

Cut, hedge, or hold the AI- infrastructure bet

*The trillion-dollar build-out, and the tripwires
that flip hold to cut*

CONFIDENTIAL AND PROPRIETARY

*Any use of this material without specific permission
of Huegoo is strictly prohibited*

Hold the demand-backed core, hedge the tail, and tilt off the leveraged edge

1

The call

hold the demand-backed core, hedge the tail with index puts, and tilt toward power and REITs off the leveraged neoclouds, with tripwires that flip hold to cut.

2

Demand is real

Nvidia data-center revenue ran toward 194 billion dollars, and customers contracted about 2 trillion of backlog.

3

The structure is fragile

capex intensity tripled to 45-to-57 percent on about 121 billion of 2025 debt, a slice off balance sheet.

4

The ROI gap

GPUs are booked over six years but age in two, and required revenue of 650 billion to 2 trillion dwarfs about 75 billion today.

5

The hedge is priced

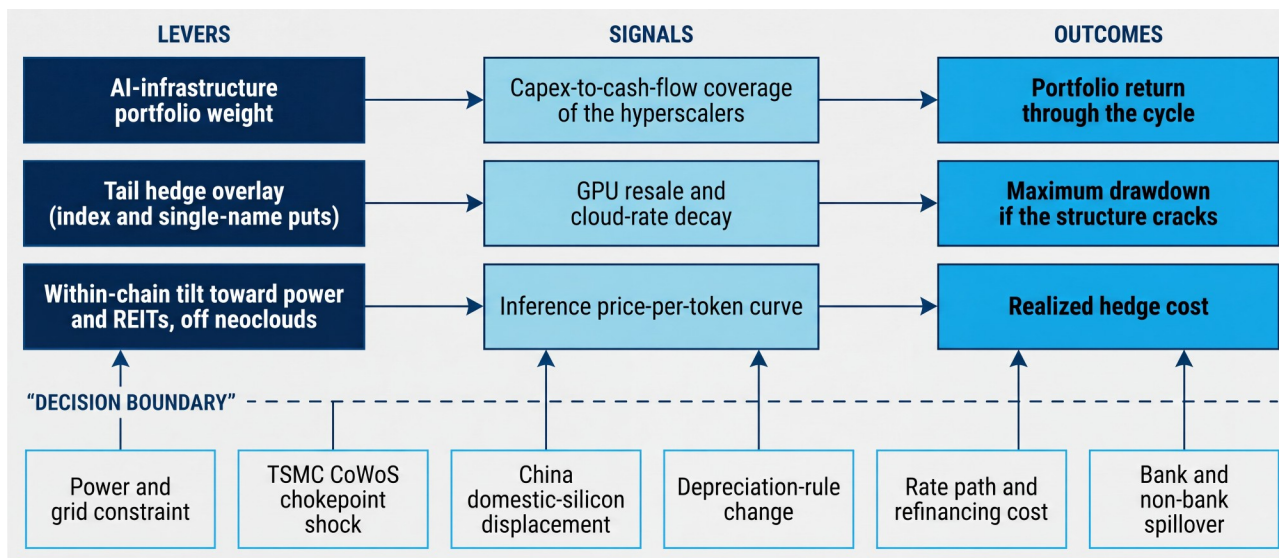
a disclosed chip-put book totals about 7.7 billion notional, and five tripwires flip hold to cut.

Source: NVIDIA Q4 FY2026 results; BIS Quarterly Review, March 2026; Princeton CITP, the \$300B question; required-revenue figures are JPMorgan and Bain estimates

Three levers we control set return and drawdown; six externals we do not set the tripwires

Exhibit 1

Levers the fund controls, leading indicators, and outcomes; externals enter below the boundary as tripwires



Source: Huegoo decision model; calibrated on the BIS Quarterly Review, March 2026 and the Chicago Fed AI tail-risk note

First, the case for holding: the demand is contracted, not hoped-for

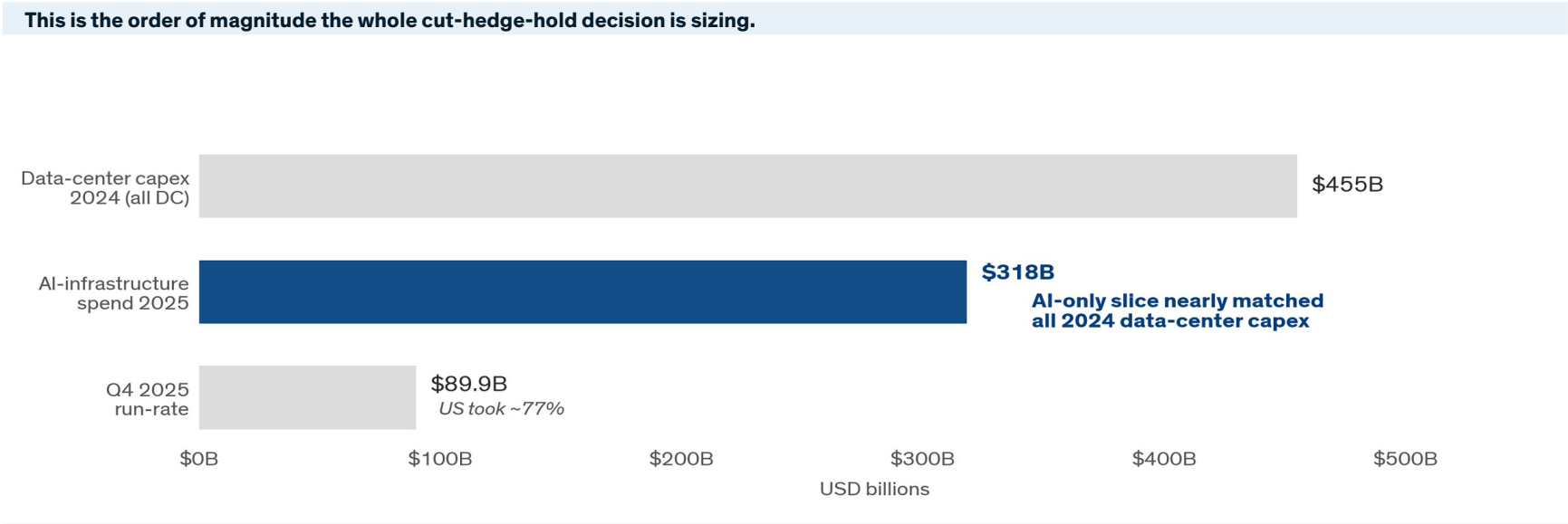
Situation. The fund holds heavy direct and passive AI-infrastructure exposure into a build-out the largest hyperscalers guide near 700 billion dollars in 2026.

Complication. The spend roughly doubled year on year and reached about 2 percent of US GDP, so a skeptical committee asks whether the demand under it is real or hope.

Question. Is the demand under the build-out contracted and real, or a spot-buying spike that unwinds?

