Duration Mismatch

Financing the AI Build-Out

The AI build-out has become an externally financed industrial project. The binding risk is not valuation but duration: three-to-six-year hardware funded with five-to-forty-year money, secured on a revenue pool concentrated in two private labs.

$800bn

HYPERSCALER CAPEX IN 2026, FIVE COMPANIES

32%

DEBT SHARE OF HYPERSCALER CAPEX, TRAILING

$266bn

AI-LINKED IG ISSUANCE BY EARLY SEPTEMBER 2026

23 OCTOBER 2026 · 22 MIN READ

EQUITYTRIBES.COM

An EquityTribes long read. Figures are as at 19 September 2026 unless dated otherwise; a note on what has changed since closes the piece.

In the eighteen months to September 2026 the AI build-out changed character. What began as the largest self-funded capital programme in corporate history has become an externally financed industrial project.

Five companies — Amazon, Microsoft, Alphabet, Meta and Oracle — will spend close to $800 billion on capital equipment this calendar year, roughly double 2025 and five times 2023. Sell-side consensus for 2027 already sits at $1.1–1.3 trillion.[1] Operating cash flow no longer covers it. Alphabet reported its first negative free-cash-flow quarter since its 2004 listing, Meta's free cash flow fell to $784 million, and Oracle's was negative $5.4 billion in the quarter to August. The debt share of hyperscaler capex has risen from 9 per cent in fiscal 2024 to 32 per cent on a trailing basis; AI-linked investment-grade issuance reached $266 billion by early September, against $29 billion in all of 2024; and Alphabet's $84.75 billion equity raise in June was the largest by a listed company on record.[2]

The usual question is whether AI is a bubble. It is the wrong question, or at least not the most useful one. The more useful question is who is holding which risk, for how long, and on what collateral. Asked that way, six things stand out.

1. **The binding risk is duration, not valuation.** Most of the money buys equipment that lasts three to six years; more and more of it is raised with instruments that last five to forty.
2. **Demand is real but concentrated.** The contracted backlog is large, and a large share of it is owed by two private companies.
3. **Circularity is a feature of the system, not an accusation.** Suppliers finance their customers, customers invest in each other, and the same revenue pool sits behind most of the paper.
4. **Physics has caught up with finance.** Power, memory and local consent now set the pace as much as capital does.
5. **The rate regime has turned.** The cost of capital rose at the moment the bet moved off balance sheet.
6. **The likeliest outcome is a productive overbuild, not a clean bubble.** The network will be valuable; the marginal capital that built it will not all be repaid.

An industrial build-out wearing a technology valuation

Capital spending by the large platforms has compounded at roughly 70 per cent a year since ChatGPT's launch and accelerated to more than 80 per cent in 2026. Amazon, Microsoft, Alphabet and Meta spent about $150 billion in 2023, $250 billion in 2024 and roughly $405 billion in 2025, including finance leases. Their calendar-2026 guidance sums to about $735 billion at the midpoints.

Add Oracle — $55.7 billion of capex in the fiscal year to May and $28.5 billion in the single quarter to August — plus the finance leases and customer prepayments that sit outside the headline line, and the cohort's 2026 outlay approaches $800 billion.[1]

Against the size of the economy the programme is already without modern precedent. Stijn Van Nieuwerburgh estimates that if US capacity grows by 200 gigawatts over 2026–32 at his unit costs, the $8.2 trillion of investment would run at about 2.8 per cent of GDP a year. That is more than the railway boom of 1865–90 (2.4 per cent), the interstate highway programme (1.6 per cent), electrification (1.1 per cent) or the telecoms build-out of 1996–2003 (0.8 per cent).[3] J.P. Morgan's Michael Cembalest finds that technology capex contributed 40–45 per cent of US GDP growth over the three quarters to the third quarter of 2025, and that in the fourth quarter investment associated with AI infrastructure accounted for "essentially all" of measured growth.[4]

The corporate spending is the visible top of a broader flow. Stanford's AI Index records $344.7 billion of private AI investment worldwide in 2025, up 127.5 per cent, of which the United States took $285.9 billion — 23 times China and 14 times Europe.[5] The first half of 2026 was larger still: Nvidia told investors that global venture funding of AI exceeded $400 billion in six months, surpassing all of 2025, and that roughly 70 per cent of it is spent on compute.[6]

What the money buys matters as much as how much of it there is. Microsoft said roughly two-thirds of its fourth-quarter capex went to CPUs and GPUs; Alphabet put about 60 per cent of its second-quarter spend into servers and 40 per cent into data centres.[7] The mix has shifted decisively towards short-lived assets.

