

THEMATIC RESEARCH · AI INFRASTRUCTURE & CAPITAL RETURNS

The AI Capex Payback Clock

Four Companies Will Spend Roughly \$700bn on AI Infrastructure This Year. The Arithmetic of What That Capital Must Earn to Clear Its Cost — and the One Condition Under Which It Ever Does

TON618 CAPITAL RESEARCH · AS OF JULY 19, 2026 ·

Every figure is drawn from the companies' own SEC filings, earnings disclosures, or dated public reporting, and is separated throughout into fact ("the filing says...") and estimate ("our model implies..."). The payback model is exactly that — a model; §7 states plainly which inputs are observed, which are inferred, and which are stress assumptions. This note is the demand-side companion to our July 17 restatement The Depreciation Question; the two share one engine — the gap between the life the hardware keeps and the life the accounting assumes. It is also the demand-side foundation of our five-note AI-infrastructure series; the reader's guide maps how the five connect and is the place to start.

§0 The Argument

~\$137bn

THE 2026 SHORTFALL: ANNUAL
OPERATING PROFIT THE YEAR'S
AI CAPEX MUST EARN TO CLEAR
ITS COST OF CAPITAL ON A
THREE-YEAR CLOCK, LESS
EVERY DOLLAR OF AI PROFIT WE
CAN CREDIT — MERCHANT AND
INTERNAL

The claim. Microsoft, Alphabet, Amazon and Meta have guided to roughly **\$700bn of capital expenditure in 2026**, up from about \$420bn in 2025 and \$230bn in 2024 (guide/tracker basis) — among the steepest concentrated capital build-outs in corporate history, faster than either the 1990s telecom build or the 2010s cloud expansion. To clear an 8.5% cost of capital *and* replace itself as the silicon wears out on the three-year clock the hardware actually keeps, the 2026 vintage alone must generate on the order of **\$221bn of annual operating profit** — equivalently, **\$440–740bn of incremental annual revenue** depending on margin. Against that, the disclosed and inferable AI revenue actually reaching these four income statements is on the order of **\$80bn**. Credit a generous, unbooked **\$60bn** for internal AI returns embedded in ads, search and productivity, and the vintage still falls roughly **\$137bn a year short**. The spend is not fraud, not hidden, and not obviously irrational. The question this note answers is narrower and harder: on the companies' own numbers, what has to be true for it to pay back — and when.

Capital expenditure is a promise. A dollar spent on a data center is an assertion that the dollar will come back with a return on top, over the life of the asset. For most of the history of these four companies that promise was kept quietly and comfortably: they invested a modest fraction of prodigious cash flow, and the returns were never in doubt. What has changed since 2023 is not the direction of the spend but its scale relative to the return that has so far appeared. The capital is arriving in full. The revenue is arriving in part.

We are not forecasting a crash, and we do not assert the promise will be broken. The most-asked version of this question — *"when does the AI trade blow up"* — is the wrong one, because it demands a date the data cannot honestly supply. The answerable version is mechanical: **how much must this capital earn, how much is it earning, and under what path do those two lines meet.** That calculation has a clean and, we think, underappreciated property. Under continued capex growth the two lines never meet — you cannot outrun your own capital deployment. They meet in exactly one family of futures: the one in which capex growth stops *and* AI revenue keeps compounding at close to its current pace. The payback case and the plateau case are the same case. That is the finding.

~\$700bn

Big-four 2026 capex guidance, up ~70% on 2025 and ~3x on 2024

\$221bn

Operating profit the 2026 vintage must earn annually to clear WACC + 3-yr economic depreciation

~\$80bn

Merchant AI revenue plausibly reaching the four P&Ls in 2026, net of intra-ecosystem double-count

2031, at earliest

First year the gap closes — and only if capex freezes now and AI revenue compounds ~70%/yr decaying

§1 The Tape

The denominator is not in dispute. The four companies have told us what they intend to spend, and the run-rate already confirms it.

