

A futures market for AI computing power



Source: AI generated

When the world's largest asset manager predicts a futures market for computing power and an exchange launches one within a week, the signal seems clear: compute has become the AI era's scarce commodity. Yet the same industry straining against that shortage is engineering its way out of it, targeting cost reductions of up to two orders of magnitude.

New asset classes typically follow a pattern: a critical resource becomes scarce, demand rises, and financial markets build instruments to manage that scarcity, as with oil futures, then carbon credits. In 2026, the same dynamic is taking shape around AI computing power, as demand for AI infrastructure continues to outpace supply.

Charles-Henry Monchau, CFA, CAIA, CMT
Chief Investment Officer
charles-henry.monchau@syzgroup.com

Hugo Morel
Syz Research Lab Team
hugo.morel@syzgroup.com

Benjamin Roth
Syz Research Lab Team
benjamin.roth@syzgroup.com

1. The shortage is real, not just one CEO’s view

At the Milken Institute Global Conference in May, BlackRock CEO Larry Fink argued that demand for compute had grown so intense that “a new asset class will be buying futures of compute,” citing shortages in capacity, chips and memory across the US. He dismissed AI bubble concerns, pointing instead to supply shortages outpacing

demand. BlackRock is backing that view with capital, including a ~\$40 billion acquisition of Aligned Data Centers and a \$10.7 billion deal for power producer AES. Brookfield CEO Bruce Flatt echoed the scale of the shift, describing a decade-long “rewiring of the global economy.”

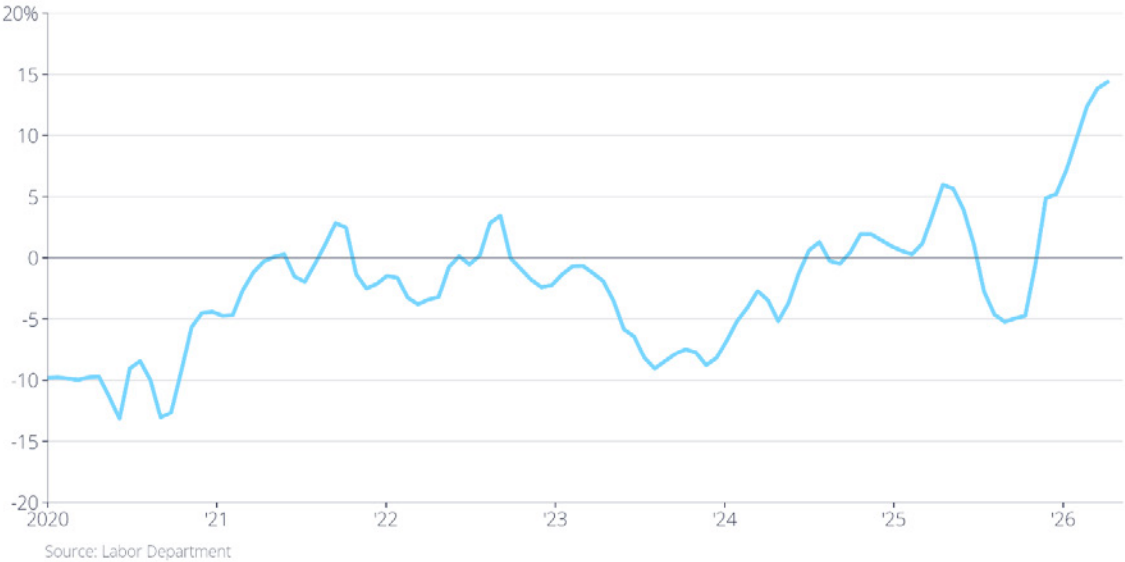
BlackRock AI & Data Acquisitions (2024–2026)

Announced Date	Target	Sector	Value (USD)
Nov 2024	CoreWeave (CRWV US)	AI cloud infrastructure / GPU data centers	\$650M
Dec 2024	HPS Investment Partners	Large-scale AI/data center infrastructure consortium	~\$12B (all-stock)
Oct 2025	Aligned Data Centers	Data Centers / Digital Infrastructure	\$40B
Feb 2026	Unnamed Target	Mega AI/data center infrastructure deal	\$122B
Feb 2026	Unnamed Target	Digital infrastructure consortium	\$30B
Mar 2026	AES Corp.	Power Generation for AI Data Centers	\$10.7B
Jul 2025	Unnamed Target (Talen Energy deal)	Power/data center infrastructure (TLN US)	\$2.33B

Ground-level research corroborates this. A June 2026 Silicon Valley trip by tech-equity investors Brook Dane and Sung Cho found a compute-constrained environment extending beyond GPUs to ASICs and memory chips, in what Cho called an unprecedented up-cycle. The strain

is driven by demand concentration: the top 5% of enterprise users consume roughly three times the tokens of the median company, with inference now overtaking training as the dominant volume driver.

Change in consumer prices for computer software and accessories from a year earlier



This structural mismatch is what makes the shortage investable: if supply can’t meet demand within a year, capex must stretch over three to five years, support-

ing Dane’s view that valuations across the supply chain remain “very attractive.”

2. A tradable market is already being built

Within a week of Fink's remarks, on 12 May 2026, CME Group announced what's billed as the first futures market for computing power. CEO Terry Duffy summed up the logic: "Compute is the new oil of the 21st century." The contracts let companies and traders hedge or speculate on GPU rental prices, much as airlines hedge fuel costs, offering enterprises a way to fix forward costs, and investors a new route to AI infrastructure exposure.