What a gigawatt costs, and how long it earns

The cost of an AI factory is quoted in dollars per gigawatt of IT load, and estimates have converged on $30–60 billion. Morgan Stanley puts the all-in figure at $30–50 billion, about three-quarters of it IT hardware, with land, shell, substation, interconnection and commissioning at $10–15 billion. Jensen Huang told analysts in August that total investment per gigawatt has gone "from, say, $30 billion about five years ago to now $60 billion today", with Nvidia's own content rising from about $18 billion (Hopper) to $25 billion (Blackwell) to $40 billion (Vera Rubin).[8]

Two features of that cost structure matter more than its level. The expensive part is the short-lived part: hardware is roughly three-quarters of the project, and a Vera Rubin server now costs an estimated $8 million, nearly double the system it succeeds a year later.[9] And the buildings are cheap relative to what they house, which is why Morgan Stanley's stress test finds construction-cost inflation barely moves returns, while token pricing is "the single most impactful driver".[10]

THE DEPRECIATION LEVER

Everything downstream depends on how long the equipment earns. Between 2020 and 2024 the hyperscalers lengthened the assumed useful life of servers from three or four years to five or six — Microsoft, Google and Oracle to six, Meta to five and a half, Amazon to five — saving, on Vashist's

estimate, about $18 billion a year of depreciation across the industry.[11] Meanwhile Nvidia's architecture cadence is 18–24 months, and each new generation offers many times the throughput per megawatt of the last.[12]

Cembalest's sensitivity is the cleanest measure of what is at stake. Restating GPU and networking depreciation on a three-year life, applied to capex since January 2023, would cut 2025 pro-forma earnings per share by 6–8 per cent at Amazon, Meta, Microsoft and Alphabet and by 17 per cent at Oracle.[13] Michael Burry's more aggressive version — that two-to-three-year lives imply depreciation understated by about $176 billion over 2026–28 — is contested.[14] The direction is not. The useful-life assumption has become an active earnings lever at exactly the moment the equipment has become collateral.

RETURNS HANG ON THE PRICE OF A TOKEN

Morgan Stanley's stress test of a hypothetical 1 GW data centre — 65 per cent inference, full utilisation — yields a return on invested capital of 46 per cent at $1.80 per million tokens. The range runs from 20 to 72 per cent as the token price moves between $1.00 and $2.50: roughly six points of ROIC for every 10 per cent change in price.[10]

The difficulty is that the price is falling. The Silicon Data usage-weighted token price index peaked above $2.00 in May 2026 and was about $1.10 by August, as open-weight models took more than half of enterprise usage. Combined with usage data, Morgan Stanley finds total user spend on tokens "has apparently flatlined since June". Nvidia's counter is that its customers now report a payback on data-centre capital of "less than a year".[15] Both can be true, if the return accrues to the operator that captured early scarcity pricing and not to the marginal capacity being financed today.

Deflation is structural. The cost per million tokens fell from about $20 in late 2022 to $0.40 by August 2025, and Chinese models are on average about six times cheaper per token than US equivalents.[16] It cuts both ways: cheaper intelligence expands use, and agentic workflows consume five to thirty times the tokens of a chatbot query. But it also compresses the revenue per gigawatt on which every project's debt service depends.[17]

For a model developer the same arithmetic is starker. Vashist's "Lab X" — roughly 1 GW and $55 billion of cumulative infrastructure — needs about $78 billion of revenue to break even at OpenAI's disclosed 2025 gross margin of 33 per cent, but only $31 billion at a 60 per cent margin.[18] The labs' most important financial number is therefore not revenue growth but the spread between what a token sells for and what it costs to serve.

The capital stack: from cash flow to everyone else's balance sheet

Until 2025 the build-out was, in Cembalest's words, "mostly financed via internally generated cash flow… until very recently". That changed in five quarters. Morgan Stanley's forward free-cash-flow estimates for the hyperscaler cohort have "flipped negative"; FactSet expects fiscal-2026 free cash

flow “close to zero or negative for all except Alphabet and Microsoft”; Amazon’s long-term debt rose 81 per cent to $119 billion between December and March, and Alphabet’s rose 111 per cent to $98 billion in the first half.[19]

The instruments now in use form a stack of ten layers. Each moves the same underlying exposure — a claim on future AI revenue, secured on depreciating hardware — to a different holder with a different idea of what it owns.

LAYER	WHO PROVIDES IT, AND HOW MUCH (2026)	RISK RETAINED
1 · Operating cash flow	Amazon, Microsoft, Alphabet, Meta, Oracle; free cash flow near zero or negative for most	ROIC dilution; buybacks as the release valve
2 · Investment-grade bonds	Hyperscalers, developers, chipmakers; $266bn to early September	Spread widening; falling cover ratios
3 · Equity and convertibles	Alphabet, Oracle, Nebius, the lab IPOs; $135bn of US converts, 44% AI-related	Dilution; the IPO window
4 · Project finance and SPV leases	Meta–Blue Owl (Hyperion), Stargate; ~90% asset-level leverage	Four-year renewals; residual-value guarantees
5 · Securitisation	Data-centre ABS and CMBS; $61bn outstanding	Refinancing at the repayment date; tenant concentration
6 · GPU-backed loans	CoreWeave, Nebius, Lambda, Crusoe, Nscale; LTV 60–70%	Two-to-five-year collateral against five-to-seven-year contracts
7 · Customer prepayments	Oracle–OpenAI; $11.4bn in one quarter	Capacity prepaid but not yet built
8 · Vendor finance	Nvidia: ~$50bn in the labs, $500bn platforms (MoU)	Circularity; residual-value backstop
9 · Private credit and insurance	Blue Owl, Apollo, Blackstone, KKR, Brookfield; life insurers	Opacity; policyholder liquidity