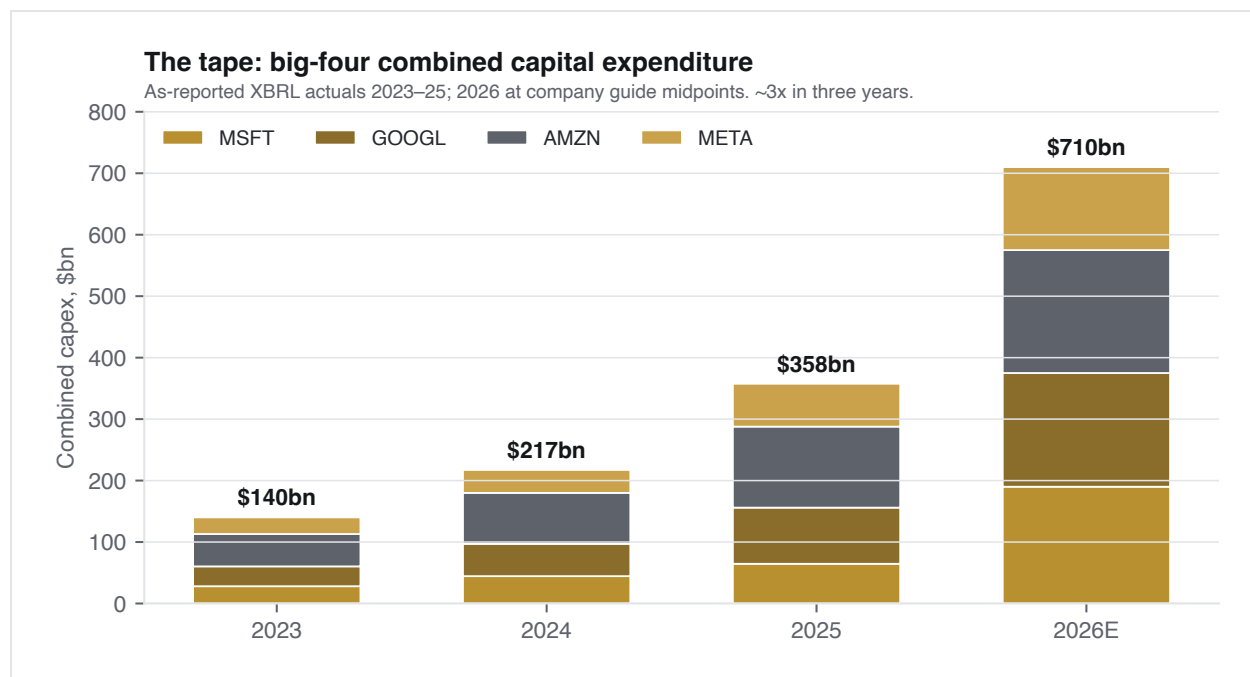
Two bases circulate for this number, and we keep them distinct. Summed from each company's own XBRL cash-flow facts *on its own fiscal year*, big-four capex was **~\$217bn in 2024** and **~\$358bn in 2025** — the bottom-up figures in the table below. The widely-cited tracker figure is higher, roughly **\$410–420bn for 2025**, because it calendarizes Microsoft's June year-end and includes finance-lease capex; that is the basis on which the four have guided **2026 to a combined \$700–725bn** — Microsoft to about \$190bn, Amazon near \$200bn, Alphabet \$180–190bn, Meta \$125–145bn. Independent trackers span the same guide band — roughly **\$620bn (RBC) to \$725bn (consensus, up ~77% on the ~\$410bn 2025 base)**, with Futurum near \$690bn. We use the guide/tracker basis for the forward comparison (it is what the guides are quoted on) and the bottom-up XBRL for the as-reported table; the reconciliation matters less than the direction, which is the same on either basis: sharply up.

The as-reported vintages, pulled directly from each company's XBRL cash-flow facts, show where the acceleration is concentrated:

AS-REPORTED CAPITAL EXPENDITURE, \$BN (SEC XBRL; AMZN = PURCHASES OF PROPERTY & EQUIPMENT INCL. FINANCE-LEASE BUILD)

COMPANY	2023	2024	2025	LATEST Q, ANNUALIZED
Alphabet (Dec FY)	32.3	52.5	91.4	~143
Amazon (Dec FY)	52.7	83.0	131.8	~177
Meta (Dec FY)	27.0	37.3	69.7	~76
Microsoft (Jun FY)	28.1	44.5	64.6	~124
Combined run-rate	—	—	—	~\$519bn

The final column is the tell. At the pace of the most recently reported quarter, the four are *already* spending at an annualized ~\$519bn, before the year's guided step-up fully lands. Oracle, on a May fiscal year, sits just outside this set and reported **\$55.7bn of FY2026 capex** — a separate name spending at hyperscaler scale, and the bridge to our depreciation note.



Not all of this dollar needs to pay back on GPU time. A data center is roughly two assets welded together: the short-life IT gear inside it — GPUs, servers, networking silicon, superseded on the ~24-month architecture cadence — and the long-life shell around it: land, steel, power substations, cooling, transformers that serve for fifteen years and more. The payback clock runs at two very different speeds on the two halves. Throughout this note we split capex **60% short-life / 40% long-life**; the companies' own PP&E disclosures put the server- and network share of capex near 41–45% (see the depreciation note's wedge work), and adding the balance of IT equipment lifts the short-life bucket to roughly three-fifths. \$7 carries the sensitivity; the qualitative result is not fragile to it.

§2 The Hurdle

A dollar of capital earns its keep only when it returns, every year, the cost of financing it plus the cost of replacing it as it wears out. For short-life AI gear, the second term dominates — and it is the term the accounting understates.

The hurdle is standard corporate finance, applied honestly to the asset lives the hardware actually keeps. For any vintage of capital C , the minimum annual operating profit that justifies it is:

$$\text{required annual profit} = C \times (\text{WACC} + \text{economic depreciation rate})$$

