AI: bubble, super-cycle, or both?



Source: iStockphoto.com/Jira Pliankharom

Why the hyperscalers' spending spree and circular AI financing deserve scrutiny, why the AI boom may still have years to run, and where the best investment opportunities may lie.

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

Malek Dahmani
Head of Equities
malek.dahmani@syzgroup.com

Assia Driss
Syz Research Lab Team Coordinator
assia.driss@syzgroup.com

Syz PRIVATE BANKING

Every great technology cycle tells two stories at once. The first is about the technology itself, which usually turns out to be even more transformative than its early champions imagined. The second is about the capital that funds it, which usually turns out to be even more excessive than its sceptics feared. Investors who confused the first story with the second often paid a heavy price. Those who dismissed the technology altogether missed a generational opportunity.

Artificial intelligence in the autumn of 2026 fits that pattern. The usage and revenue figures coming out of the frontier labs have no precedent in the history of software. Yet the largest US hyperscalers are on course to spend close to USD 1 trillion on capex next year, increasingly financed through debt. Much of their contracted backlog also depends on two frontier AI companies with enormous infrastructure commitments, themselves partly funded by their own suppliers. AI may have years of growth ahead, but it remains uncertain who will ultimately earn a return on the capital being deployed.

The super-cycle: demand is real, and it is being monetised

Revenue growth without precedent

Anthropic's annualised revenue has surged from roughly USD 9 billion in December 2025 to an estimated USD 74 billion in September 2026, with expectations approaching USD 100 billion by year-end. OpenAI is generating around USD 50 billion on an annualised basis. No software company has ever scaled at anything close to this pace. Estimated token demand has increased roughly eighteenfold since December.

Enterprise adoption is still at an early stage. Median token spending remains below USD 11 per employee per month, while the average enterprise use case generates approximately USD 55 in benefits for a token cost of just USD 2 to 5. On OpenRouter, frontier models account for just 11% of tokens processed, yet generate roughly three-quarters of implied revenue.

The economics of the build-out

Provided capacity is used, the economics are favourable. Broadcom CEO Hock Tan, who leads the US semiconductor and infrastructure software giant, estimates that each gigawatt of AI computing capacity generates around USD 30 billion in annual revenue, of which about USD 10 billion goes to power, cloud, chips and memory, and about USD 20 billion to the models and applications built on top. The relevant question is therefore how many gigawatts the market needs, and how quickly.

For now, demand continues to exceed supply. Microsoft's CFO has said that the shortage is visible even in GPU spot prices, and Azure growth has accelerated to 43%. Nvidia's quarterly revenue doubled year on year to USD 96 billion, and the company expects the coming year to be supply-constrained. Most capital expenditure also goes into short-lived servers, which can be slowed quickly if demand weakens.

Agents and the next wave of demand

The transition from chatbots to autonomous AI agents could extend this cycle beyond current expectations. It also changes the computing infrastructure required. Training typically uses one central processing unit (CPU) for every seven or eight graphics processing units (GPUs), compared with one CPU for every one to three GPUs in inference. Agentic artificial intelligence workloads can require one CPU per GPU or even more. Advanced Micro Devices (AMD), a US technology company that designs computer processors and other chips, now estimates that the server CPU market could reach USD 220 billion by 2030, up from approximately USD 26 billion in 2025.

Agents also communicate with one another, supporting demand for networking, optics and memory. Memory prices have increased sevenfold from their cyclical lows.

Why this is different from the dot-com bubble

Three differences set this cycle apart. First, the largest AI investors have highly profitable core businesses, and corporate earnings remain strong, with S&P 500 profits expected to grow approximately 30% year on year in Q3 2026, led by a 65% increase in technology. Second, demand is spreading beyond the frontier labs. Palantir reported 93% revenue growth in Q2, Databricks is growing at more than 80%. Third, the productivity gains from general-purpose technologies have always taken time to materialise. Factories needed decades to reorganise around the electric motor.

Super-cycles rarely end because valuations look stretched. They end when demand stalls or when financing dries up. Today demand is still accelerating, and financing remains available, if tighter and more creative. This argues for remaining invested, while keeping a clear view of the vulnerabilities.