LAYER	WHO PROVIDES IT, AND HOW MUCH (2026)	RISK RETAINED
10 · Sovereign capital	MGX, SoftBank, PIF/HUMAIN, EU InvestAI	Policy horizon; oil-linked budgets

BONDS: THE WORKHORSE, AT A PRICE

J.P. Morgan's tracker records $266 billion of AI capex-related dollar high-grade issuance in 2026 to 9 September — $182 billion from hyperscalers, $42 billion each from data-centre companies and semiconductor firms — and estimates $2.1 trillion of AI capex financing demand over the next five years.[20] Individual deals have set records: Alphabet's $32 billion in under 24 hours in February, Nvidia's $25 billion in June, Amazon's $25 billion eight-part sale in July.[21]

The market is absorbing it, but at a price. Hyperscaler bonds have "begun trading more like A-rated debt, particularly at longer maturities"; the insurers who are the natural thirty-year buyers have shown "isolated indigestion"; and Insight Investment calculates that the average initial cover ratio on hyperscaler deals fell from about 4.5 times in September 2025 to 1.8 times by July 2026.[22] Amazon's July sale was 2.5 times covered, against 3.2 times on its March issue. AI-tied debt is already an estimated $1.2 trillion, about 15 per cent of the investment-grade market.[23] Oracle is the outlier: S&P cut it to BBB-, one notch above speculative grade, on 9 July, and its 15-to-19-year lease obligations starting between fiscal 2027 and 2029 total almost three times its fiscal-2027 capex guidance.[24]

HYPERION: THE TEMPLATE

The transaction that defines the cycle is Meta's Hyperion campus in Richland Parish, Louisiana: roughly 2 GW and about $30 billion. Meta sold 80 per cent of the equity in a joint venture to Blue Owl, and the venture raised $27 billion of debt in October 2025, rated A+ — one notch below Meta — at a yield of 6.58 per cent. Meta is sole tenant under five successive four-year leases running from 2029 to 2049, may terminate portions at each renewal, and stands behind a residual-value guarantee that declines to zero over twenty years. Under current accounting, as Van Nieuwerburgh notes, Meta recognises neither the leases nor the guarantee on its balance sheet, "even though one of these two liabilities is certain to materialize".[25]

Hyperion matters less for its own risk than as a template. Asset-level leverage is about 90 per cent against Meta's 25 per cent book leverage, and the premium paid for the structure — at least 100 basis points over Meta's own debt on most estimates — adds more than $5 billion of interest over the life of the bond, passed back to Meta as rent.[26]

THE LEDGER NOBODY CONSOLIDATES

Generalise the mechanism and you get an off-balance-sheet ledger that no single disclosure captures. Moody's finds the hyperscalers carry about $970 billion of lease commitments, $660 billion of which is not on balance sheet. Goldman, as reported by the FT in August, puts the figure at $1.5 trillion, $1 trillion of it off balance sheet. Morgan Stanley counts a further $500 billion or so of hardware purchase commitments and about $400 billion of signed leases that have not yet

started, which it thinks unlikely to be reflected in consensus earnings.[27] The estimates use different definitions and are not additive. What they share is the order of magnitude: the true leverage of the build-out sits in footnotes.

SECURITISATION: THE EXIT FOR FINISHED BUILDINGS

Stabilised, leased data centres are refinanced in the securitisation markets, which have grown from $4 billion outstanding in 2020 to $61 billion in 2026. Structures are conservative — loan-to-value of no more than about 70 per cent, an anticipated repayment date of around five years, final maturities of 25–30 years — and deal sizes are far smaller than a single AI campus, which is why bonds, project finance and private credit carry the construction risk before securitisation takes the finished asset.[28]

On 29 July the SEC's Division of Corporation Finance confirmed that whole-business data-centre securitisations are not "asset-backed securities" under the Exchange Act, because the collateral is not self-liquidating. Sponsors are therefore free of the 5 per cent risk-retention rule and related disclosure obligations the market had observed voluntarily since 2018. The letter is narrow and fact-specific, but its direction is to lower the cost of securitising the physical layer just as volumes accelerate.[29]

NEOCLOUDS: WHERE THE MISMATCH IS SHARPEST

The specialist GPU clouds are where duration risk is most concentrated. CoreWeave's total debt reached about $35 billion by mid-2026, it guides to $35–39 billion of capex for the year against revenue of $12.4–13.2 billion, and its second-quarter net loss of $626 million was driven by $640 million of net interest expense. Its facilities are structured one special-purpose vehicle per cluster, ring-fenced against a named customer contract, with loan-to-value of 60–70 per cent on GPU collateral; the investment-grade rating of its Meta-backed loan comes from the tenant, not the borrower. Microsoft, OpenAI and Meta were more than 60 per cent of its 2025 recurring revenue.[30]

Morgan Stanley describes the model plainly: an SPV raises money from GPU investors, buys servers and chips, and receives five-year lease payments from a hyperscaler, which gets compute without upfront investment — "leasing 'cars' (the GPUs) rather than buying them". The contracts run five to seven years; the collateral is on an 18–24-month product cycle; and a meaningful share of the debt matures between 2026 and 2028.[31] CME Group's plan to list futures on H100 and B200 rental indexes would, if it trades, give lenders a public forward curve for the rent their collateral earns for the first time.[32]