Global AI-infrastructure spend more than doubled to \$318 billion in 2025

Exhibit 2 Data-center and AI-infrastructure capex, USD billions, with the Q4 2025 run-rate



Source: Dell'Oro via PR Newswire, \$455B 2024; IDC AI Infrastructure Tracker; 2024 figure is all data-center capex, 2025 is AI-specific

¹ 2024 is total data-center capex (Dell'Oro); 2025 is AI-specific infrastructure (IDC), not a like-for-like series.

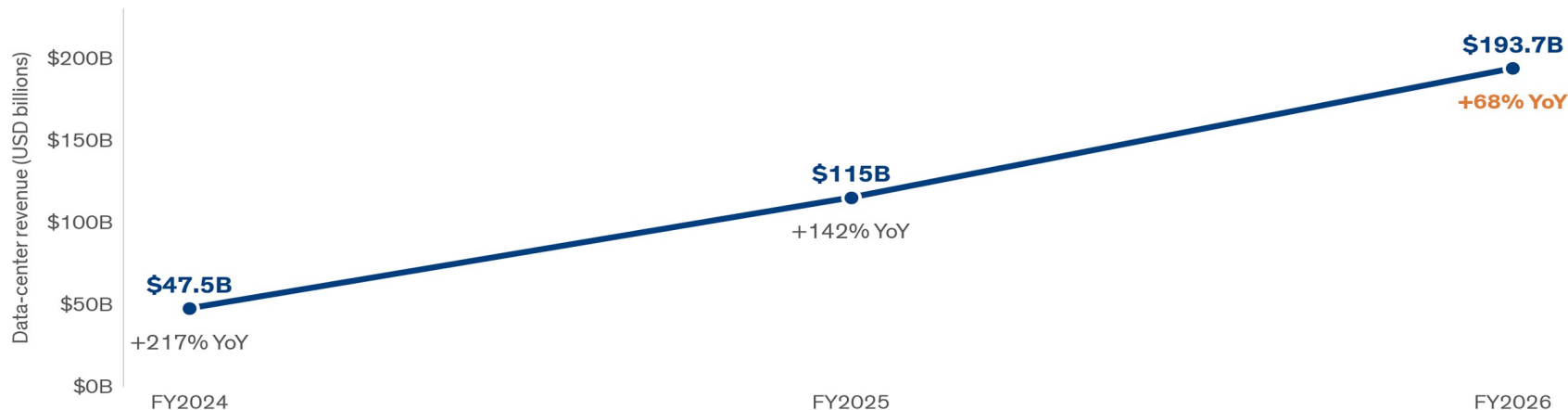
² Q4 2025 run-rate of \$89.9B, with the US taking about 77%.

Nvidia data-center revenue ran from \$47.5B to \$193.7B in two fiscal years

Exhibit 3

NVIDIA data-center segment revenue, USD billions, fiscal years 2024 to 2026

The revenue is real, and the deceleration from 217 to 68 percent is the first crack to watch.

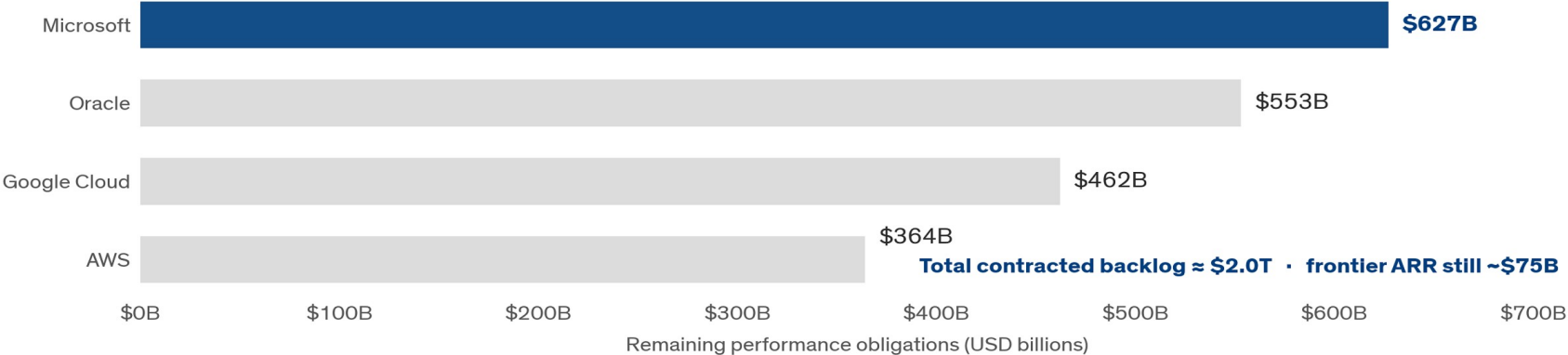


Source: NVIDIA Q4 FY2024 results; NVIDIA Q4 FY2026 results; about 90 percent of global accelerator spend

Customers contracted about \$2.0 trillion of backlog, and frontier revenue is scaling fast

Exhibit 4 Remaining performance obligations by provider, USD billions, May 2026, with frontier ARR noted

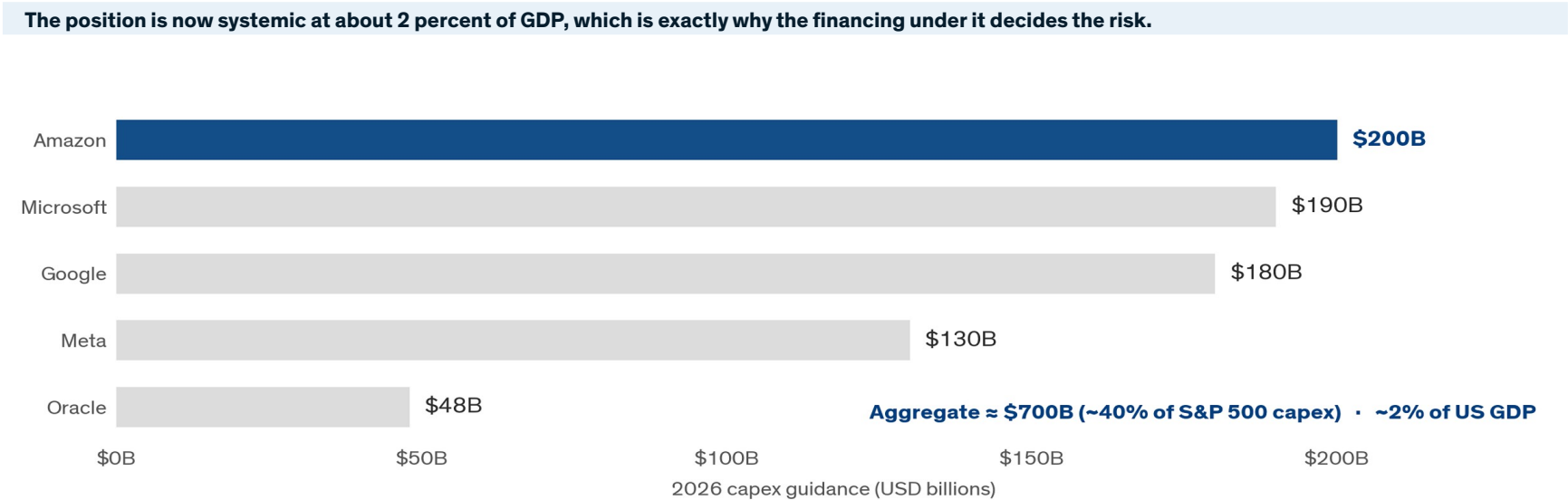
The bull foundation is solid, but the roughly 75 billion of ARR is still far below the revenue the build will require.