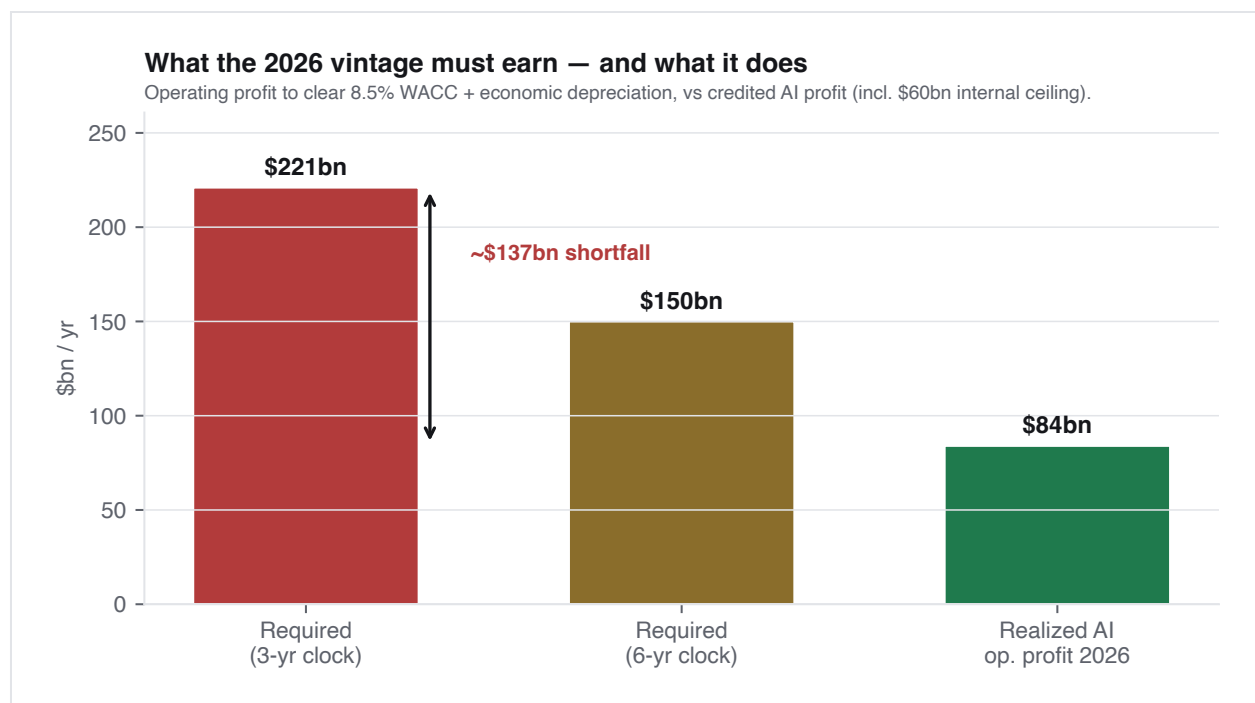
The first term is the cost of the money. The second is the cost of the asset aging — because unlike a bond, a GPU does not return your principal at maturity; it returns nothing, so the capital must earn back its own replacement. With a WACC near **8.5%** and our 60/40 split, the blended hurdle rate depends entirely on the life assumed for the short-life half:

BLENDED ANNUAL HURDLE RATE ON A MIXED AI-INFRASTRUCTURE DOLLAR

SHORT-LIFE (IT) ECONOMIC LIFE	SHORT-LIFE DEP.	BLENDED HURDLE RATE	ON 2026'S ~\$710BN
Three years — the clock the silicon keeps	33%	31.2%	\$221bn / yr
Six years — the clock the books assume	17%	21.2%	\$150bn / yr

The two rows are the same engine as the depreciation note, seen from the other side. There we showed that booking servers over five to six years while they are superseded every two understates depreciation and overstates earnings. Here that same wedge sets the hurdle: on the six-year clock the accounting uses, the 2026 vintage must earn **\$150bn** a year; on the three-year clock the hardware enforces, it must earn **\$221bn**. The **\$71bn difference is the annual cost of the optimistic schedule** — the amount by which the reported books understate what this capital has to produce to be worth owning.

Translate the three-year hurdle into revenue, and the scale becomes legible. At a 30% incremental operating margin — generous for cloud-delivered compute once power and depreciation are loaded — the 2026 vintage needs **~\$738bn of incremental annual revenue**. At an aggressive 50% margin, **~\$443bn**. Either way the requirement is measured in hundreds of billions of *new* annual revenue, for a single year's capital, on top of everything these companies already sell.