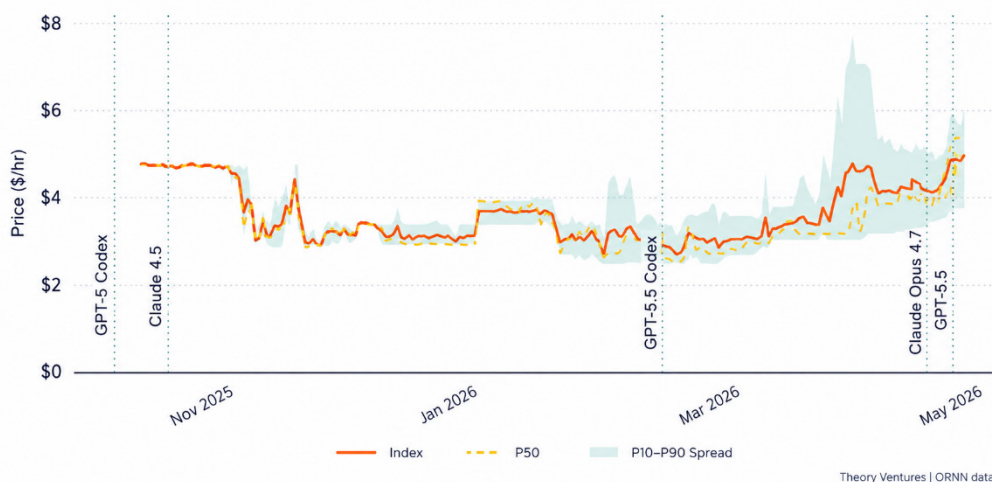
The speed of the launch is itself telling: an exchange only lists a contract when the underlying is scarce, volatile

and significant enough to generate hedging demand and liquidity. That a major derivatives venue reached this judgment within days suggests compute-as-commodity has moved from metaphor to consensus.

But a futures market embeds a directional bet: that today's scarcity persists long enough to sustain multi-year hedging demand. That assumption is exactly what the industry's own engineering roadmap is working to undermine.

B200 GPU Pricing Over Time

Index value with P10–P90 price spread on Ornn marketplace



3. Will the scarcity that justifies the market persist?

Engineering analysis of the AI inference stack suggests costs per token could fall sharply, by up to two orders of magnitude. Levers already in deployment include model quantisation and pruning (85 to 95% cost reduction), FP16-to-INT4 precision shifts (2 to 4x throughput gains), advanced 3D packaging (80 to 90% reduction), custom inference ASICs (70 to 80%), and co-packaged optics (50 to 65% lower energy per bit). These aren't distant gains: hyperscalers are committing over \$700 billion in 2026 capex, much of it toward industrialising these efficiencies.

This creates real tension: futures markets bet compute stays scarce; the engineering roadmap bets unit economics collapse. The two can coexist short term, but point in opposite directions long term. There's a further wrinkle: efficiency gains could let countries with older chip technology, notably China, run optimised models cheaply, turning a disadvantage into a competitive edge and complicating the "US capacity shortage" narrative.

Conclusion

The futures curve won't just price compute. It will reveal the market's honest view on how long scarcity lasts. A steep multi-year premium backs a structural shortage; a flat curve bets on efficiency winning out. Either way, the

contract becomes a live referendum on whether compute remains AI's defining constraint or becomes a commoditised input.

Welcome to Syzerland®

For further information

Banque Syz SA

Quai des Bergues 1
CH-1201 Geneva
T. +41 58 799 10 00
syzgroup.com

Charles-Henry Monchau, CFA, CAIA, CMT

Chief Investment Officer
charles-henry.monchau@syzgroup.com

Hugo Morel

Syz Research Lab Team
hugo.morel@syzgroup.com

Benjamin Roth

Syz Research Lab Team
benjamin.roth@syzgroup.com

This marketing document has been issued by Bank Syz Ltd. It is not intended for distribution to, publication, provision or use by individuals or legal entities that are citizens of or reside in a state, country or jurisdiction in which applicable laws and regulations prohibit its distribution, publication, provision or use. It is not directed to any person or entity to whom it would be illegal to send such marketing material.

This document is intended for informational purposes only and should not be construed as an offer, solicitation or recommendation for the subscription, purchase, sale or safekeeping of any security or financial instrument or for the engagement in any other transaction, as the provision of any investment advice or service, or as a contractual document. Nothing in this document constitutes an investment, legal, tax or accounting advice or a representation that any investment or strategy is suitable or appropriate for an investor's particular and individual circumstances, nor does it constitute a personalized investment advice for any investor.

This document reflects the information, opinions and comments of Bank Syz Ltd. as of the date of its publication, which are subject to change without notice. The opinions and comments of the authors in this document reflect their current views and may not coincide with those of other Syz Group entities or third parties, which may have reached different conclusions. The market valuations, terms and calculations contained herein are estimates only. The information provided comes from sources deemed reliable, but Bank Syz Ltd. does not guarantee its completeness, accuracy, reliability and actuality. Past performance gives no indication of nor guarantees current or future results. Bank Syz Ltd. accepts no liability for any loss arising from the use of this document.