NVIDIA AS BANKER TO ITS CUSTOMERS

Nvidia's balance sheet has become load-bearing. On its 26 August call the company said it had invested "nearly $50 billion" in the frontier labs, had provided "selective credit enhancement for nearly 2 gigawatts" for a second lab, and expects to "leverage our balance sheet to contribute toward roughly a quarter of our business next year". For neoclouds it offers a take-or-pay floor on part of a facility's capacity — "a minimum revenue guarantee that gives lenders the confidence to underwrite the project" — in exchange for a share of revenue above it. "We're not making loans," the company said. "In this model, we get paid twice."[33]

On 10 August Nvidia announced platforms with Apollo, BlackRock, Blackstone, Brookfield, Goldman Sachs and KKR to raise more than $500 billion of third-party capital for AI factories with compute as collateral, and agreed to cover up to 25 per cent of any shortfall in the value of GPUs pledged. The platforms are memorandums of understanding, not committed capital.[34] Colette Kress addressed the obvious objection directly: "we know some will call this circular financing. We see it differently."

PREPAYMENTS, PRIVATE CREDIT AND THE INSURANCE CHANNEL

Oracle has pioneered a third mechanism: making customers pay upfront. Its operating cash flow of $23.1 billion in the quarter to August included $11.4 billion of customer prepayments, which it now deducts to report a "net cash outlay" for capex of $18.0 billion against $28.5 billion gross.[35] Prepayments align capacity with demand but concentrate counterparty risk. Cembalest's description of the anchor contract remains the sharpest: OpenAI committing "to pay Oracle $60 bn per year, an amount OpenAI doesn't earn yet, to provide cloud computing facilities that Oracle hasn't built yet, and which will require 4.5 GW of power".[36]

Private credit is the layer official statistics see least clearly. A 2025 Morgan Stanley estimate allocated about $800 billion of the $1.15 trillion of debt needed for hyperscaler compute over 2025–28 to private credit.[37] Its ultimate funders are increasingly insurers: life insurers own about 42 per cent of the $2 trillion private-credit universe, and private-equity-controlled insurers place half of new investment in privately placed instruments, against 14 per cent for independents.[38] The Financial Stability Board warned in May that private credit's links with banks, insurers and private equity "could amplify stress in adverse scenarios"; the BIS noted in June that financing of the AI boom "looks increasingly reliant on debt and complex funding structures across the supply chain".[39] The policyholder has become a silent partner in the build-out.

Demand versus commitments

The bull case rests on contracted demand, and the contracted numbers are large. Microsoft's commercial remaining performance obligations reached $678 billion at the end of June, up 84 per cent; AWS's backlog stood at $496 billion after growing $132 billion in one quarter; Alphabet's exceeded $460 billion; Oracle's reached $664 billion in August.[40] Nvidia's summary — "the cloud industry's backlog now exceeds $2 trillion" — is, if anything, conservative.

Concentration is the caveat. By Myrmikan's tally of company disclosures, half of the $2 trillion of firm orders comes from OpenAI and Anthropic, and half of Microsoft's backlog is attributable to OpenAI. Vashist's test is the right one: "What proportion of AI infrastructure spending is supported by unaffiliated, recurring, cash-paying end customers?" The honest answer is that nobody outside the companies knows.[41]

TWO LABS, TWO TRAJECTORIES

The two companies at the centre of the backlog diverged sharply in 2026. Anthropic generated $11.6 billion of revenue in the second quarter, more than double the first, with its first positive operating result. OpenAI's revenue rose from $5.7 billion to $6.7 billion over the same period, on an

operating loss of $12.3 billion; it has cut its infrastructure plan from $1.4 trillion to about $600 billion through 2030 and expects to be cash-flow positive no earlier than 2030.[42] Both companies' commitments dwarf their revenues: OpenAI's roughly $600 billion includes about $250 billion to Azure and $300 billion to Oracle; Anthropic has committed about $200 billion to Google Cloud over five years. Prepaid multi-year cloud contracts, rather than owned data centres, "simply relocate[s] the depreciation and financing risk… onto the hyperscaler and neocloud balance sheets".[43]

THE CIRCULAR FLOW

The transactions linking the labs to their suppliers and financiers form a closed loop. Hyperscalers buy chips from Nvidia. Nvidia invests in the labs. The labs commit to buy compute from the hyperscalers and neoclouds, partly with that capital. The hyperscalers invest in the labs and book the commitments as backlog. The backlog supports the bonds that fund the next round of chip purchases.

The accounts show the loop. Nvidia's four largest direct customers were 61 per cent of its fiscal-2026 revenue.[44] Amazon's second-quarter net income of $62.6 billion included a $53.4 billion pre-tax gain on its Anthropic stake, and Alphabet's $112 billion included $98 billion of mostly unrealised gains on the same holding — the investee's valuation, set in a round the investors anchored, flowing through the investors' reported profits.[45] BCG counts more than $1.5 trillion of deals between US labs, chip designers, hyperscalers and capital providers since 2024, and warns that "a default, write-down, or supply disruption at any major player could ripple quickly across the rest".[46]

None of this makes the demand fictitious. ChatGPT has about 900 million monthly users; generative AI reached 53 per cent population adoption within three years, faster than the personal computer or the internet; enterprise adoption is close to universal.[47] What the loop does is make demand look more independent than it is, and make the system's expectations move together.