§3 The GPU-Dollar

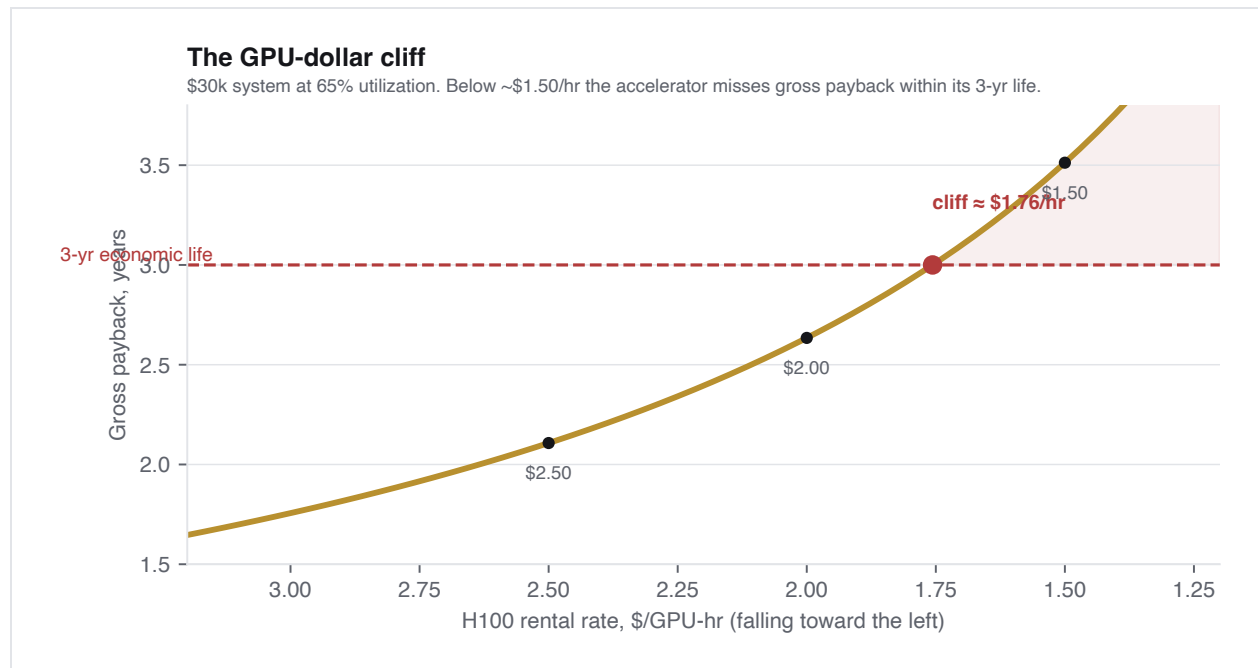
Zoom all the way in, to one accelerator. The macro hurdle is abstract; the unit economics are not, and they are moving the wrong way.

Take a single H100-class accelerator, all-in system cost — the chip plus its share of server, networking, and integration — of roughly **\$30,000**. Run it at a realistic **65% sustained utilization** over a year. What it must earn to pay back within its economic life is a function of one price the market sets in public: the hourly rental rate.

GROSS PAYBACK ON A \$30K GPU SYSTEM AT 65% UTILIZATION, BY RENTAL PRICE

RENTAL RATE	GROSS REVENUE / YR	GROSS PAYBACK	CLEAR 3-YR LIFE?
\$2.50 / GPU-hr	\$14,200	2.1 yr	Yes
\$2.00 / GPU-hr	\$11,400	2.6 yr	Barely
\$1.50 / GPU-hr	\$8,500	3.5 yr	No

These are *gross* paybacks — revenue against hardware only, before the power to run it, the cooling to keep it alive, the data-center shell it sits in, the financing on all of it, and any profit whatsoever. And the rental price has been falling hard. H100 rates are down **64–75% from their scarcity peak**, with market medians now around **\$2.30–3.10/GPU-hr** and floor listings near **\$1.40**; neocloud and marketplace supply routinely prices 3–6× below hyperscaler list. Newer Blackwell-class parts still command a scarcity premium — B200 hyperscaler *on-demand list* rates run to ~\$14/hr on AWS, though open-market *spot* is already ~\$2–5/hr — but the H100 curve is the preview: each generation is scarce, then abundant, then cheap, on roughly the same 24-month clock that supersedes it.



The unit economics and the macro hurdle are the same statement at two scales. When a GPU rents for \$1.50/hr, it does not return its *gross* cost inside the three-year life the silicon keeps — and the macro model's \$137bn gap is that same shortfall, summed across millions of accelerators. The falling-price trend is not a side detail. It is the mechanism by which the payback clock runs down: supply catches demand, price falls, and the revenue side of every vintage compresses while its cost was fixed at purchase.

§4 The Revenue Ledger

The numerator is where the honest work is. AI revenue is real, large and growing fast — and systematically double-counted across the ecosystem. This is the ledger, netted.

The most-disclosed number belongs to Microsoft: an AI business at a **\$37bn annualized run-rate** as of the March-2026 quarter, up **123% year-on-year** and up from roughly \$13bn only fifteen months earlier. That

growth rate is not in question and it is extraordinary. Alphabet and Amazon disclose less cleanly — AI revenue is folded into Google Cloud and AWS growth and commentary rather than broken out — but a defensible read puts Google Cloud's AI-attributable revenue near **\$18bn** and AWS's near **\$16bn** on a 2026 run-rate basis. Meta books **essentially no external AI revenue**: its AI spend serves its own advertising and engagement, which is precisely the internal-ROI question this section turns to below.

