

TECHNOLOGY | 4 MIN READ | SEPTEMBER 2026

Tokenomics: The Hidden Economics of AI Adoption

Navigating AI's value challenge by prioritizing spending



Kieran White

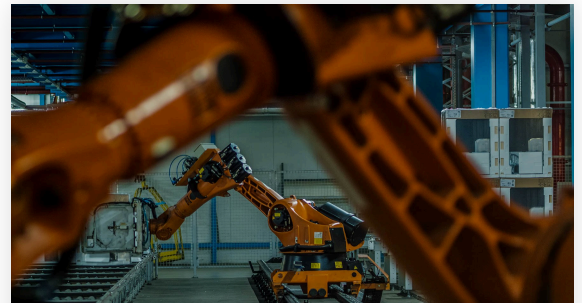
International Head, AI Center of Excellence

- Massive spending: AI investment surge, yet only 13% of organizations report significant enterprise-level value from their GenAI investments and 40% of projects are failing to meet goals.
- Hidden costs: Tokens drive is non-linear, exposing organizations to a pricing cliff.
- Discipline beats speed: Success shifts from deployment velocity to cost governance—rightsizing models, on-premises inference, and treating tokens as managed capital.

Challenges amid transformation

Artificial intelligence (AI) is transforming industries. From banking to healthcare, companies globally are investing

Suggested views



Jump to all insights on
Technology →



Beyond the Headlines: AI Has Added Jobs in Asia – So Far

TECHNOLOGY | 5 MIN READ
AUGUST 2026

billions to integrate AI across their operations, seeking to boost efficiency and gain insights that create competitive advantages.

Yet the results haven't been linear, underscoring the need to prioritize spending on opportunities that can deliver the highest returns.

The following highlights the challenges companies face.

AI is seeing massive investment as illustrated below:

- In 2025, 80% of banks increased their AI budgets, with 41% of C-level executives expecting AI to exceed \$25million^{1, 2}.
- Global corporate AI investment surged to \$582 billion, with generative AI alone reaching \$110 billion in trailing 12-month revenue—growing at triple the pace of prior technology waves^{3, 4}.

However, companies are struggling to immediately achieve the planned results:

- Only 13% of organizations report significant enterprise-level value from their GenAI investments.
- 40% of implemented projects fail to meet planned results^{5, 6}.

Clearly the gap between investment and value is widening—and a significant reason lies in tokenomics, a strategic enabler for sustainable AI adoption.

What are tokens?

Tokens are the bricks of AI—small, discrete units that form the foundation of every interaction. Just as a building isn't constructed from a single slab, language models process text by breaking it into tokens, each representing roughly three-quarters of a word.

These tokens are more than technical abstractions; they are the economic currency of AI, behaving like volatile