IS ANYONE MAKING MONEY DOWNSTREAM?

The evidence on returns to AI users is improving but remains thin relative to the capital deployed. On one side: productivity gains of 14–26 per cent in customer support and software development, 30 million paid Copilot seats, nearly twenty AI-native companies above $1 billion of recurring revenue. On the other: only 37 per cent of organisations attribute any positive operating-profit impact to AI at enterprise level, 93 per cent report exceeding their AI budgets, and agent deployment at scale remains in single digits across nearly every business function.[48]

The distributional point matters more for financiers. US consumer surplus from generative AI is estimated at $172 billion a year, and Nordhaus's finding that innovators historically capture about 3 per cent of the social returns to their innovations is the base rate.[49] Value can be created on a vast scale while the suppliers of capital to the physical layer earn less than their cost of capital — as they did in railways, electrification and fibre.

Physics has caught up with finance

ELECTRICITY

The IEA projects data-centre electricity consumption roughly doubling from 485 TWh in 2025 to 950 TWh in 2030, about 3 per cent of global demand, and judges that "around 20% of planned data centre projects could be at risk of delays" from grid constraints.[50] In the United States the constraint binds first in PJM, the largest grid. Its 2028/29 capacity auction on 14 July cleared at the regulatory cap of $325 per megawatt-day for the third year running — uncapped, it would have cleared at $554.72 — and procured 6,831 MW less than the reliability requirement, the second shortfall in PJM's history.[51] Its market monitor put it bluntly: "Current capacity in PJM is not adequate to meet demand from large data center loads and will not be adequate in the foreseeable future. This is a simple factual issue."[52]

The equipment to fix it is scarce. New combined-cycle gas turbines cost $2,200–2,500 per kilowatt against $722 in 2022, with delivery in 2029–30; transformer lead times exceed two years; UK grid connections take eight to ten years, and EirGrid has halted new Dublin connections until 2028.[53] Operators are going around the grid — more than 55 GW of behind-the-meter capacity is planned for US data centres — at a cost: Morgan Stanley estimates self-generation roughly doubles power costs, from $0.10–0.12 to $0.20–0.22 per kWh.[54]

MEMORY AND THE SUPPLY CHAIN

The second constraint is components. Shortages that began in high-bandwidth memory have spread to DRAM and NAND; DDR5 prices are up about 110 per cent this year, HBM capacity is sold out through 2026, and memory may absorb about 30 per cent of hyperscaler AI data-centre spend this year against 8 per cent in 2023–24. Nvidia expects "extreme pricing conditions in memory" to push its gross margin from 75 per cent to a trough of 71–72 per cent.[55] Upstream concentration is structural: one foundry makes virtually every leading AI chip, three companies supply all HBM, and one supplies EUV lithography.[56]

CONSENT

The third constraint is political. Data Center Watch counted at least 75 US projects worth about $130 billion blocked or delayed in the first quarter of 2026 — roughly all of 2025 in three months — and opposition groups more than doubled to 833 across 49 states. Seventy-one per cent of Americans oppose a data centre near them.[57] The Carnegie data-centre model shows why consent matters more than energy prices: a one-year delay costs an illustrative 100 MW facility more than 5.5 per cent of its lifecycle value; a doubling of energy prices, 4.5 per cent.[58]

The rate regime has turned

The build-out was designed at a cost of capital that no longer exists. The US ten-year yield touched 5.01 per cent on 14 September, its highest since 2007, and the Federal Reserve raised its target range to 3.75–4.00 per cent two days later, its first increase since 2023. Chair Warsh named "more

competition for capital" among the drivers of higher yields; private AI capex, public borrowing for infrastructure and defence, and an energy-driven inflation shock are all bidding for the same savings.[59]

The arithmetic for the stack is unforgiving. A fifty-basis-point move in the ten-year takes 8–10 per cent off the present value of an eight-to-ten-year hardware cash-flow stream. Hyperion's 6.58 per cent coupon and Alphabet's 6.125 per cent century bond were struck before the latest leg up. Morgan Stanley's consolation — "a potentially higher-for-longer cost of capital acts as a self-curing agent, preventing irresponsible capital flows" — is correct, and cold comfort for the marginal project.[60]

Where the losses would surface

The layers fail in a predictable order. Vashist's transmission sequence runs from weak end-user monetisation, through lower lab revenue and lower GPU-cloud utilisation, to renegotiated cloud commitments, slower building, capex moderation and lower demand for chips and memory. The weak points are where "capital intensity, limited diversification, weak contractual protection, and uncertain end-market pricing overlap": the leveraged GPU cloud, the single-tenant development, the lab with commitments far ahead of revenue.[61] The hyperscalers themselves can absorb AI losses for a long time without a credit problem. The first casualties would be the vehicles that borrowed against the hyperscalers' credit without owning it.