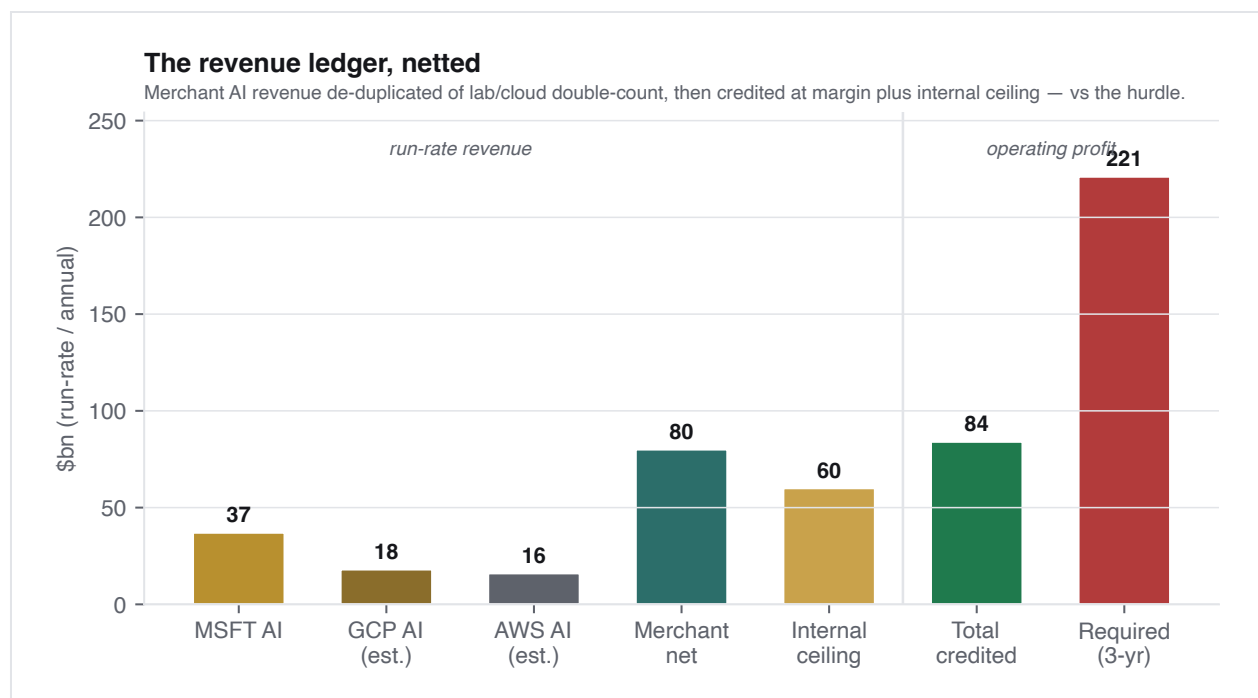
Then there is the downstream check — the frontier labs, whose revenue is the clearest evidence that demand for this compute is real. OpenAI reached roughly **\$24–25bn** of annualized revenue by spring 2026; Anthropic's run-rate ran **~\$30bn in April 2026 and a reported ~\$47bn by May**, with third-party estimates higher still by mid-year — a pace that is itself the clearest evidence the compute demand is real. But here is the trap that inflates every naive AI-revenue tally:

The labs' revenue is, in large part, the clouds' revenue counted a second time. A dollar a customer pays OpenAI is substantially a dollar OpenAI pays Microsoft for compute. Adding "cloud AI revenue" and "lab AI revenue" double-counts the same underlying dollar as it passes through the stack.

THE NETTING PRINCIPLE APPLIED THROUGHOUT THIS SECTION

Netted for that circularity — counting each dollar once, where it lands on one of the four income statements — the **merchant AI revenue actually reaching the big four in 2026 is on the order of \$80bn** of annualized run-rate. That is the honest numerator, and it is the number to weigh against a \$221bn hurdle.

The strongest bull objection, quantified rather than waved away. Much of this capital was never meant to be rented to strangers. It runs Meta's feed ranking, Google's search and ad targeting, Microsoft's Copilot-in-Office — returns that show up as *protected and expanded margins in the core franchises*, not as a line called "AI revenue." This internal ROI is real, and dismissing it would be dishonest. So we bound it rather than ignore it: a **generous ceiling of ~\$60bn a year** of incremental operating profit across the four plausibly attributable to internal AI — roughly the profit on a mid-single-digit-percent uplift to their combined advertising and productivity franchises. It is a ceiling, not an estimate; the true figure is smaller and largely unverifiable from outside. Crediting the full ceiling is the most bull-friendly assumption in this note.



Stack the numerator honestly, then: ~\$80bn of merchant AI revenue at a 30% operating margin is ~\$24bn of operating profit; add the full \$60bn internal-ROI ceiling; the total realized AI operating profit we can credit the four in 2026 is about **\$84bn**. Set against the \$221bn three-year hurdle, the gap is **~\$137bn**; against the \$150bn booked-clock hurdle, **~\$66bn**. Both are the shortfall *after* granting the bull case its best number.

§5 The Clock

Now run it forward. The gap is a snapshot; the clock is the trajectory. It has one property that decides everything.