Source: Cloud Wars, hyperscaler backlog ~\$2.0T; The AI Corner, Anthropic ARR; token-growth metrics are attributed, not independently re-verified

The largest hyperscalers guide to roughly \$700B in 2026, about 2% of US GDP

Exhibit 5 2026 capex guidance by hyperscaler, USD billions, with the share-of-GDP anchor



Source: Fortune, \$700B with no clear end; Yahoo Finance, \$725B aggregate +77%; Fortune/Pantheon, ~2% of GDP; total varies by basket; the 40%-of-S&P figure is attributed; 75% of growth is single-quarter

¹ Aggregate ranges \$600–725B by basket (CreditSights narrow vs four-name Yahoo Finance); ~\$700B is the midpoint.
² ~2% of US GDP is the verified Pantheon anchor; the ~40%-of-S&P-500-capex share is attributed, not independently re-verified.

Now the case against holding blind: the financing flipped from cash to leverage

Situation. The demand is real and contracted, which is the case for holding the core.

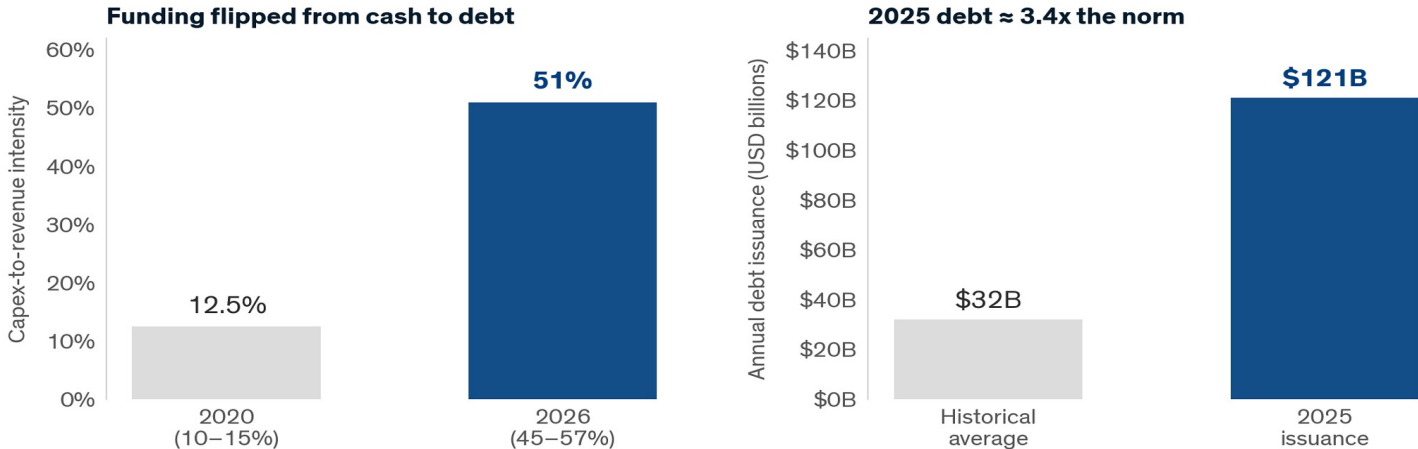
Complication. But the funding flipped from internal cash to 45-to-57 percent capex intensity and roughly 121 billion of debt, a slice of it off balance sheet, and the chips are booked far slower than they age.

Question. What is actually funding the build, how visible is it, and does the revenue repay it?

Capex intensity tripled to 45–57% of revenue, forcing about \$121B of 2025 debt

Exhibit 6 Capex-to-revenue intensity, then versus now, and 2025 debt issuance against the historical average, percent and USD billions

The internal-cash cushion that absorbed past tech cycles is gone, so the bet now carries financing risk.

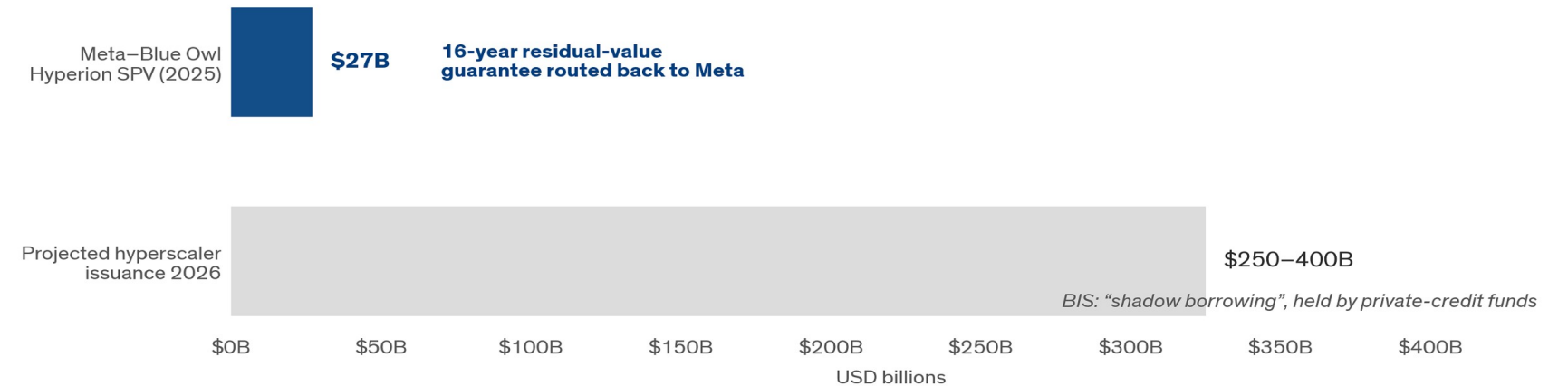


Source: Introl hyperscaler capex and debt; BNY Mellon, record AI debt issuance; MUFG financing the AI supercycle; Amazon free cash flow projected negative in 2026

A growing slice of the debt sits off balance sheet, in private-credit vehicles

Exhibit 7 Verified off-balance-sheet and projected hyperscaler debt, USD billions, 2025 to 2026

Leverage you cannot see on the balance sheet is still leverage the fund owns through its holdings.