The system has already rehearsed. CoreWeave's late-2025 construction delays left part of its fleet idle while interest accrued; Nvidia responded with a roughly $2 billion equity injection.[62] Oracle's shares fell about 35 per cent and its spreads widened 90 basis points in the months after its leverage was flagged in September 2025, and it has since lost its BBB rating.[63]

The exposure is transmitted through equity markets that have never been more concentrated on one theme. Forty-two AI-linked S&P 500 companies delivered 78 per cent of the index's price return and 66 per cent of its earnings growth from November 2022 to December 2025.[64] Goldman's framing is the one to hold on to: "the risk for investors is less about valuation and more whether there is an unsustainable earnings bubble". The analogue is the banks in 2007, not the dot-coms in 2000.[65] And data centres no longer diversify it: the beta of data-centre REITs has risen from about 0.5 to about 1. "Data centers no longer behave as relatively defensive infrastructure investments," Van Nieuwerburgh concludes.[66]

WHAT TO WATCH

INDICATOR	READING, SEPTEMBER 2026	WHAT WOULD CHANGE THE PICTURE
Hyperscaler capex vs operating cash flow	Capex above cash flow at Alphabet, Meta, Oracle; debt share 32%	Debt share above 50%, or buybacks cut to fund capex

INDICATOR	READING, SEPTEMBER 2026	WHAT WOULD CHANGE THE PICTURE
AI-linked IG issuance and cover ratios	$266bn; cover 4.5x → 1.8x	A pulled hyperscaler deal; spreads 100bp wider than AA peers
Token prices vs GPU rents	~$1.10/M tokens, down from >$2.00; H100 rents rising	Token spend falling while capacity keeps growing for two quarters
Neocloud refinancing	CoreWeave ~$35bn debt; $640m quarterly interest	A GPU-backed facility amended, extended or foreclosed
Frontier-lab economics	OpenAI −$12.3bn operating loss on $6.7bn; Anthropic profitable	Renegotiation of the Oracle or Azure commitments
Off-balance-sheet commitments	$0.66–1.0tn of leases off balance sheet	Rating agencies consolidating leases and guarantees
Power	PJM 6.8 GW short; capacity price at the cap	A large campus missing energisation by more than a year
Consent	$130bn blocked or delayed in Q1	State-level moratoria in Virginia or Texas
Rates	10-year 4.9–5.0%; Fed hiking	10-year sustained above 5.25%; insurers leaving 30-year IG
Private-credit liquidity	Redemption gates at large managers in H1	Gating at an infrastructure-heavy vehicle

Three paths

Scenarios for this cycle have to be specific to place, workload, contract and hardware generation: the same year can bring overbuild in Northern Virginia and scarcity in Frankfurt, idle training capacity and scarce inference capacity, worthless Hopper fleets and appreciating grid-connected land. With that caveat, three paths bound the range.

- **Disciplined scaling.** Inference and agentic demand grow fast enough to keep new capacity busy, power stays scarce enough to hold pricing, and cloud and application revenue catch up with depreciation. Signposts: token spend rising again; hyperscaler free cash flow turning positive; new

capacity clearing in PJM. Losses fall only on marginal capital.

- **Productive overbuild.** Capacity grows faster than lab revenue, token prices fall further, leveraged operators restructure and older fleets lose value — while cheap compute accelerates adoption downstream. Signposts: token prices below $1.00 for two quarters; neocloud contract renegotiations; capex growth slowing sharply. Losses fall on neocloud equity, GPU-backed lenders, single-tenant landlords and late-vintage GPU owners; users inherit the surplus.
- **Financing squeeze.** A demand disappointment coincides with tighter money: a lab misses targets, a GPU-backed vehicle defaults, spreads widen and insurers retreat from the long end. Losses spread to private-credit investors and, through affiliated vehicles, to insurance policyholders, long-dated bondholders and concentrated equity indices.

The paths are not exclusive, and the second is the one every previous general-purpose-technology boom has followed: the network is built, the builders' capital is impaired, and the users inherit the surplus. The question that follows — whether losses concentrated in private credit and insurance become a public problem — has no settled answer yet. Van Nieuwerburgh frames it best: "the relevant question is therefore not whether risks are already systemic, but under what conditions they could become so".[67]

Where durable value sits

The map is stable across sources. Value is concentrated upstream — foundry gross margins of 60–68 per cent, Nvidia at 71–75 per cent, a memory oligopoly with pricing power for the first time in a decade — and the financial vulnerability sits in the middle: neoclouds, single-tenant developers, financing vehicles and labs with commitments ahead of revenue.[68] Four positions look durable within it:

- **Scarce physical inputs** — grid connections, transformers and switchgear, high-bandwidth memory, advanced packaging, optical networking — earn while shortages last, subject to the old capacity cycle in which high margins invite new supply.
- **Full-stack efficiency** — custom silicon, scheduling, routing, quantisation — creates value without proportional power.
- **Distribution** — the platforms and application vendors that own a workflow and proprietary data capture AI revenue without owning depreciating hardware, which is why September's sell-off rotated into application software rather than out of technology.
- **The financial plumbing itself** — securitisation platforms, collateral valuation, compute futures, the fees of the asset managers in the financing platforms — earns a spread on the mismatch however it resolves.[69]

For anyone underwriting a data centre or a compute facility, the questions follow directly. Are power, permits, water and long-lead equipment secured, or planned? Are shell-and-power and IT economics modelled separately, with lives of roughly fifteen years and four? Who is the true end customer, and is its demand independent of financing from the project's own suppliers? Have utilisation, token price, power cost, delay, refinancing cost and hardware residual value been

stressed together rather than one at a time? Where does every lease, guarantee and termination right ultimately land? And does a 2026–28 maturity secured on 2024-vintage GPUs sit anywhere in the structure? If so, that is the weakest point in the system.

