INTERMARKET UNIVERSE · EST. 2026

IU DIGEST

FLAGSHIP INSTITUTIONAL RESEARCH

JULY 31, 2026

The Economics of AI Infrastructure

What Today's Valuations May not Reflect.

RETURN ON INVESTED CAPITAL

Measuring Efficiency. Driving Value.

COST OF CAPITAL

The Benchmark for Every Decision.

GAP & SPREAD ANALYSIS

Where Value Is Created — or Overlooked.

INSTITUTIONAL GLOBAL MARKET ANALYSIS

iudigest.com

THIS ISSUE

Contents

The Executive Summary and Key Findings open the report; a plain-terms primer on the ROIC–WACC framework — Return on Invested Capital and Weighted Average Cost of Capital — follows for readers new to it. Each analytical section then stands alone — those pressed for time may begin with Section V (the four-hyperscaler Q2 2026 earnings window) or Section VIII (the peer-contagion scenario) without loss of context.

REPORT PARAMETERS

COVERAGE UNIVERSE	MSFT, GOOG/GOOGL, AMZN, META, ORCL; CoreWeave, Nebius, IREN; NVDA supply chain
ANALYTICAL FRAMEWORK	Dual-life EVA matrix; ROIC–WACC sensitivity across four assumptions
EDITORIAL DATE	July 31, 2026

EXECUTIVE SUMMARY

The AI Trade's Blind Spot

The book-versus-economic depreciation gap on AI infrastructure is not a transient accounting artifact hyperscalers will grow out of — it is a structural feature of the compute cycle they have entered. Three independent lines of evidence converge on that conclusion, and none can be dismissed as a personality-driven bear thesis.

First, the depreciation debate is no longer academic: Michael Burry's understated-depreciation estimate ($176B across 2026–2028), footnotebrief's parallel reconstruction (~$228B rounded to ~$200B), and Goldman Sachs' finding that a 5→3-year life shift raises implied annual compute depreciation by nearly $1 trillion (from ~$3T to nearly $4T) all triangulate the same order of magnitude. Second, the engineering reality — Meta's own Llama 3 study implying a ~9% annualized H100 failure rate, GPU rental rates that fell 64–75% from their 2024 peak, and Nvidia's confirmed annual architecture cadence — is inconsistent with the 5–6 year lives most hyperscalers book. Third, the physical bottlenecks that would let hyperscalers "cascade" old chips into productive second lives (power availability, data-center shell reusability) are 5–10 years behind the compute-deployment curve.

The investment implication is not that the hyperscalers are uninvestable — Microsoft, Alphabet, Amazon, and Meta remain among the highest-return franchises in market history, and their core businesses subsidize the AI build. The implication is that single-point ROIC estimates on the AI infrastructure increment are analytically inadequate, and that the marginal 2026 capex dollar is being deployed at a return that is far more scenario-dependent than the sell-side constructive frame admits. Readers new to the ROIC–WACC framework will find a plain-terms primer on page 5.

> *A single parameter — the assumed useful life of a GPU — swings the ecosystem depreciation burden by hundreds of billions to more than a trillion dollars. That is the size of the analytical gap this report exists to bound.*

$433.9B	$149B	~9%
TRAILING FOUR-QUARTER CAPEX, FOUR US HYPERSCALERS	TTM DEPRECIATION RECOGNIZED AGAINST THAT CAPEX	ANNUALIZED H100 FAILURE RATE, META LLAMA 3 STUDY

KEY FINDINGS

Three Signals That Bound the Debate

— The reported **~10–12 percentage-point positive ROIC–WACC spread** on hyperscaler AI infrastructure rests on four separable and individually fragile assumptions — GPU/TPU depreciation life, value-cascade capture, power availability, and data-center shell reusability. Under a *dual-life EVA framework* — Economic Value Added, run in parallel using both the book-reported useful lives and lives closer to the engineering reality — the spread on the marginal 2026 capex dollar compresses from double digits to roughly **200–500 basis points**. That is not a bearish scenario; it is what emerges when engineering-consistent depreciation is applied to a normal revenue path.