Source: BIS Quarterly Review, March 2026; Meta-Blue Owl Hyperion JV; Global Data Center Hub on the +225 bps RVG; the aggregate off-balance-sheet figure is not a quoted BIS number; magnitude is sized on these verified deals

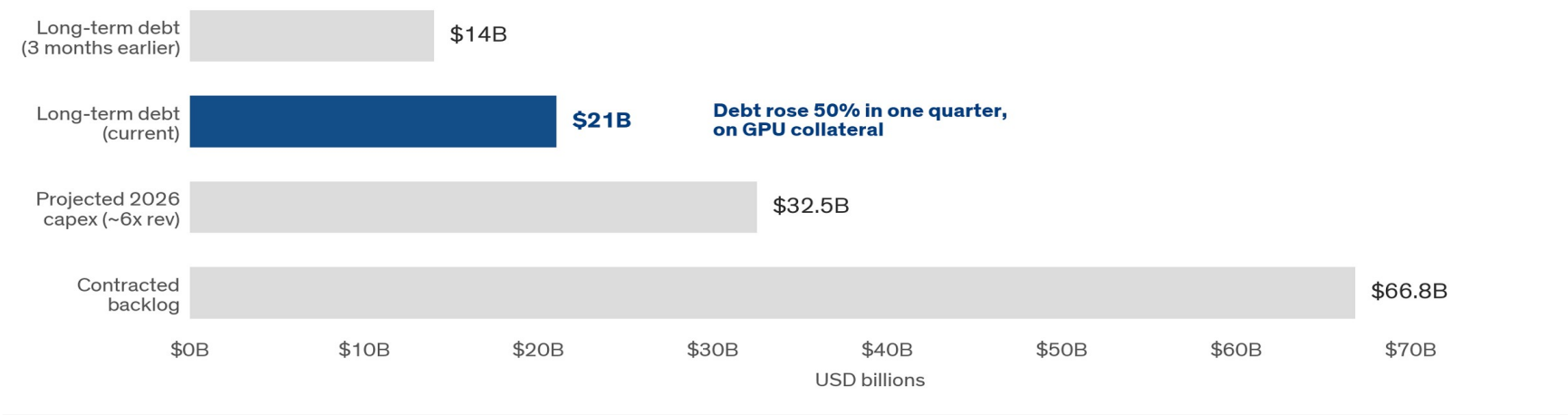
¹ BIS quotes only the mechanism ("shadow borrowing", private-credit-held) and >\$100B 2025 bond issuance; the off-balance-sheet aggregate is sized on the verified Hyperion \$27B and Morgan Stanley \$250-400B figures, not a quoted BIS number.

² The Hyperion vehicle carries a +225 bps spread (~2x Meta's normal cost) and a 16-year residual-value guarantee back to Meta.

Neoclouds are the leveraged edge: CoreWeave carries \$21B of debt on a \$66.8B backlog

Exhibit 8 CoreWeave long-term debt, backlog, and projected 2026 capex, USD billions

This is the part of the chain to tilt away from: collateral that can depreciate faster than the loan amortizes.



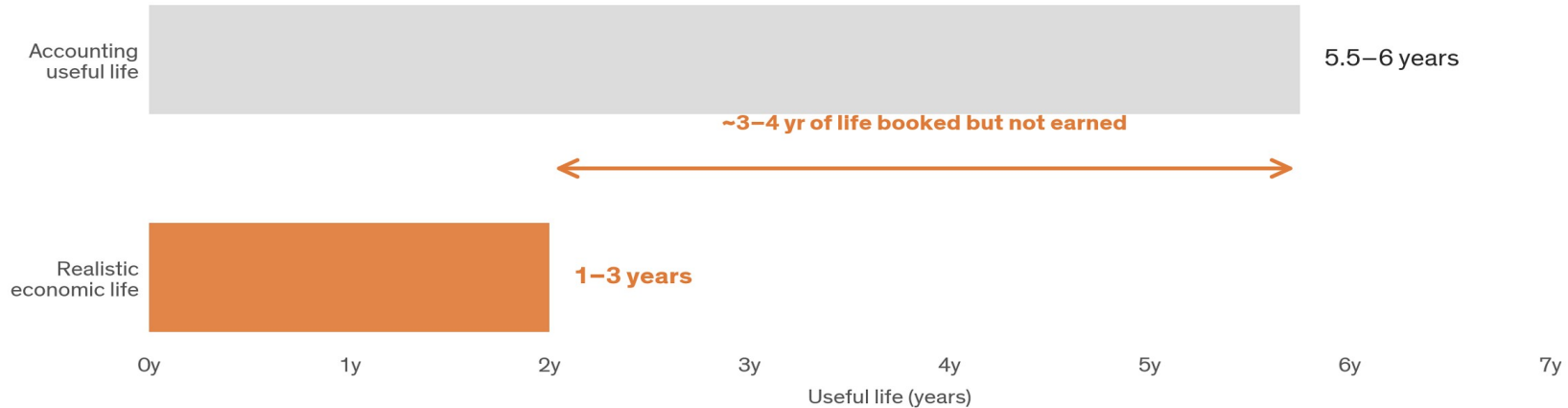
Source: CoreWeave \$8.5B IG GPU-backed facility; Agent Market Cap, CoreWeave debt and backlog; Insurance Journal, JPMorgan securitization \$30-40B

GPUs are booked over 5.5–6 years but obsolesce in 1–3: the \$300 billion question

Exhibit 9

Accounting useful life versus realistic economic life of an AI GPU, years

If the chips age faster than the books say, reported hyperscaler earnings are overstated, which is the crux.



Source: Princeton CITP, the \$300B question; Fortune on Burry, Meta 5.5yr; Deep Quarry, Amazon 6 to 5 years; useful life is contested; Burry's 176 billion understatement is his attributed estimate

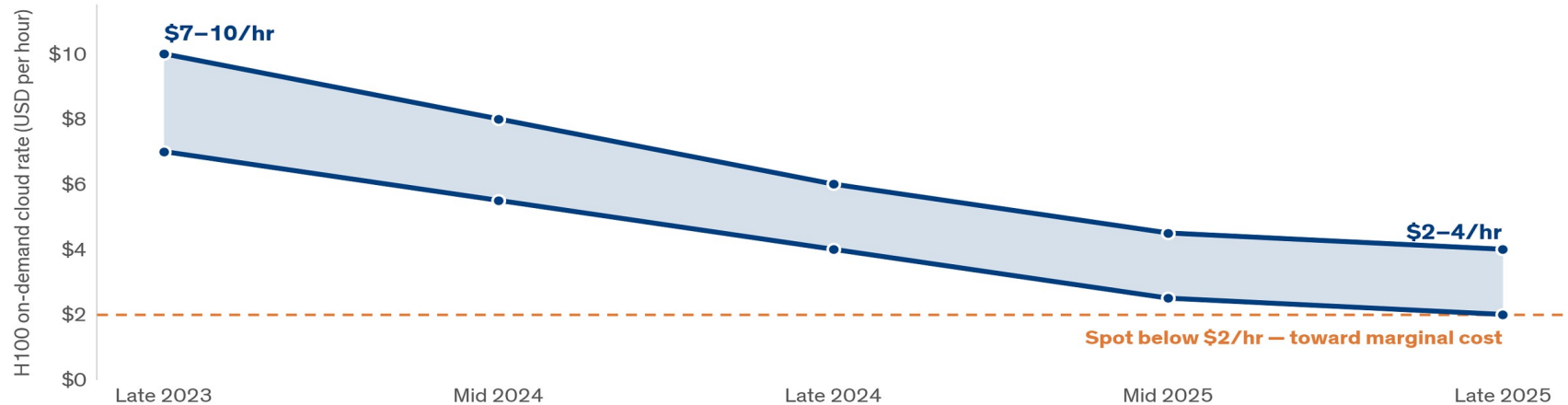
¹ Useful life is the contested crux: Princeton puts economic life at 1–3 years against 5.5–6-year books; the \$300B is the 2025 capital at risk, not a quantified accounting subsidy.