Since 19 September

Three developments in the weeks after this analysis was closed sharpen rather than change it.

- **Vendor finance has spread beyond Nvidia.** On 2 October Bloomberg reported that Broadcom's banks were assembling $60 billion of AI chip financing for Anthropic and others, in a $42 billion senior and an $18 billion junior tranche, the latter led by Blackstone; a week later it reported early talks on a further $30 billion for OpenAI.[70]
- **Run-rates turned out to be definitions.** On 8 October the FT reported OpenAI's annualised revenue at close to $50 billion at the end of September, about $20 billion below the figure that had circulated. The two labs count partner sales differently, so a single headline comparison between them says as much about accounting as about demand.[71]
- **Private valuations reached retail.** A US exchange-traded fund began offering exposure to OpenAI and Anthropic through swaps priced off pre-IPO perpetual futures — a daily public price for valuations no audited account yet supports.[72]

Each is a further step along the same line: the further the exposure travels from the operator's balance sheet, the less the holder can see of what secures it.

This analysis is independent editorial research for information and education only. It is not investment, legal or tax advice, a personal recommendation or an offer or solicitation to buy or sell any security. Figures from company guidance, leaked documents and press reports are unaudited; where sources conflict, audited disclosures and official statistics are preferred.

Footnotes

1. Calendar-2026 capex guidance after second-quarter reporting: Amazon ~$220bn; Alphabet $195–205bn; Microsoft ~$175bn on a calendar basis after a lease reclassification; Meta $135–145bn. FactSet puts calendar-2026 guidance for the five "close to $800bn" including finance leases and prepayments. Morgan Stanley's sell-side consensus is $800bn for 2026 and $1.1tn for 2027; Nvidia cited $1.3tn for 2027 on its 26 August 2026 call.
2. FactSet Insight, "Hyperscalers Tap External Financing as AI Capex Outruns Cash Flow", 23 July 2026; J.P. Morgan weekly credit report, 9 September 2026; Alphabet equity raise priced 1 June 2026.
3. Stijn Van Nieuwerburgh, "Financing the AI Buildout", draft of 20 March 2026, p. 4.
4. J.P. Morgan Asset & Wealth Management, "Smothering Heights", Eye on the Market Outlook 2026, 1 January 2026, pp. 1 and 38.
5. Stanford HAI, AI Index Report 2026, chapter 4.2.
6. Nvidia Q2 FY2027 earnings call, 26 August 2026, prepared remarks.
7. Microsoft FY26 Q4 call (29 July 2026); Alphabet Q2 2026 call (22 July 2026).
8. Morgan Stanley Wealth Management, "AI Data Centers: The Heart of the Ecosystem", 25 August 2026, pp. 4, 5 and 11; Nvidia Q2 FY2027 call.
9. Morgan Stanley (25 August 2026), p. 6.
10. Morgan Stanley (25 August 2026), pp. 2, 9 and 11.
11. J.P. Morgan Outlook 2026, p. 16; Vashist, "The AI Investment Value Chain" (September 2026), §9.
12. Nvidia Q2 FY2027 call, p. 4.
13. J.P. Morgan Outlook 2026, p. 16.
14. Vashist (2026), §9, citing Burry (November 2025).
15. Nvidia Q2 FY2027 call, p. 15.