AI's Labor Market Impact: Much Noise, Little Disruption

ECONOMICS | 3 MIN READ
JULY 2026



Quantifying Asia's AI Exposure

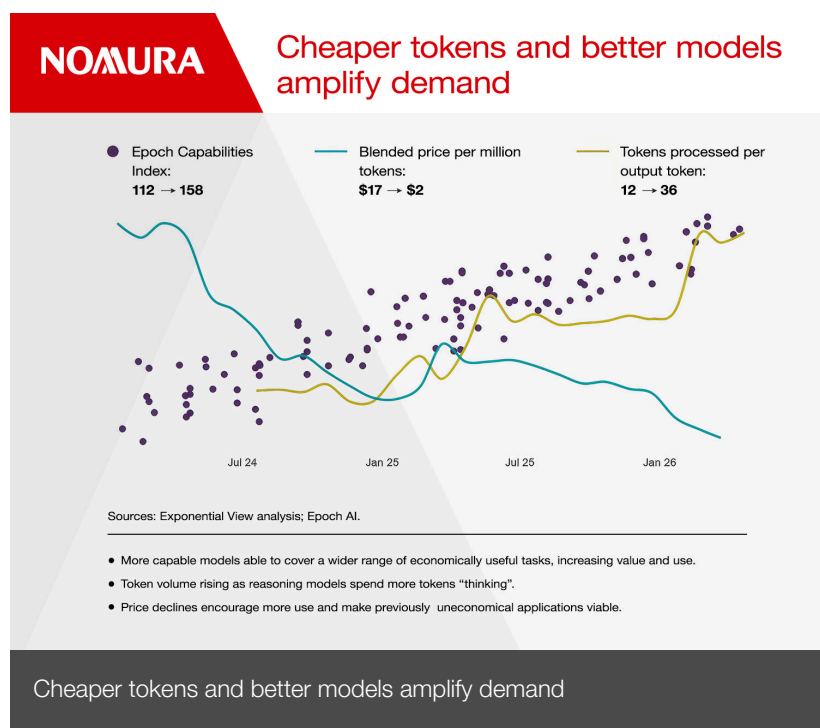
ECONOMICS | 2 MIN READ
FEBRUARY 2026

input costs rather than predictable subscriptions^{7, 8}.

Efficiency in token usage mirrors construction: just as architects optimize brick counts to balance cost and design, organizations must refine prompts and instructions to maximize value. The shift from possessing intelligence to applying it efficiently has turned tokens into a competitive battleground—a resource that demands strategic allocation, much like capital or talent⁹.

The race to the bottom—and what comes next

The AI pricing war of 2023–2025 compressed token costs dramatically. Research shows every 10% price reduction yields 12–18% more tokens consumed, with total spend still rising⁴. Enterprise AI budgets grew ~320% between 2024 and 2026¹⁰.

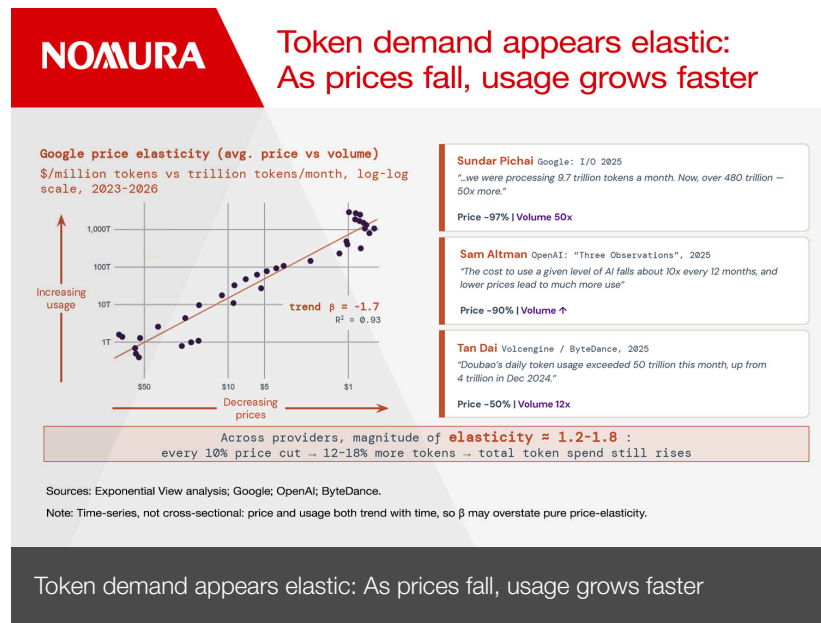


But this compression was deliberate market strategy—not equivalent reductions in compute costs. We are already seeing the correction:

- A leading e-commerce player removed token leaderboards.
- A major tech firm cancelled an AI code subscription.

- Organizations questioning frontier model economics at scale¹¹.

The dynamics are non-linear: tokens consumed per task have surged even as unit prices fell, keeping effective cost per task often flat or rising¹².



The hidden cost problem

Context inflation

A single enterprise query can consume over 6,000 tokens in system prompts and retrieved documents before the user's question is even processed—over 85% of the cost remains invisible to the end user^{8, 13}.

Agentic multiplication

Agentic AI—systems chaining multiple autonomous steps—compounds this dramatically⁷. As organizations move from chatbots to autonomous workflows, consumption grows exponentially.

The pricing cliff

Frontier deployment is increasingly constrained by cost, capacity, and marginal returns. The shift is from "what models can do" to the "the price and scarcity of inputs"¹¹. Organizations built on artificially low prices will face a significant adjustment.