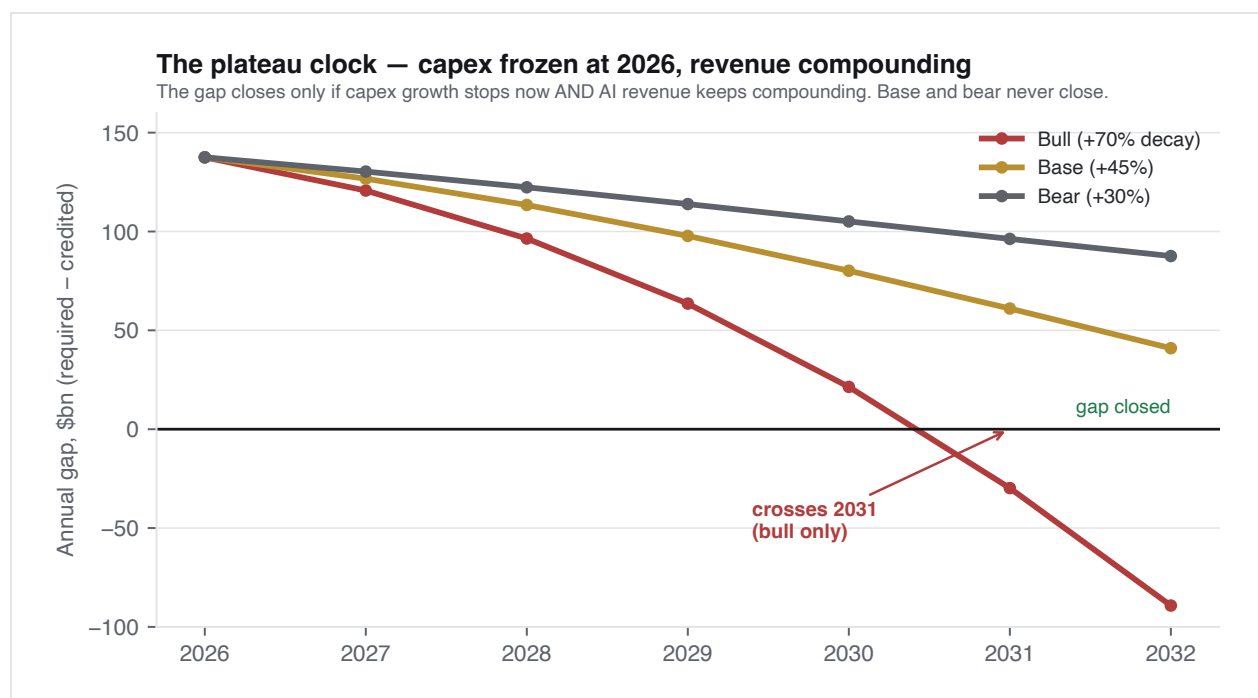
The naive question — *when does compounding AI revenue catch the required return?* — has a clean and slightly counterintuitive answer: **under continued capex growth, never**. If capex keeps compounding at 18–25% a year, the required return compounds with it, and even AI revenue growing at 70% off a small base cannot catch a hurdle that is itself accelerating off a base an order of magnitude larger. You cannot out-earn capital you are still deploying faster than the earnings can grow. In every continued-growth scenario we run, the annual gap *widens* through 2030.

The gap closes in exactly one family of futures: **capex growth stops, and AI revenue keeps compounding**.

Freeze capex at its 2026 level — \$710bn a year, no further growth — and let the revenue engine run:

THE PLATEAU CLOCK — CAPEX FROZEN AT THE 2026 LEVEL, MERCHANT AI REVENUE COMPOUNDING, \$BN/YR

REVENUE PATH	2026 GAP	2028 GAP	2030 GAP	GAP CLOSES
Bull — AI rev +70%/yr, decaying	\$137	\$96	\$21	2031
Base — AI rev +45%/yr, decaying	\$137	\$113	\$80	Not by 2032
Bear — AI rev +30%/yr, decaying	\$137	\$122	\$105	Not by 2032



Read the table honestly and it lands in the middle, not at either pole. **Even in the bull case — capex frozen today, AI revenue compounding at a decaying 70% — the 2026 vintage does not clear its three-year cost of capital until 2031.** In the base and bear cases it does not clear within the horizon at all, even with capex frozen and the

full internal-ROI ceiling granted. And the freeze itself is the catch: capex is not frozen — it is guided *up* another 20%+ next year. The plateau the payback depends on is the opposite of what the companies are doing.

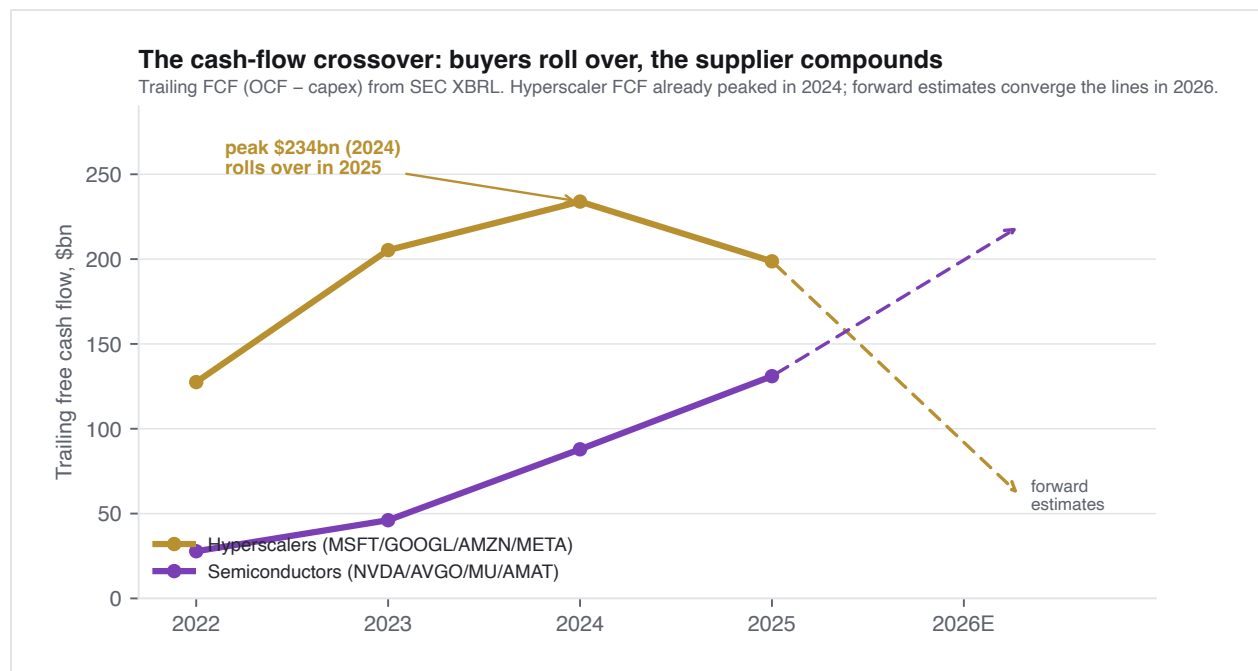
This is the internal contradiction at the center of the trade, stated as neutrally as we can: **the only path to payback runs through a capex plateau, and the capex plateau is the one thing the bull case insists will not happen.** A reader can believe the AI revenue ramp and still see the problem — because the ramp only rescues the economics if the spending that produced it stops growing. A reader can believe the spending is justified by future revenue and still see the problem — because "future" in even the optimistic arithmetic means 2031, five more years of a widening cash gap first.