² Burry's ~\$176B 2026–28 understatement is his attributed estimate; the schedule changes (Meta to 5.5yr, Amazon to 5yr + \$920M Q4-24 charge) are disclosed facts.

The secondary market already prices it: H100 cloud rates fell from \$7–10 to \$2–4 an hour

Exhibit 10 H100 on-demand cloud rate, USD per hour, late 2023 to late 2025

The market is voting that GPU economics decay on the 1-to-3-year clock, not the 5-to-6.

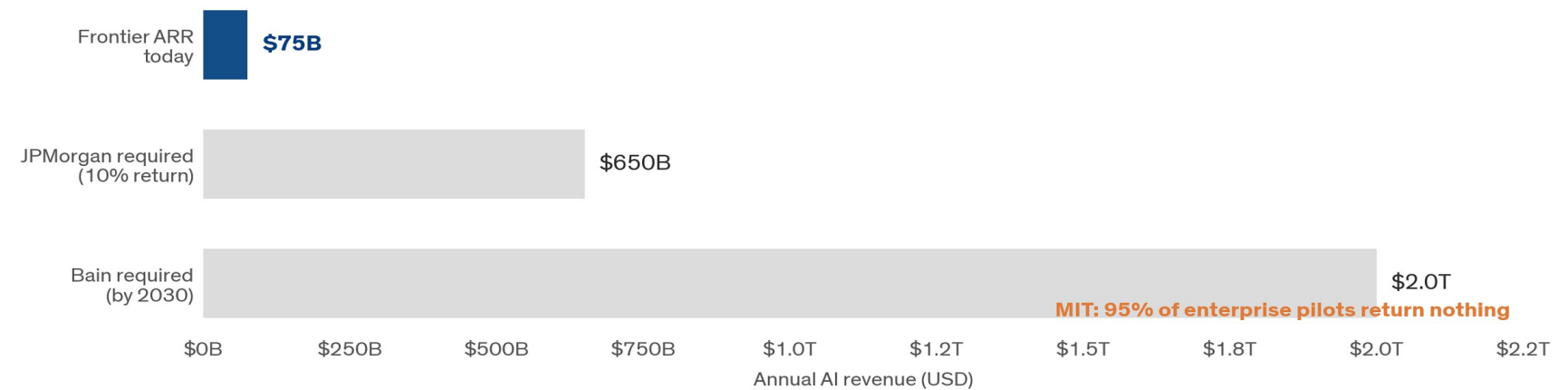


Source: Hashrate Index, used-GPU and cloud-rate pricing; secondary H100 fell from about a 50,000 dollar peak toward 6–15,000

Required revenue of \$650B to \$2T dwarfs the ~\$75B of AI revenue earned today

Exhibit 11 Annual AI revenue required for a return versus combined frontier ARR today, USD billions

The build is priced for revenue 9 to 30 times today's, a wide gap that the bear says will not close.



Source: Tom's Hardware, JPMorgan \$650B; Bain \$2T by 2030; Fortune/MIT, 95% of pilots; required-revenue figures are named-house estimates, not consensus

¹ Required-revenue figures are named-house estimates: JPMorgan \$650B/yr in perpetuity for a 10% return; Bain \$2T by 2030. Not a consensus number.

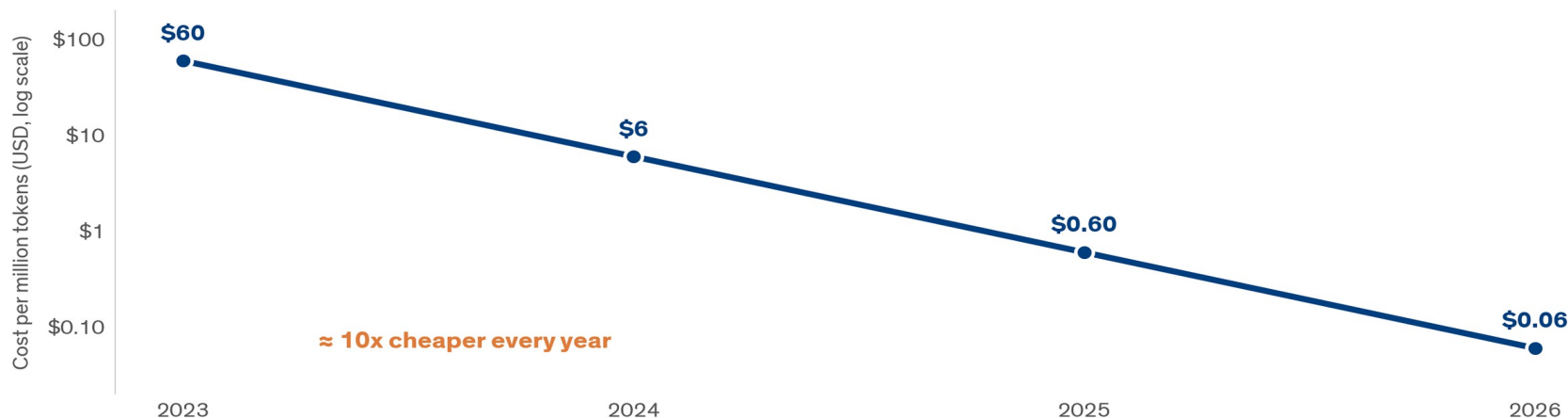
² ~\$75B is combined frontier ARR today; MIT finds 95% of enterprise GenAI pilots return no measurable ROI.

Inference prices fell about 1,000x in three years: good for use, hard for payback

Exhibit 12

Inference cost per million tokens at constant quality, USD, log scale, 2023 to 2026

The same force that wins users shrinks the margin that repays the leverage behind the build.