— Amazon has already walked its server life back from six to five years (effective January 2025, disclosed February 6, 2025) — a costly reversal that generated a ~$700M operating income hit and a separate ~$920M accelerated depreciation charge, the kind of move a company does not make voluntarily unless its own utilization data leaves no choice. If a second hyperscaler follows, the coordination mechanism that ran in one direction from 2020–2023 reverses. A modeled peer-contagion outcome implies a **1.5–2.5 turn P/E compression** on Alphabet alone — roughly **$300–500B in enterprise value**, with proportionate hits across the peer set.

— The July 22–30 four-hyperscaler earnings window converged on the same picture. Alphabet posted its first-ever negative quarterly free cash flow of **–$5.9B** with capex guidance raised to **$195–205B**. Microsoft's CFO acknowledged on the call that **roughly two-thirds of Q4 capex went to short-lived assets** — a first-party concession that GPUs and CPUs are earning over a much shorter economic life than the six-year book life implies. Meta beat revenue but **missed EPS by 14%**, with depreciation and infrastructure costs pressuring the income statement. And Amazon raised full-year capex from **$200B to $220B** mid-cycle. Every hyperscaler that reported after Alphabet either raised capex further or absorbed an earnings miss; none walked back the AI infrastructure commitment. The book-versus-economic gap is beginning to surface in the tape.

These three signals point the same way. The first sets the size of the gap between what is being reported and what is likely being earned. The second shows how quickly that gap could widen if a single hyperscaler moves. The third shows it is already beginning to appear in the numbers of the group's most disciplined operator. None of this requires a bearish view of AI demand — only that the accounting eventually catch up with how the hardware actually behaves. The market has begun to ask for that reconciliation.

READER'S PRIMER · ¶

How Analysts Actually Measure Value

Every capital allocation decision in corporate finance reduces to a single question: does the return a company earns on the money it deploys exceed the cost of raising that money? The gap between those two numbers — the yield on invested capital minus the price of financing that capital — is the compass by which professional allocators judge whether a business is compounding wealth or eroding it. This primer offers a plain-terms grounding for readers who want to refresh the framework before working through the analytical sections that follow.

Return on Invested Capital (ROIC)

ROIC measures how efficiently a company converts every dollar of capital it has invested — in property, plants, equipment, working capital, and intangibles — into after-tax operating profit. A company that earns 20% ROIC generates roughly $0.20 of after-tax operating profit for every $1.00 tied up in the business. The numerator is operating profit adjusted for taxes; the denominator is the sum of debt and equity capital actually at work in the business. ROIC is the numerator of value creation.

Weighted Average Cost of Capital (WACC)

WACC measures what the same company pays, in blended terms, for the debt and equity it uses to fund itself. Debt has an explicit cost — the interest rate on borrowings, adjusted for the tax deductibility of interest. Equity has an implicit cost — the return shareholders require for the risk they bear, typically estimated from the equity risk premium and the stock's sensitivity to market moves. WACC is the hurdle rate: the minimum return any new dollar of investment must clear to be worth deploying.

The Spread as Compass of Value

The difference between the two — ROIC minus WACC — is the compass of long-term value. When ROIC exceeds WACC, every reinvested dollar creates economic profit; shareholders should want more capital deployed. When ROIC falls below WACC, every reinvested dollar destroys value regardless of how much revenue the deployment generates. In strict financial theory, it is this spread — not revenue growth, not headline earnings per share — that determines whether a business is compounding wealth for its owners or eroding it. A company can grow rapidly and destroy value simultaneously if it is reinvesting billions at returns below its cost of capital.

Why This Matters for the AI Capex Debate

Microsoft, Alphabet, Amazon, and Meta are collectively deploying more than $400 billion per year of capital expenditure into AI infrastructure — GPUs, networking, data-center shells, and power. The consensus institutional view is that this capital is earning ROIC well above WACC: the spread is real, and every dollar of capex is value-creating. The bear thesis, which this report engages seriously, is that the reported spread rests on depreciation accounting that understates the true cost of the capital being deployed. If GPU and networking useful lives are shorter than the balance sheet claims, reported operating profit is overstated, ROIC is overstated, and the spread — the very justification for the capex — narrows or reverses.