§6 The Order of Pain, and What to Watch

The clock does not strike for everyone at once. If the gap forces a repricing, it arrives in a sequence, and the sequence is observable before it completes.

The break, if it comes, is not a revenue event — it is a *financing* event. The clock stops not when revenue fails to catch required return (that may take a decade) but when the market stops funding the gap in the meantime. And on this measure the inflection is not a future signpost — it has already begun, and it is visible in the filings.

The cash-flow crossover. Summed from SEC XBRL, the four hyperscalers' combined trailing free cash flow rose from ~\$128bn (2022) to a **peak of ~\$234bn in 2024** — then fell to **~\$199bn in 2025**, its first decline of the buildout, as capex outran even their surging operating cash flow. Over the same window the four largest semiconductor names (Nvidia, Broadcom, Micron, Applied Materials) compounded FCF from ~\$28bn to ~\$131bn. The two lines are converging, and forward consensus (Bank of America / Bloomberg, whose "hyperscaler" set swaps Oracle for Microsoft) has them crossing outright in 2026, with hyperscaler forward FCF collapsing toward zero as the semiconductor line runs past \$200bn. **The capex is not disappearing; it is transferring — from the buyers' cash flow onto the supplier's.** That transfer is the single clearest picture of both halves of this thesis: the demand side (this note) and the supplier that collects it (our companion NVIDIA valuation).



The sharper gauge is capex as a share of operating cash flow — the self-funding ratio. It has inflected up hard: **Amazon reached 0.94 in 2025** (capex nearly the whole of operating cash flow), Alphabet 0.42→0.56, Meta

0.41→0.60, Microsoft 0.38→0.47. On 2026's guided capex — roughly double 2025's — those ratios cross **1.0**, at which point the marginal data center must be financed with debt or off-balance-sheet structures rather than internal cash. Amazon crosses first. The order of pain runs from the outside in:

1. **Neocloud counterparties first.** The pure-plays — the CoreWeaves and their smaller peers — carry the same three-year-life assets against the thinnest earnings and the most leverage, and they rent into the very price curve (§3) that is falling. They break first because they have no ad franchise to cross-subsidize the gap. This is the direct link to our neocloud credit work.
2. **Hyperscaler ROIC and margins next.** The four can fund the gap far longer, but not for free: the cost shows up as compressing return on invested capital and, eventually, as the depreciation wave from today's capex hitting the income statement — the subject of the depreciation note — just as the revenue it was meant to earn is still ramping.
3. **The supplier last, and hardest to read.** NVIDIA sits at the top of the stack collecting the capex as revenue; its order book is the last thing to turn and the first thing the market will scrutinize for the turn. That is the subject of our companion valuation note.

Signposts — observable, and dated where possible. In the spirit of a falsifiable thesis rather than a forecast, these are the things whose movement would confirm or refute the clock, not predictions of when:

- **Capex guidance revisions.** The single cleanest signal. A *plateau* in 2027 guidance is, paradoxically, the *bullish* tell for payback (§5); continued 20%+ growth widens the gap. Watch the direction, not the level.
- **AI revenue disclosure cadence.** Microsoft broke out its run-rate because it was flattering. The quarter a company *stops* disclosing an AI run-rate it previously volunteered is a signpost in itself.
- **GPU rental prices.** A sustained break below ~\$1.50/hr on H100-class parts pushes the median vintage past its gross payback (§3). Public rental indices make this checkable weekly.
- **Neocloud credit spreads and lease terms.** The first-to-break cohort; widening spreads or shortening lease tenors lead the rest.
- **Free-cash-flow prints.** Aggregate hyperscaler FCF has already peaked (2024) and rolled over (2025); the next markers are the self-funding ratio crossing 1.0 (Amazon first, on 2026 guides) and the quarter aggregate FCF turns outright negative — or is held positive only by finance-lease reclassification. That is when the financing event moves from abstract to reported.