What does this mean for the adoption race

Controlling token economics

Organizations should manage tokens like capital—tracking return on intelligence for each AI project and allocating spend to opportunities with the highest returns⁹. Token governance should match the governance of capital and revenue¹³.

Small models are a strategic hedge

Compact models can outperform larger frontier models on specialized tasks like mathematical reasoning. Running it on dedicated hardware costs ~\$50/day for 100M tokens versus ~\$1,560/day on a frontier API — a 32x difference¹⁰. The practical lever is rightsizing models to tasks rather than defaulting to the most capable option¹².

On-premises deployment is returning

Self-hosted inference offers potentially over 50% cost savings versus API approaches over three years¹⁴. For high-volume, predictable workloads, the economics are compelling—and provide greater control over cost predictability.

Conclusion

The AI adoption race is entering its second phase. The first rewarded velocity—deploying models at scale. The second will reward discipline:

- Who sustains investments without succumbing to cost inflation?
- Who demonstrates genuine ROI beyond pilot projects?
- Who navigates the non-linear economics of tokens, compute, and marginal returns?

Tokenomics is not a technical curiosity. It is the economic framework that will determine which

strategies survive. Organizations that treat tokens like capital, models like strategic hedges, and governance like a competitive advantage will outlast those chasing speed alone.

The future belongs to those who measure, optimize, and govern, not just those who spend.

Sources

1. *American Banker, 80% of Banks Have Upped AI Spending — Here's What They're Buying, 2025*
2. *BCG / Accenture, via Business Insider, How Wall Street Banks Can Find ROI on AI Investments, 2024*
3. *Stanford HAI, AI Index Report 2026, Chapter 4: Economy, 2026*
4. *Exponential View, The State of the AI Economy, 2025*
5. *Accenture, via American Banker, Why Banks Aren't Seeing High Returns on Generative AI, 2025*
6. *EY, Gen AI in Retail and Commercial Banking, 2025*
7. *Deloitte, CFO Guide to AI Token Economics, 2025*
8. *CIO.com, Tokenomics in Enterprise AI, 2026*
9. *BCG, The Era of Token-Based Competition Is Here, 2026*
10. *PracticalLogix, Small Language Models Eat the Edge: The 32x Cost Disruption Reshaping Enterprise AI, 2026*
11. *Citadel Securities, Tokenomics, 2025*
12. *Bain & Company, How Token Economics Will Change Opex, 2025*
13. *Flexera, Tokenomics: Economics of AI, 2026*
14. *Deloitte, AI Tokens: How to Navigate Spend Dynamics, 2025*

Tags In

Artificial Intelligence

Technology

Contributor



Kieran White

International Head, AI Center of Excellence

Disclaimer

This content has been prepared by Nomura solely for information purposes, and is not an offer to buy or sell or provide (as the case may be) or a solicitation of an offer to buy or sell or enter into any agreement with respect to any security, product, service (including but not limited to investment advisory services) or investment. The opinions expressed in the content do not constitute investment advice and independent advice should be sought where appropriate. The content contains general information only and does not take into account the individual objectives, financial situation or needs of a person. All information, opinions and estimates expressed in the content are current as of the date of publication, are subject to change without notice, and may become outdated over time. To the extent that any materials or investment services on or referred to in the content are construed to be regulated activities under the local laws of any jurisdiction and are made available to persons resident in such jurisdiction, they shall only be made available through appropriately licenced Nomura entities in that jurisdiction or otherwise through Nomura entities that are exempt from applicable licensing and regulatory requirements in that jurisdiction. For more information please go to <https://www.nomuraholdings.com/policy/terms.html>.

BACK TO TOP



Share this article



TWITTER



LINKEDIN



EMAIL

[nomura.com](https://www.nomura.com)

[nomuranow.com](https://www.nomuranow.com)

[Global Offices](#)

[Careers](#)

[Disclaimer](#)

[Cookies](#)

[Feedback](#)

[Privacy Policy](#)

[Accessibility Policy](#)