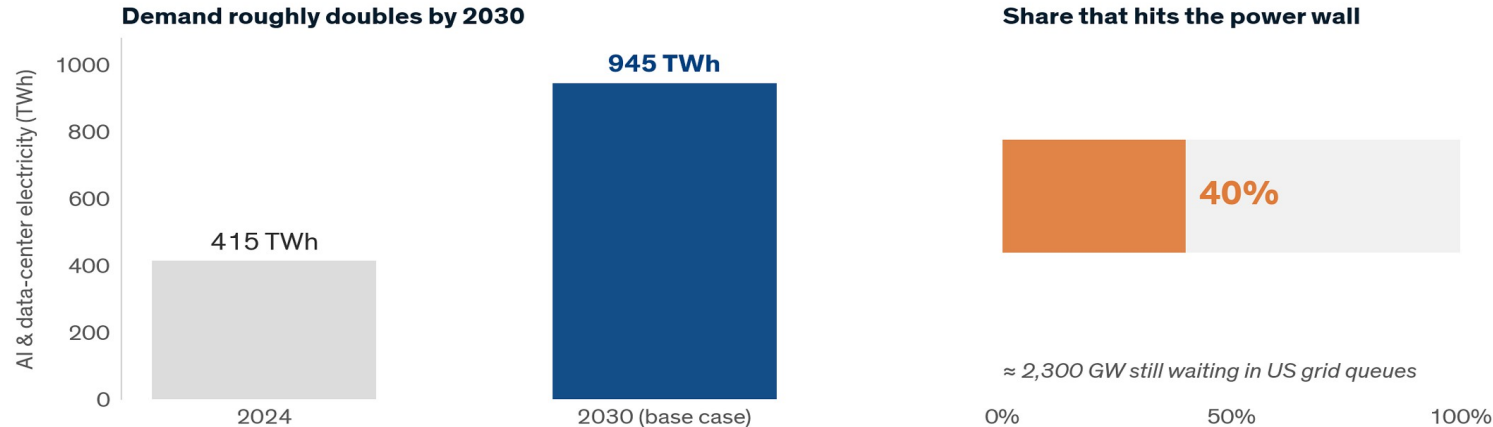


Source: a16z, LLMflation; Mirantis, inference about 80% of opex; about 10x a year, roughly 1,000x over three years at constant quality

Power is the binding constraint: 40% of AI data centers are power-limited by 2027

Exhibit 13 Global AI and data-center electricity demand and the share of capacity power-constrained, TWh and percent

Power, the grid, and memory are uncontrollable externals the allocator must watch as tripwires, not solve.



Source: IEA Energy and AI; Data Center Knowledge, Gartner 40% constrained; LBNL Queued Up; memory sub-figures and the capacity-slip estimate are attributed; the US 2026 share is directional

History cuts both ways, the tail is real, and the hedge is already priced

Situation. Demand is real and the financing is fragile, and every capital boom claims a precedent that flatters it.

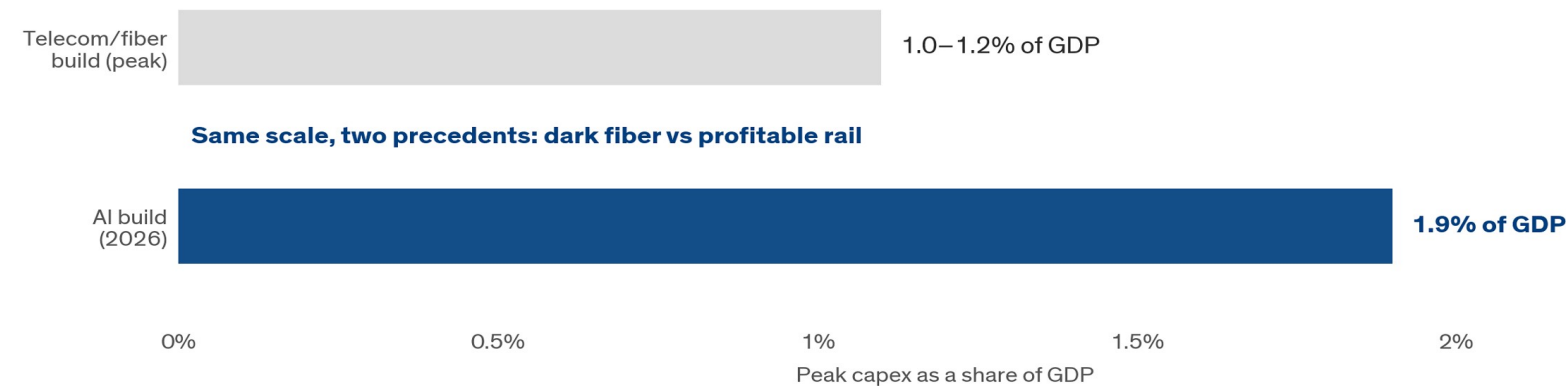
Complication. The telecom build busted while railroads and the cloud paid off, and new tail risks emerged in banks, in one packaging chokepoint, and in China, which shifted the burden onto a hedged, tripwired position.

Question. Which precedent is this, what flips the bet, and what do we actually do?

AI capex outweighs the telecom bust as a share of GDP, yet railroads overbuilt and still paid off

Exhibit 14 Peak capex as a share of GDP, telecom build versus AI build, percent

Overbuild is not the same as wasted: the question is which precedent this is, which is why the call hedges rather than cuts.



Source: 7GC, AI capex and the telecom bubble; Wikipedia telecoms crash; FHWA railroad returns ~15%; LongYield, cloud paid off

¹ Bear precedent: telecom left 85–95% of laid fiber dark, ~\$1T of debt recovered at ~20%. Bull precedent: railroads overbuilt yet earned ~15%, and the 2008–23 cloud build paid off.

**Real demand. Fragile
structure. The decision is the
gap between them.**

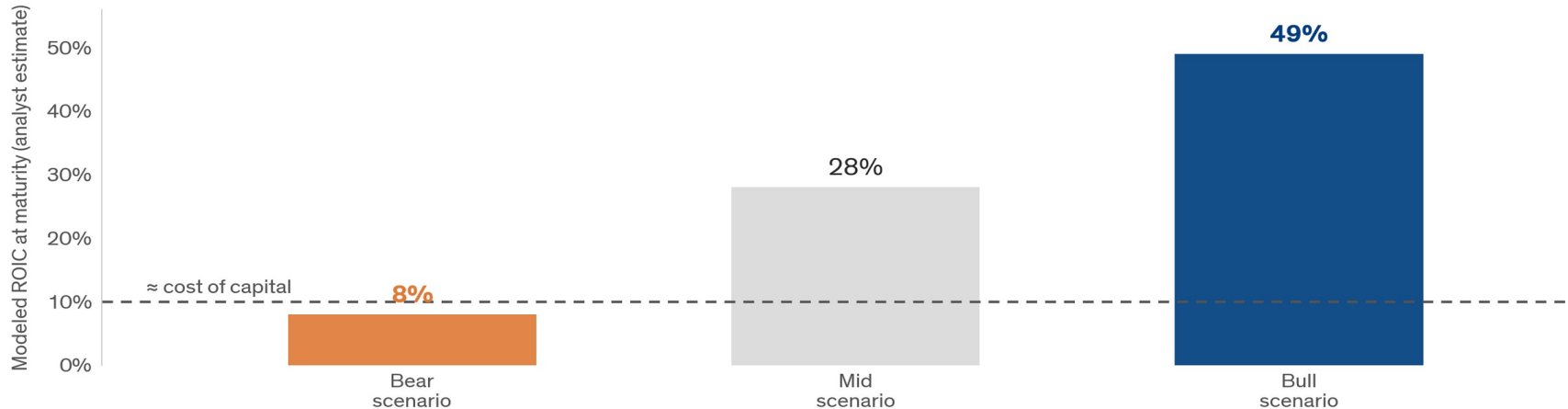
Hold the core, hedge the tail, and tilt off the leveraged edge, with tripwires that flip hold to cut.