Two Terms You'll See Throughout

Two terms recur throughout the report. **Book ROIC** refers to the return calculated using each company's reported financials as they stand today — five- and six-year GPU depreciation lives, current tax treatment, disclosed capex allocations. **Economic ROIC** refers to the same calculation performed with useful-life assumptions closer to the engineering and secondary-market evidence: three-year GPU lives, four-year networking lives, and a residual value on the components once retired. The gap between the two — and how it should be modeled — is the size of the analytical debate this report exists to bound.

SECTION · I

The Reported ROIC–WACC Consensus

28.6%

CONVEQUITY TWO-YEAR FORWARD ROIC, BOOK BASIS

The bull case, stated carefully, is quantitatively serious. Convequity — an independent research firm covering the AI and semiconductor complex, operating outside the big banks — is a useful starting point. Its "AI Bubble Barometer," among the most disciplined boutique frameworks in the space, isolates AI-specific invested capital on the three core hyperscalers (Microsoft, Amazon, Google) and pairs it against AI-attributable contracted revenue, producing a two-year forward ROIC of **28.6%** against an estimated WACC of ~16–18% — a positive spread of roughly 10–12 percentage points on a book basis. Convequity itself notes the exercise "inherently involves a high degree of estimation and uncertainty."

Bessemer Venture Partners' state-of-the-cloud framing supplies the top-down scale check. At an unlevered 15% ROIC, an **$8T cumulative capex build-out** requires roughly **$3T of incremental annual hyperscaler revenue** — about 7x the ~$450B expected in 2026, translating to ~$6T of customer-facing software revenue against ~$1.4T today. Bessemer plausibly frames this as "one of the largest revenue and value creation events in history"; the framing is aspirational, a description of what would need to be true, not what is.

The buy-side sorts into three camps on how to read this. The first — the Convequity base case, InvestorPlace's mid-case decomposition — argues the spread is real and value-creative. The second sees it as substantially an accounting mirage: Michael Burry's November 2025 argument, backed by disclosed put positions on Nvidia and Palantir, is that hyperscalers depreciate GPUs over five to six years when the true economic life is closer to two or three, understating depreciation by roughly $176B across 2026–2028 and inflating reported operating income at names like Oracle (+26.9% by 2028) and Meta (+20.8%). The third camp — Michael Cembalest's *Eye on the Market* and JPMorgan — offers a nuanced defense: longer lives may be partly defensible via the value cascade, but the risk is real and the disclosure inadequate. Cembalest models the income-statement impact of applying three-year lives to GPU and networking capex since January 2023, and finds an earnings hit in the −6% to −8% range for most hyperscalers, larger for Oracle.

Goldman Sachs' *Tracking Trillions* framework quantifies the sensitivity at the ecosystem level. Against a baseline of roughly **$7.6T cumulative AI capex over 2026–2031**, shortening the assumed chip useful life from five years to three raises implied annual compute depreciation by nearly $1 trillion — from about $3T to nearly $4T — while a three-year versus seven-year comparison represents a **$1.76 trillion variance** in the cumulative spending forecast. A single parameter swings the ecosystem depreciation burden by hundreds of billions to more than a trillion dollars. That is the size of the analytical gap this report exists to bound.

SECTION · II

The Depreciation Debate Framework

$228B

FOOTNOTEBRIEF CUMULATIVE SUPPRESSED DEPRECIATION, 2026–2028

The cleanest natural experiment the debate has is what happened at the start of 2025 — when two hyperscalers moved in opposite directions on the same question within the same quarter. Effective January 1, 2025, Amazon shortened the useful life of a subset of its servers and networking equipment from six years back to five. CFO Brian Olsavsky told investors on the February 6, 2025 call, "we're decreasing the useful life for a subset of our servers and networking equipment from six years to five years, beginning in January 2025," citing "the increased pace of technology development, particularly in the area of artificial intelligence and machine learning." The company guided to a **~$700M reduction in 2025 operating income**, plus a separate **~$920M accelerated depreciation charge** in Q4 2024 for early-retired equipment. In the same quarter, Meta moved the opposite direction, extending the useful life of certain servers and network assets to 5.5 years — a change it expected to reduce full-year 2025 depreciation expense by **approximately $2.9 billion**.