Mapping the risks

To assess the bear case more clearly, eight key risks can be grouped into four categories according to how they would affect share prices: demand may disappoint, value may escape those who invest, supply may fail to keep pace, or a broader shock may affect the entire sector at once.

Demand

One overlooked short-term risk is tokenmaxxing, the possibility that AI consumption per employee reaches a plateau. Given how low spending remains relative to the benefits, however, the risk appears limited for now. Employment presents a longer-term challenge. According to research from Stanford University and ADP, a US company specialising in payroll and human resources services, employment among 22- to 25-year-olds in occupations exposed to artificial intelligence has declined by around 6% since the end of 2022. The impact on corporate earnings remains limited, but the political consequences could become more significant in 2027 and 2028.

Value capture

This is where the principal risks lie. Consensus estimates suggest that capital expenditure across the four largest hyperscalers will increase from approximately USD 723 billion in 2026 to USD 993 billion in 2027. According to Jefferies, a global investment banking and financial services firm, their combined quarterly free cash flow has fallen from USD 60 billion in the fourth quarter of 2025 to just USD 7 billion in the second quarter of 2026. Debt issuance is accelerating accordingly. Hyperscalers, including Oracle, have raised approximately USD 229 billion in investment-grade bonds so far this year, compared with USD 108 billion during the whole of 2025. The build-out is moving from being funded by profits to being funded by promises.

Competition is also shaping the economics of AI. Chinese open-weight models have evolved from low-cost alternatives into genuine competitors, now processing more than half of the tokens on OpenRouter at prices up to 150 times lower than US frontier models, while capturing only about 4% of revenue. Orchestration platforms are giving customers greater bargaining power. By routing each request to the cheapest model capable of completing the task, businesses can reduce costs by 40% to 50%. Both trends could put pressure on pricing and on the share of revenue retained by frontier model developers.

The combined effect is already visible in returns. The hyperscalers' realised return on invested capital is estimated to have fallen from around 40% to around 25%.

The capital loop: who is paying whom

The most concerning feature of the current cycle is the circular nature of its financial arrangements. Chipmakers such as Nvidia invest in the frontier AI labs. These labs then sign multi-year computing contracts with cloud providers and neoclouds, which buy chips to fulfil them and in several cases also hold equity stakes in the same labs. Nvidia alone has invested around USD 50 billion in frontier labs and has set up some USD 500 billion of financing platforms with partners.

A closer look at individual transactions shows how far suppliers now extend into their customers' capital structure:

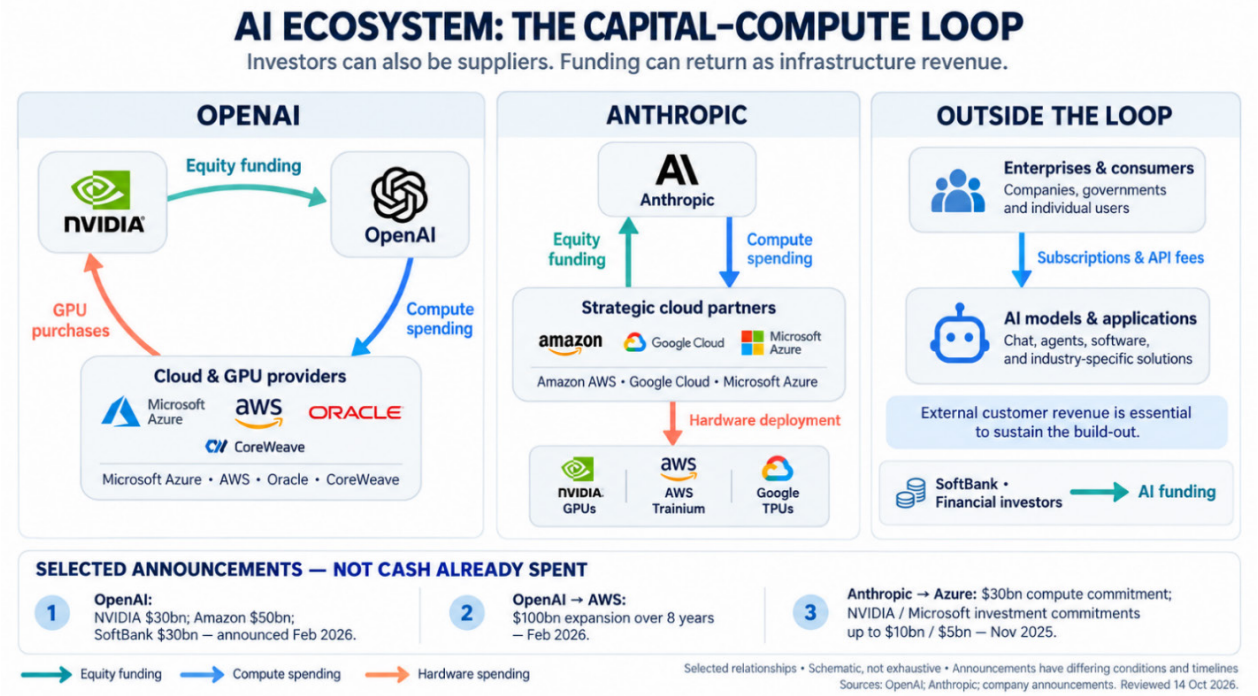
- **Suppliers are becoming shareholders:** Nvidia's letter of intent to invest up to USD 100 billion in OpenAI ultimately became a USD 30 billion equity investment.
- **Suppliers are also lenders:** Broadcom holds up to USD 42 billion in notes convertible into Anthropic shares, equivalent to roughly one-third of Anthropic's USD 125 billion TPU lease commitment to Broadcom.
- **Suppliers are providing guarantees:** Nvidia guarantees up to USD 105 billion of residual value on the 20-year lease of OpenAI's Ohio campus.
- **Private credit is taking on part of the risk:** Blackstone leads the USD 18 billion junior tranche of a USD 60 billion financing package involving Broadcom and Anthropic.

The sums involved are extraordinary. Remaining performance obligations at Microsoft, Alphabet, Amazon and Oracle reached approximately USD 2.3 trillion at the end of Q2 2026, up 186% in a year. Jefferies' investment strategist Chris Wood estimates that roughly half of this backlog is owed by OpenAI and Anthropic, equivalent to around ten times their combined annualised revenue. The cloud providers have, in practice, extended large unsecured credit to a small group of AI companies with substantial future financing needs. Most of these obligations are also difficult to unwind: the xAI contract is the only Anthropic compute commitment that can be cancelled on 90 days' notice.

Part of the exposure sits off balance sheet. Moody's counted USD 662 billion of data-centre lease commitments that had not yet commenced at the start of the year, up from USD 152 billion at the end of 2023.

These arrangements also have implications for reported earnings. Over the four quarters to Q2 2026, the four hyperscalers' annualised earnings increased by approximately USD 222 billion. Over the same period, other non-operating income, which includes unrealised gains on investments in private AI companies, rose by USD 224 billion. In other words, the increase in investment revaluations was roughly equivalent to the entire increase in reported earnings.

Two qualifications are important. Several headline figures represent maximum commitments, letters of intent or press reports, and some projects never materialise. For example, a planned 600-megawatt expansion of the Stargate AI data centre campus in Abilene, Texas, was abandoned by Oracle and OpenAI, although the existing campus continues to be developed. And as long as demand exceeds supply, the capacity being built is used. Circularity becomes a problem when one link in the chain weakens, whether through a failed funding round, a poorly received IPO or slower revenue growth at one of the labs. The telecoms crash of 2001 offers a historical parallel: Lucent Technologies, a major US telecommunications equipment maker, and Nortel Networks, its Canadian counterpart, suffered heavy losses as telecom operators they had helped finance cut spending. Microsoft chief executive Satya Nadella has recently recommended reading 1873, a history of the financial panic triggered by railway investment that ran ahead of revenues.



Source: Bank Syz, company announcements (2026)

Supply

The infrastructure build-out also faces significant physical constraints. Morgan Stanley estimates a potential US grid-access shortfall of 38 GW between 2026 and 2028, before accounting for solutions such as gas turbines and fuel cells. Memory is currently the most pressing semiconductor bottleneck. These shortages may slow infrastructure deployment, but they also create substantial pricing power for companies controlling scarce resources.