The bull's forward case: inference monetizes the build, with ROIC modeled at 8% to 49%

Exhibit 15

Modeled AI-infrastructure ROIC at maturity, bear, mid, and bull scenarios, percent

The bull case is a real range, not a fantasy, but its floor of about 8 percent sits below the cost of capital, so the tail must be hedged.



Source: Computerworld, inference 50% to 67%; InvestorPlace ROIC 8/28/49%; Goldman tracking trillions; ROIC and trajectory are analyst estimates, not independently verified

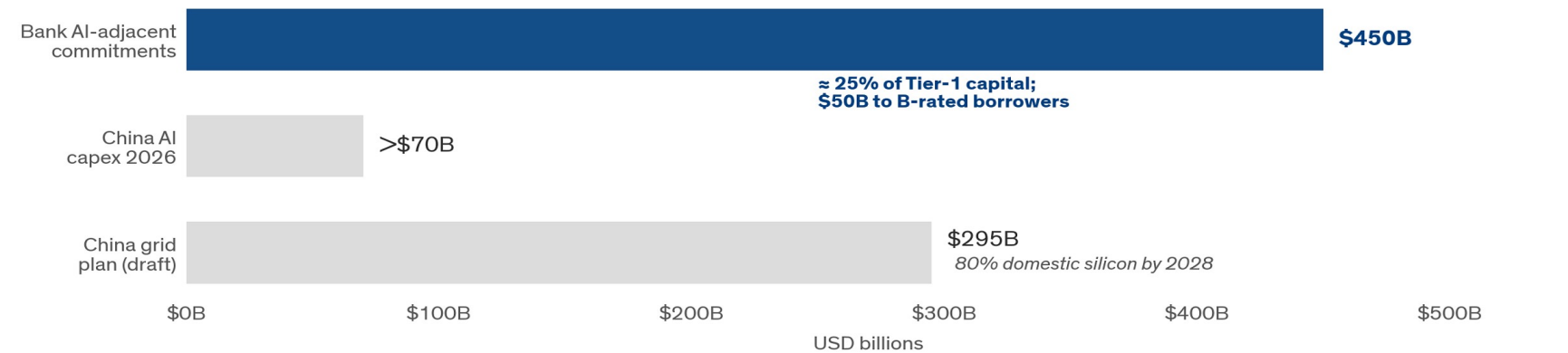
¹ ROIC scenarios (~8% bear / ~28% mid / 49% bull) are analyst estimates at maturity, not independently verified.

² Forward case assumes inference rising 50% ~67% of compute and TSMC 3nm staying booked through 2028.

The leverage has entered the banking system, and China is building a walled-off stack

Exhibit 16 Bank AI-adjacent commitments versus China AI data-center capex, USD billions

The leverage is not contained to tech, and a walled-off Chinese pool caps the bull ROIC fan.



Source: Chicago Fed, AI tail risk for banks; Central Banking on the study; Tom's Hardware, China \$295B plan; the \$1.2T AI-debt and 45% survey figures are attributed; China's \$295B is a draft plan

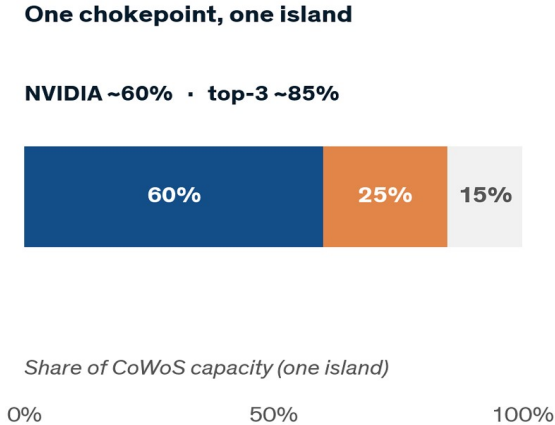
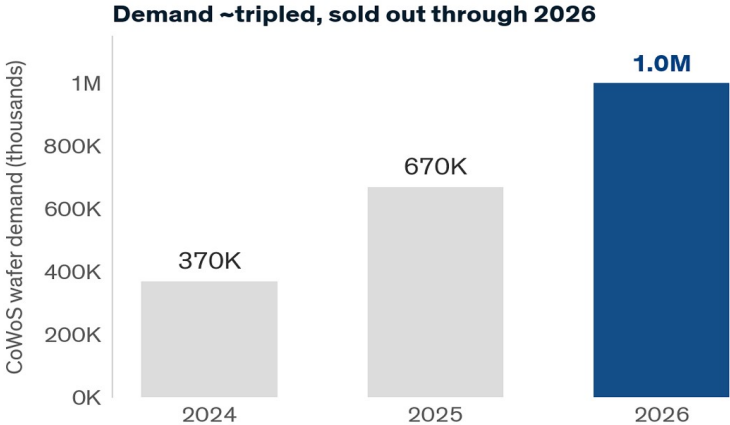
¹ Bank C&I commitments ~\$450B (~25% of Tier-1 capital) are corroborated on live secondaries; the ~\$1.2T AI-debt and 45%-top-tail-risk survey figures are attributed to JPMorgan and BofA.

² China's \$295B / 80%-domestic-by-2028 grid is a draft national plan, not committed spend.

Every frontier accelerator funnels through one packaging line, on one island, sold out through 2026

Exhibit 17 CoWoS advanced-packaging wafer demand and concentration, thousands of wafers and percent

A Taiwan-centered shock hits all hyperscaler accelerator supply at once, which is the un-hedgeable tail the basket put exists for.