Two Hyperscalers, Opposite Directions

Two hyperscalers, the same auditors, the same hardware cycle, opposite conclusions within the same month. Amazon's reversal cost real reported income with no offsetting tax or narrative benefit — the kind of hit a company does not take voluntarily unless its own utilization data leaves it no choice.

The cumulative picture of walk-backs and extensions matters. Amazon's Q3 2025 shortening drove a $392M depreciation increase and a $298M net income reduction in the quarter alone, with a nine-month cumulative $889M depreciation and $677M net income effect. Meta has extended useful lives three times in three years, culminating in the January 2025 $2.9B depreciation reduction. The coordination ran one direction from 2020–2023: AWS extended first (3→4 in 2020, 4→5 in 2022), and Google and Microsoft followed onto a uniform six-year assumption by 2023–2024. footnotebrief's reconstruction, benchmarking a four-year counterfactual against the affected base and projected 2026–2028 capex, produces roughly $228B in cumulative suppressed depreciation (~$46B in 2026, ~$75B in 2027, ~$107B in 2028), which it rounds conservatively to ~$200B — in the same neighborhood as Burry's $176B.

There is also a framing question that is not academic: which number should investors trust when the two diverge, free cash flow or reported earnings? Alphabet's Q2 2026 negative FCF print exposes exactly why the answer matters — a company can post record operating income and simultaneously bleed cash, and the depreciation debate determines which number is "real."

One further wrinkle sharpens the picture. Accelerated depreciation is paradoxical: it depresses reported net income and book ROIC, but it is cash-flow protective in the near term. Shorter useful lives push more of the depreciation deduction into earlier years, reducing taxable income and cash tax payments — a real tax shield worth roughly the depreciation increment times the marginal tax rate. Amazon's February 2025 walk-back to five-year lives illustrates the mechanics: the ~$920M accelerated charge hurt reported income but delivered offsetting near-term cash tax relief. That makes Alphabet's Q2 2026 negative FCF print harder to dismiss. Even without a useful-life change — and even with the tax shield from higher depreciation running in Alphabet's favor — free cash flow crossed into negative territory. When the accounting cushion is intact and cash flow still turns negative, capex volume itself is the culprit.

The Dual-Life EVA Response

The methodology needed here is straightforward. Rather than a single-point ROIC, run two parallel cases. The **book case** applies the disclosed 5–6 year useful lives and produces Convequity's ~28.6% ROIC. The **economic case** applies a 3-year GPU life and a 4-year networking life to an AI-share of the asset base — approximately 70% for Microsoft, Google, and Meta and roughly 50% for Amazon (mirroring Cembalest's AI-share assumptions of 70%/50%) — with a 10% salvage assumption. Under the economic case, the ROIC–WACC spread compresses from ~10 percentage points to roughly 200–500 basis points on the marginal dollar. The right output is a matrix — ROIC across life columns (3/4/5/6 years) against revenue rows (bull/base/bear) — not a single headline number.

SECTION · III

Engineering Reality vs. Accounting Fiction

27%

THREE-YEAR CUMULATIVE H100 FAILURE RATE, META LLAMA 3 STUDY

The engineering evidence carries the economic-life case, and it is more concrete than the accounting debate usually acknowledges. Meta's own published study of the Llama 3 405B training run — **16,384 H100 GPUs over 54 days** — recorded 419 unexpected interruptions, of which 148 (30.1%) were GPU failures including NVLink and 72 (17.2%) were HBM3 memory failures. That is roughly one failure every three hours. Tom's Hardware's derived math implies an approximately **9% annualized failure rate**, which compounds toward roughly **27% cumulative failure over three years**. This is not a fringe estimate — it is arithmetic off a first-party disclosure from one of the four hyperscalers whose depreciation policy is under scrutiny.