Broader market risks

Politics is becoming increasingly relevant ahead of the US midterm elections in November, with rising electricity bills attracting attention. Export controls and potential local taxes on data centres add further uncertainty. Market concentration is another concern. Asia-Pacific semiconductor stocks fell 21% between 22 June and 17 July without a major deterioration in fundamentals. During periods of stress, semiconductors, AI-related utilities and hyperscalers increasingly move together.

Category	Risk	Horizon	Probability	Impact	Most exposed
Demand	Tokenmaxxing	6–18 m	Low	Moderate	AI labs, SaaS multiples
Demand	Employment & macro	2–5 yrs	Medium	High, slow	Consumer, credit; regulatory backlash
Value capture	Capex return & circularity	12–36 m	Medium	Very high	Hyperscalers, neoclouds, AI credit
Value capture	Open-weight commoditisation	Ongoing	High	Selective	Labs without distribution
Value capture	Orchestration	Ongoing	High	Selective	Frontier pricing on simple tasks
Supply	People, power, chips	0–36 m	Very high	Two-sided	Deployers lose; bottleneck owners gain
Broader market	Politics & “Two Worlds”	6–24 m	Medium+	Moderate to high	Data centres, semis under export controls
Broader market	Concentration & technicals	Permanent	High	Episodic	The whole AI complex

Source: Bank Syz, based on Morgan Stanley, Jefferies, Stanford/ADP, OpenRouter and company reports (2026).

Taken together, these eight risks suggest that value capture deserves greater attention than a near-term slowdown in demand. Current usage data continue to support the growth story, but the distribution of profits across the AI value chain, and the circular financing behind it, remain much less certain.

Where the opportunities lie

The hyperscalers

The hyperscalers will very likely remain dominant franchises, but their investment case has changed. The main risk is a gradual deterioration in profitability, with returns settling well below historical levels while valuations continue to reflect the previous asset-light business model. Companies with diversified enterprise customer bases and competitive proprietary models appear better positioned than those relying heavily on one or two customers to support their contracted backlog. Oracle and the neoclouds sit at the riskier end of that spectrum.

Picks and shovels

Leadership rotates as bottlenecks shift. GPUs and custom accelerators led in 2023–24. Memory has been the trade of 2026. The cycle has turned in favour of CPUs, and networking and optics look set to be a key theme of 2027. Hardware nevertheless remains a cyclical business: in the 2021–22 downturn, memory stocks fell about 50% from peak to trough, networking 47% and optics 37%. This argues for diversified exposure across infrastructure bottlenecks, with a preference for companies benefiting from genuine supply constraints and strong pricing power.

Applications and the middle layer

The distribution of revenues across the AI value chain remains heavily concentrated in hardware. Based on forward revenue estimates, hardware accounts for approximately 70% of the market and applications for just 4%. As computing becomes cheaper, however, the economics could gradually shift in favour of applications. Valuations nonetheless remain demanding. In listed markets, preference goes to incumbents that are visibly turning their data and distribution into AI revenue, and to the "middle layer" of data platforms, orchestration and cybersecurity, where growth is re-accelerating and margins accumulate.

Conclusion

Because the end of a capex cycle tends to come from the financing side rather than from equity valuations, the signposts that matter most are:

- Credit spreads of AI-related borrowers and the health of the data-centre securitisation and SPV market, including the pricing of junior tranches.
- Capex guidance for 2027, and whether its growth begins to slow.
- Frontier labs' cash burn and fund-raising, including the reception of any IPO.
- Hyperscaler free cash flow and non-operating income, and how much of earnings growth is real cash.
- The share of non-cancellable commitments and any calls on supplier guarantees.

Both stories can be true for a long time, until financing conditions change. The conclusion is to stay in the theme and to be deliberate about where risk is taken within it: cautious on the ROIC expectations embedded in the hyperscalers, owning the next generation of picks and shovels with cyclical discipline, and building exposure gradually to the application layer. The railway boom ruined many of its financiers and transformed the world anyway. The aim is to be on the right side of both stories.

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

Malek Dahmani
Head of Equities
malek.dahmani@syzgroup.com

Assia Driss
Syz Research Lab Team Coordinator
assia.driss@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.