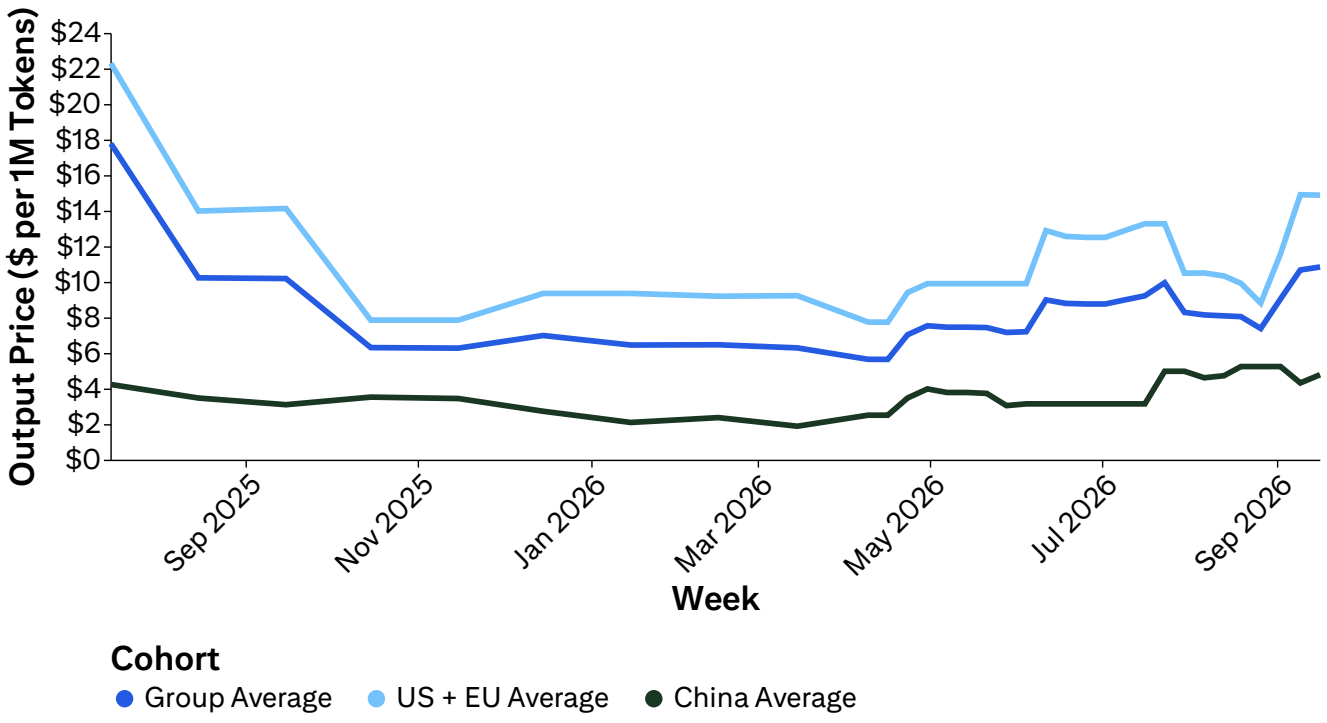
Figure 16. Output Token Pricing by Creator