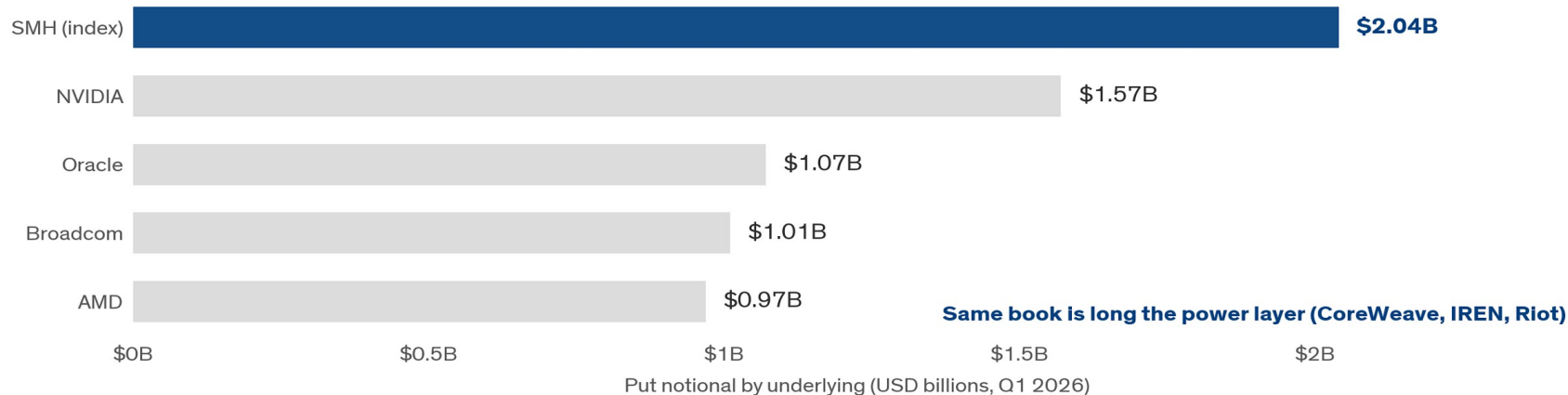
Source: Silicon Analysts, CoWoS allocation Q1 2026; CNBC, Nvidia books packaging capacity; CoWoS demand and shares are TrendForce and Morgan Stanley estimates

The hedge and tilt are already real money: a disclosed ~\$7.7B chip-put book

Exhibit 18

Disclosed chip-complex put positions by underlying, USD billions notional, Q1 2026

Index puts such as SMH and QQQ are the cleanest tail hedge per dollar; single-name puts target the depreciation-exposed edge.



Source: Benzinga, Aschenbrenner put book; Quiver Quantitative, Situational Awareness 13F; itemized positions total about \$7.7B notional; the round \$8.5B figure includes unitemized positions

Hold the core, hedge the tail with index puts, and tilt toward power and REITs

THE CALL

- Hold the demand-backed core: \$2.0T of contracted backlog (C1–C7).
- Hedge the tail with index and sector puts (SMH, QQQ), single-name puts on the depreciation-exposed edge (C26).
- Tilt toward contracted-power and data-center-REIT names, off the leveraged neoclouds (C9–C13).

WHAT WE REJECT

- Cut the sleeve wholesale: it forfeits a demand-backed bet to avoid a hedgeable financing risk.
- Hold blind, unhedged: it leaves the financing and depreciation tail uncapped.

Source: Huegoo decision synthesis on the dossier-verified evidence base; positioning precedent from the Situational Awareness 13F

Five tripwires flip hold to cut, each tied to an external we already watch

Exhibit 19

The signal, its 2026 level, the trigger, and the pre-committed action

Signal	2026 level	Trigger	Pre-committed action
Power constraint	Gartner: 40% of AI data centers power-limited by 2027	≥40% confirmed power-constrained	Cut single-chokepoint-exposed names; let the index put run
TSMC / CoWoS	Sold out through 2026; NVIDIA ~60%, top-3 ~85%	Any Taiwan-centered CoWoS shock	Flip hold to cut on the leveraged-neocloud edge
Depreciation	Booked 5.5–6 yr vs 1–3 yr economic life	Material accelerated charge or H100 below \$2/hr	Cut the depreciation-exposed edge; re-size the tail hedge up
Bank spillover	Bank C&I ≈ \$450B, ~25% of Tier-1 capital	AI debt spills into bank stress	Cut the most leveraged exposure; lift the index put
China displacement	Draft \$295B grid, 80% domestic silicon by 2028	Plan funded and accelerating	Trim names most exposed to the walled-off pool

Source: Gartner via Data Center Knowledge; Silicon Analysts CoWoS; Chicago Fed bank tail risk; thresholds are Huegoo planning calibrations

The recommendation is bounded: each failure mode has a mitigation and a reopening trigger

Exhibit 20

Failure modes, why each is bounded, the mitigation, and what reopens the decision

Failure mode	Why it is bounded	Mitigation	Reopens the decision
False alarm; hedge carry idles	Carry is $\sim 0.5\text{--}1.5\%$ of the sleeve a year	Resize the put overlay at the quarterly review	Two false fires in 12 months recalibrate the dashboard
The bull runs away	You still hold the demand-backed core	The tilt keeps the contracted-backlog upside	Core underperforms the unhedged sleeve over a full cycle
A tail fires	The index and single-name puts pay	Tripwires flip hold to cut before the drawdown compounds	Drawdown exceeds the hedged bound
Regime change in financing	Tripwires escalate it as systemic, not sector	Lift the index put; cut the most leveraged exposure	Bank or SPV refinancing stress confirmed

Source: Huegoo decision synthesis; hedge-cost and drawdown bounds are planning estimates calibrated on the disclosed put book

Approve the hold-hedge-tilt program this quarter, with quarterly tripwire review

Exhibit 21

The four moves of the hold-hedge-tilt program; costs are Huegoo planning estimates

Move	Owner	When	Cost
Size the AI-infrastructure sleeve to a target weight: hold the demand-backed core, trim the most leveraged neocloud exposure	CIO	This quarter	Within mandate
Put on the tail hedge with index and sector puts (SMH, QQQ) sized to the sleeve, and single-name puts on the depreciation-exposed edge	Head of Derivatives	This quarter	~0.5–1.5% of sleeve / yr
Execute the within-chain tilt toward contracted-power and data-center-REIT names, away from leveraged GPU-collateral neoclouds	Portfolio Manager	Next rebalance	Turnover only
Arm the tripwire dashboard (power, CoWoS, China, depreciation, bank spillover) and review the bet against it quarterly	Head of Risk	Quarterly	\$0.3M / yr

Source: Huegoo decision synthesis; hedge cost is a planning estimate; instrument precedent from the Situational Awareness 13F

The three hardest objections an IC will raise, answered before they come up

Pre-empted before they come up; see the backup exhibits where indicated

If the financing is this fragile, why not just cut the whole sleeve now?

Because demand is real and contracted, Nvidia near 194 billion dollars and about 2 trillion of backlog, and the bust precedent is matched by paid-off ones: railroads earned about 15 percent. A wholesale cut forfeits a demand-backed bet to avoid a hedgeable financing risk.

The bubble call has been wrong for two years, so why is the depreciation argument different now?

Because it is now in the data, not the rhetoric: GPUs are booked over six years against a one-to-three-year life, the secondary market moved H100 rates from 7-to-10 to 2-to-4 an hour, and two hyperscalers reset schedules in opposite directions in one year.

If the worst tail is a TSMC or Taiwan shock we cannot diversify around, what good is a hedge?

That is exactly why the hedge is index and sector puts, not a within-chain reshuffle. A single CoWoS chokepoint, Nvidia about 60 percent and the top three about 85 percent, hits the whole sleeve at once, so the cleanest protection per dollar is a basket put.

Source: Dossier corpus ai-capex-bubble-when-it-pops; see backup exhibits where indicated

THE ASK

The AI build-out is a real, demand-backed bet wrapped in a fragile financing and accounting structure, so hold the core, hedge the tail, and tilt off the leveraged edge, with named tripwires that flip hold to cut.