16. Storm, INET Working Paper 244 (December 2025), p. 20; Invesco, “An alternative model: China’s low-cost AI strategy”, July 2026, pp. 5–6.
17. McKinsey, Technology Trends Outlook 2026, trend 02; J.P. Morgan Outlook 2026, p. 18.
18. Vashist (2026), §5.2.
19. J.P. Morgan Outlook 2026, p. 3; Morgan Stanley (25 August 2026), p. 5; FactSet (23 July 2026); CNBC, 28 July 2026.
20. J.P. Morgan weekly credit report, 9 September 2026, as reported.
21. Reuters and CNBC, 7 July 2026; Fortune/Bloomberg, 8 July 2026, citing Bank of America.
22. Morgan Stanley (25 August 2026), p. 10; Insight Investment, September 2026.
23. Bloomberg estimate as cited by Myrmikan Research, 14 August 2026, p. 15.
24. FactSet (23 July 2026); CNBC, 10 September 2026.
25. Van Nieuwerburgh (2026), pp. 8–10; J.P. Morgan Outlook 2026, p. 4.
26. Storm (2025), p. 24; J.P. Morgan Outlook 2026, p. 4.
27. Van Nieuwerburgh (2026), p. 10, citing Moody’s (2026); Morgan Stanley (25 August 2026), p. 5; Financial Times, 10 August 2026, as cited by Myrmikan (p. 11).
28. Structured Finance Association, Research Corner, 23 July 2026, citing Barclays Credit Research.
29. SEC Division of Corporation Finance, interpretive letter of 29 July 2026 on certain data-centre securitisations; O’Melveny client alert, 18 August 2026.
30. CoreWeave Q2 2026 results (11 August 2026); Vashist (2026), §§4.4 and 7.
31. Morgan Stanley (25 August 2026), p. 7; Vashist (2026), §§4.4 and 11.
32. CME Group press release, 11 August 2026 (launch planned for 5 October 2026, pending regulatory review).
33. Nvidia Q2 FY2027 call, transcript pp. 5–7.
34. Nvidia press release, 10 August 2026; TechCrunch, 13 August 2026.
35. Oracle Q1 FY2027 press release and financial tables, 10 September 2026.
36. J.P. Morgan Outlook 2026, p. 4.
37. Van Nieuwerburgh (2026), p. 7, citing Morgan Stanley Research (2025).
38. Myrmikan (14 August 2026), pp. 15–17, citing the Chicago Fed and the Federal Reserve Financial Stability Report (May 2026).
39. Financial Stability Board, May 2026; BIS Annual Economic Report, 28 June 2026.
40. Company reports: Microsoft (29 July 2026), Amazon (30 July), Alphabet (22 July), Oracle (10 September); Nvidia Q2 FY2027 call.
41. Myrmikan (14 August 2026), p. 8; Vashist (2026), §7.
42. Bloomberg, 15 August 2026, from internal documents; Vashist (2026), §5.
43. Vashist (2026), §5.2; QBC Analyst, “The AI Financial Ecosystem: Fragile or Foundational?”, 10 September 2026, p. 2.
44. Vashist (2026), §3, from Nvidia’s FY2026 10-K.
45. Company Q2 2026 results; Myrmikan (14 August 2026), pp. 9–10.
46. BCG, “The Great Divide”, 30 June 2026, p. 8.
47. Deutsche Bank Research Institute, “AI IPOs 101”, 1 July 2026, p. 2; AI Index 2026, chapter 4.3.
48. AI Index 2026, chapter 4.4; McKinsey, “The State of AI in 2026” (25 August 2026), as cited in McKinsey (2026).
49. AI Index 2026, figure 4.2.22 (Brynjolfsson et al., 2026), citing Nordhaus (2004).
50. IEA, Key Questions on Energy and AI (2026); IEA, Energy and AI (2025).
51. PJM Interconnection, 2028/2029 Base Residual Auction results, 14 July 2026.
52. Monitoring Analytics, 13 November 2025, as quoted in J.P. Morgan Outlook 2026, p. 25.
53. Storm (2025), pp. 29–30; J.P. Morgan Outlook 2026, pp. 22–23; WEF/Bain, “Rethinking AI Sovereignty” (January 2026).
54. ING via Benzinga, June 2026; Morgan Stanley (25 August 2026), pp. 4 and 12.
55. Morgan Stanley (25 August 2026), p. 6; Nvidia Q2 FY2027 call, p. 7.
56. AI Index 2026, §1.3; OECD Competition Committee roundtable summary (18 May 2026), p. 4.
57. Data Center Watch, Q1 2026 report; Deutsche Bank (1 July 2026), pp. 11–13, citing Gallup.
58. Bruegel Policy Brief 18/2026, citing the Carnegie Endowment International AI Data Center model.
59. WSJ, 14 September 2026; FOMC statement, 16 September 2026; Goldman Sachs, “Competition for Capital”, 17 September 2026.
60. Morgan Stanley (25 August 2026), p. 13.
61. Vashist (2026), §§3.1 and 6.
62. Vashist (2026), §§4.4 and 8.
63. J.P. Morgan Outlook 2026, p. 3; FactSet (23 July 2026).
64. J.P. Morgan Outlook 2026, pp. 1–2.
65. Goldman Sachs (17 September 2026), pp. 3–4 and 7–8.
66. Van Nieuwerburgh (2026), p. 12.
67. Van Nieuwerburgh (2026), p. 15.
68. Vashist (2026), §§2–3.
69. McKinsey (2026), trend 05; Morgan Stanley (25 August 2026), pp. 6 and 13.
70. Bloomberg via Advisor Perspectives, 2 October 2026; The Star (Bloomberg), 9 October 2026.
71. Malay Mail (AFP), 9 October 2026; The Next Web, 9 October 2026.
72. Corgi Invest (PR Newswire via Morningstar), 5 October 2026.

FROM EQUITYTRIBES

TINO, every week

TECTONICS · INFLECTIONS · NARRATIVES · OUTLIERS

One short edition a week that reads capital, energy, technology and culture through four lenses: the structural shifts, the turning points, the stories that do or do not hold up, and the ideas at the edge. Free, by email.

equitytribes.com/tino

REPLY TO THIS PIECE

Disagree, or know something we missed? Write to info@etiic.com. The most useful replies are published, with permission, as letters.

READ IT ONLINE

equitytribes.com/analysis/duration-mismatch-financing-the-ai-build-out/
with linked sources and any updates since publication.

© 2026 EquityTribes. Independent editorial research for information only — not investment, legal or tax advice.