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

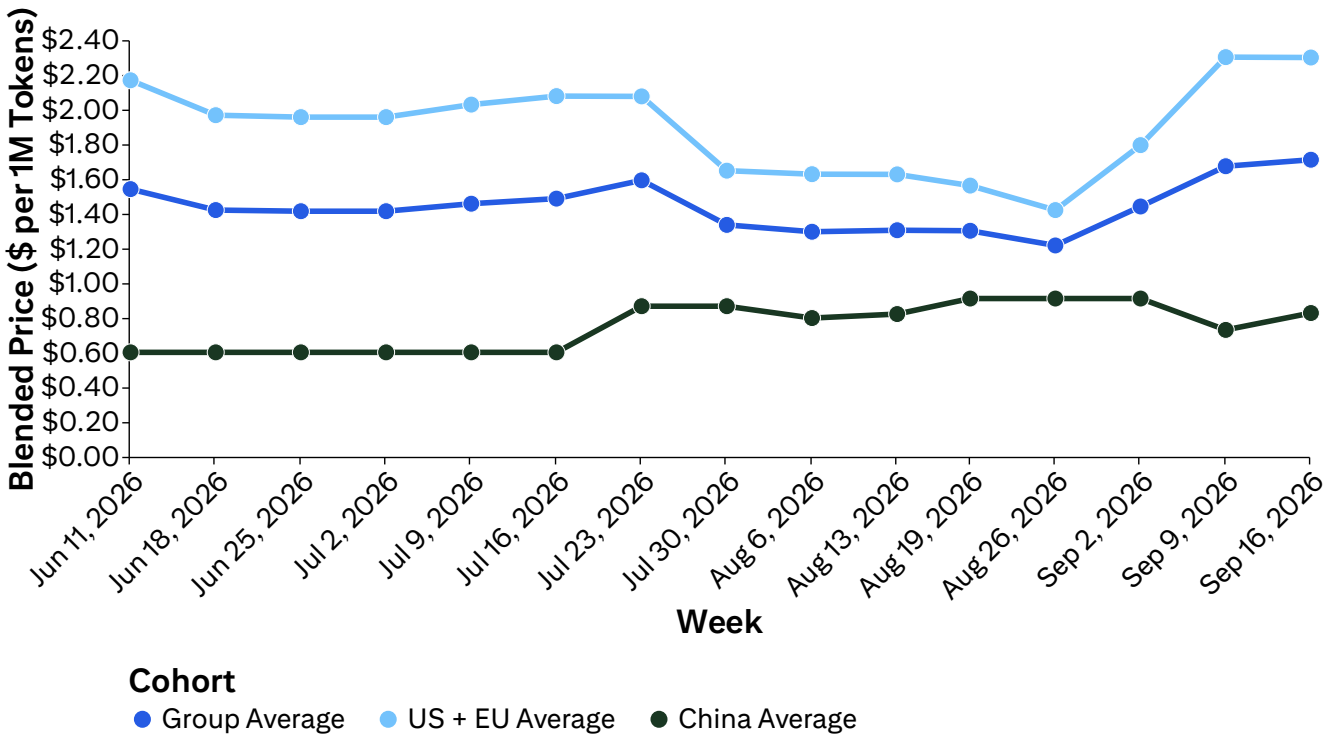
Source: Citi Research, Artificial Analysis

Figure 17. Output Token Pricing



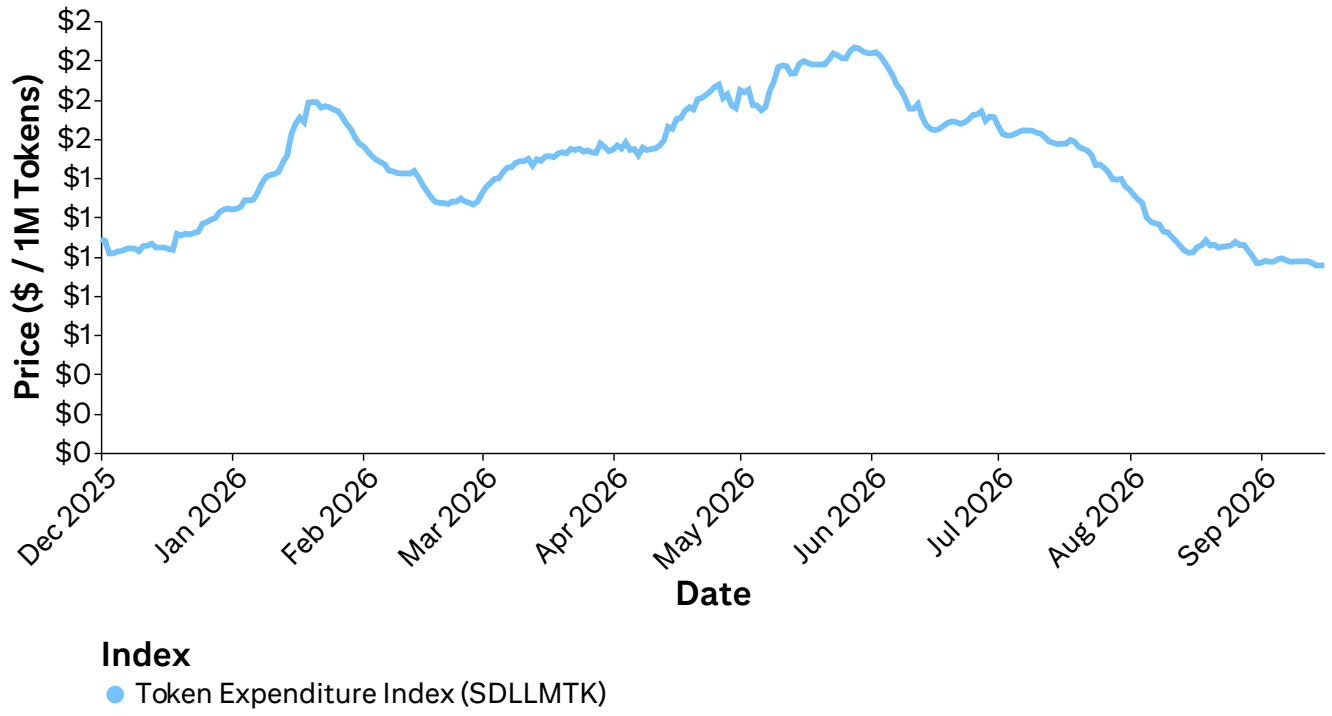
© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Citi Research, Artificial Analysis