§7 What Is Observed, Inferred, and Assumed

A model is only as honest as its disclosure of its own seams. Here are ours.

Observed (from filings and dated disclosure). All capex actuals (SEC XBRL cash-flow facts); the 2026 capex guides (company earnings disclosures); Microsoft's \$37bn AI run-rate and its growth rate; the OpenAI and Anthropic run-rates; the H100 rental-price range and its decline (public rental indices and sector research). These are facts, as-of dated.

Inferred (estimated from partial disclosure). Google Cloud's and AWS's AI-attributable revenue (~\$18bn, ~\$16bn) — these are not broken out and are our estimates from segment growth and commentary. The netted merchant figure (~\$80bn) depends on the double-count adjustment, which is directional, not precise. The 60/40 short/long capex split is anchored to the companies' server-share disclosures but extends them.

Assumed (stress inputs, not claims). WACC of 8.5%; three-year economic life for short-life IT (the depreciation note's central estimate, defended there); 30% incremental operating margin on merchant AI; the \$60bn internal-ROI ceiling; the revenue-growth decay paths in the scenarios. The single most consequential assumption is the three-year life — at six years the hurdle falls from \$221bn to \$150bn and the gap from \$137bn to \$66bn. The gap does not vanish on any life the hardware plausibly keeps; it narrows. **Nothing in the qualitative finding —**

that payback requires a capex plateau — depends on the exact life. It depends only on the required return rising with capex, which is arithmetic.

One fitted parameter, flagged. None. Unlike the depreciation note, this model contains no parameter solved to reproduce a disclosed figure; every input is either observed, independently estimated, or a stated stress. Where an estimate carries a wide band (the inferred revenue lines), the conclusion is tested against both ends in \$5's scenarios.

§8 Sourcing and Method

Capex actuals were pulled from each company's as-reported XBRL cash-flow facts via the SEC's company-concept API (`PaymentsToAcquirePropertyPlantAndEquipment` ; Amazon reported under

`PaymentsToAcquireProductiveAssets` , which includes its finance-lease equipment build and is the correct comparable). Forward guides are from Q1 -calendar- 2026 earnings disclosures and are cited as guides, not actuals. AI revenue figures are company disclosures (Microsoft) or dated public reporting (the labs, Google Cloud, AWS commentary). GPU rental prices are from public rental indices and sector research current to July 2026. The required-return engine, its scenario projections, and every figure in this note are reproducible from the accompanying `engine.py` and `results.json` .

This note makes no price target and initiates no position; it is thematic research into the arithmetic of a capital cycle, and like the depreciation note it is deliberately kept off our signal ledger. It is the demand-side foundation for the supplier-side valuation that accompanies it. The method throughout has been to defer to the disclosure, bound the bull case at its most generous, and report the gap that remains.

Reference data. All filing and market data as of July 19, 2026; capex actuals from SEC XBRL cash-flow facts, forward figures from Q1 - 2026 company guidance and dated public reporting. This note carries no target price and takes no position in any name.

§9 Disclosures

Information only. TON618 Capital. This report is for information purposes only. Nothing here is an offer to sell or a solicitation of an offer to buy any security, fund interest, or digital asset, and nothing here is personalized investment advice or a recommendation regarding any instrument.

Publisher's exclusion. All research is published solely as general, impersonal information of regular circulation. It is not tailored to the objectives or circumstances of any individual and is not issued in connection with compensation from any client. The Fund has no clients and distributes all research free of charge. On that basis it publishes in reliance on the publisher's exclusion from the definition of "investment adviser" under the Investment Advisers Act of 1940 (§202(a)(11)(D); cf. *Lowe v. SEC*, 472 U.S. 181 (1985)).

Registration & conflicts. TON618 Capital is not registered as an investment adviser or broker-dealer in any capacity. The Fund is a Bitcoin fund and may hold or transact in the securities or digital assets it discusses. This note discusses Microsoft (MSFT), Alphabet (GOOGL), Amazon (AMZN), Meta Platforms (META) and NVIDIA (NVDA); the Fund holds no position, long or short, in any of them, and has no economic interest in the price of any security named here. The Fund receives no compensation from any party in connection with its research.

Use of AI. Artificial intelligence is used in the creation of this research. All methodology and data integrity are reviewed and approved before publication by TON618 Capital's Chief Investment Officer, **Keyth Beck**; errors may nonetheless occur, and readers should verify independently.

CFA. This report was prepared to align with CFA Institute analytical standards (methodology only). CFA® and Chartered Financial Analyst® are registered trademarks owned by CFA Institute. That reference describes the

analytical framework applied; it does **not** imply the report was prepared, reviewed, or authored by a CFA charterholder, and the report is **not** issued, reviewed, endorsed, certified, or approved by — nor affiliated with — CFA Institute.

Risk & feedback. Past performance is not indicative of future results. Digital assets and equities are volatile and may result in total loss of capital. Corrections and feedback are welcome — please direct them to CIO **Keyth Beck** at keyth@ton618capital.com.