Figure 18. Blended Token Pricing



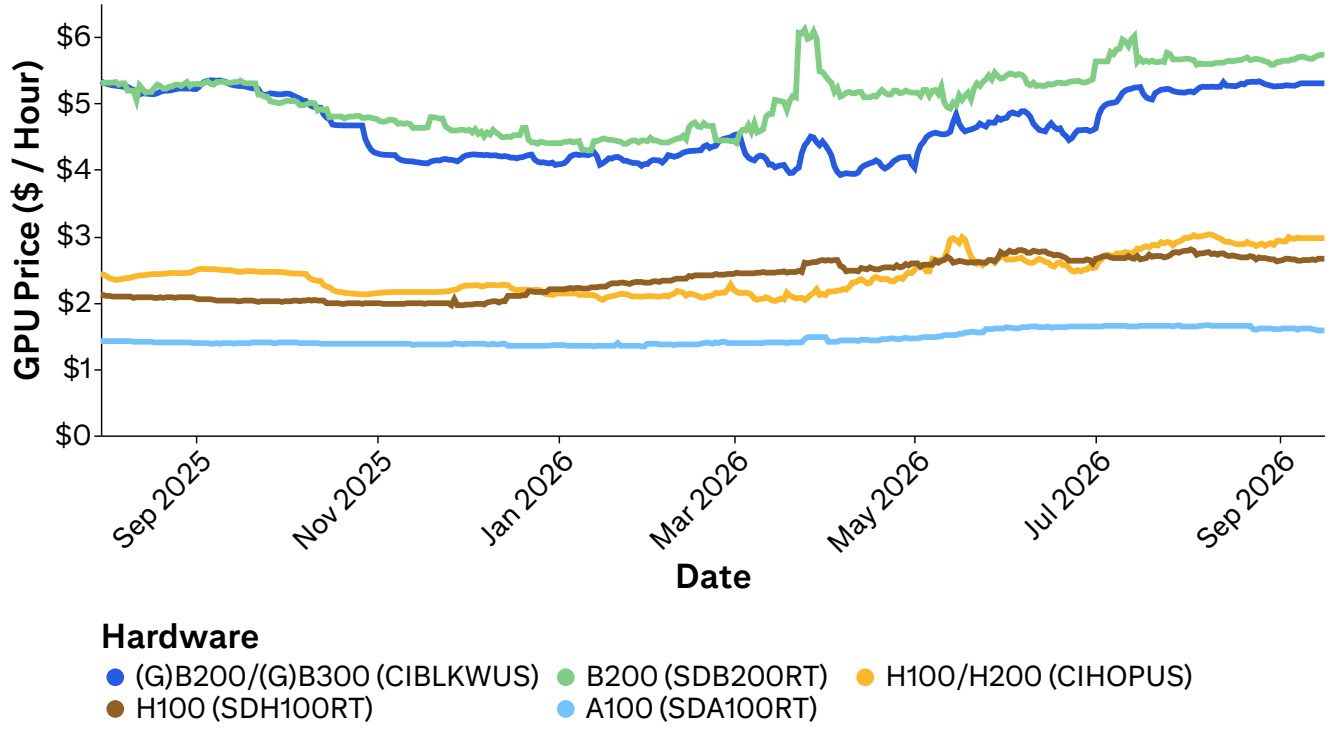
© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Citi Research, Artificial Analysis

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
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	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	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	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