MTBF vs. Economic Life

The distinction between how long a chip physically survives and how long it earns its keep is where naive analysis goes wrong. Epoch AI estimates a single H100 fails roughly once every 50,000 hours — about six years of continuous operation. But mean time between failures is not a survival guarantee. Under standard exponential reliability, a component with a 100,000-hour MTBF has only about a 36.7% probability of actually surviving to that mark. And economic life is shorter still than physical life: at the 60–70% utilization rates typical of cloud providers, thermal cycling, sustained 700W+ power dissipation, solder-joint microscopic damage, silicon-substrate stress, and power-delivery-component degradation all accelerate wear. An unnamed source described as a "GenAI principal architect at Alphabet," relayed via Tom's Hardware in October 2024, put expected service life at "one to two years, three years at the most" under constant high utilization. That specific claim is anonymous, third-hand, and explicitly not endorsed by Tom's Hardware — treat it as directional, not dispositive. But it points the same direction as the hard failure data and the rental-rate collapse.

The secondary-market data is the market's own verdict on residual value, and it is genuinely two-sided. Used H100s trade at roughly 60–70% of new (about $15,000–$28,000 versus $25,000–$40,000 new), having peaked near $50,000 during the mid-2024 scarcity before falling as supply normalized. Used A100s retain roughly 60–80% of original value. These retention rates superficially support the bull case — until you look at the cash-yield metric that actually matters. **H100 hourly rental rates fell from a 2024 peak of ~$8/hour to $1.70–$3.50/hour through 2025**, a decline of 64–75%, which is the single cleanest real-world proxy for economic-life decay. That headline decline is not the whole picture. Per SemiAnalysis's H100 1-Year Rental Price Index in April 2026, H100 1-year GPU rental contract pricing rebounded nearly 40% from a low of $1.70/hour in October 2025 to **$2.35/hour** by March 2026, with on-demand capacity sold out across GPU types. The rental market is telling a value-cascade story and an obsolescence story at the same time — which is precisely why any single-point depreciation number is inadequate.

The Value Cascade's Vulnerability

The value cascade is the bulls' best defense: GPUs waterfall from frontier training (years 1–2) to real-time inference (years 3–4) to batch inference and analytics (years 5–6), earning a declining but positive margin the whole way. The evidence is real — Azure retired K80/P100/P40 instances only after seven-to-nine-year service lives, and CoreWeave reportedly rebooked 2022-vintage H100s at 95% of original contract pricing on expiration. But the cascade has a specific vulnerability the bulls underweight: **custom inference silicon**. AWS Inferentia and Trainium, Microsoft Maia, Meta MTIA, and Google's own TPUs are all on annual cadences and are purpose-optimized for exactly the inference workloads the cascade assumes will keep old training GPUs employed. If the batch-inference tier is increasingly served by cheaper, more power-efficient custom ASICs, the back half of the cascade — the part that justifies years 5–6 of the depreciation schedule — erodes. Satya Nadella's stated procurement-staggering strategy at Microsoft ("each year you ride the Moore's Law, you add, you use software, and then you optimize") is a rational response, but it is also an implicit admission that any single vintage's

economic life is compressing.

SECTION · IV

The Nvidia Cadence as Forcing Function

~600 kW

RUBIN ULTRA KYBER RACK DRAW VS 5–10 KW/RACK LEGACY SHELLS

Nvidia has explicitly converted its product cadence into a financial weapon. Jensen Huang confirmed an annual architecture cadence at GTC 2025 and reaffirmed it through CES and GTC 2026: **Blackwell Ultra B300 (H2 2025) → Vera Rubin R100 (H2 2026) → Rubin Ultra R300 (H2 2027) → Rosa Feynman (2029–2030)**. Each generation compresses the economic half-life of the prior one. Vera Rubin is not an incremental refresh: it moves to a TSMC 3nm node, a new HBM4 memory generation, and roughly a 2.8x jump in memory bandwidth — the true bottleneck for LLM inference — with the VR200 platform specified around 50 PFLOPS FP4 and 22 TB/s of memory bandwidth, roughly a 3.3x compute leap over B300. Memory capacity has climbed from 80GB on H100 to 180GB on B200 to 288GB on B300 and Rubin-class parts; the precision floor moved from FP8 on Hopper to NVFP4 on Blackwell at roughly 2x FP8 throughput; NVLink bandwidth went from 900 GB/s on Hopper to 1.8 TB/s on Blackwell to NVLink 6 on Rubin.

Crucially, the cadence is co-designed at the rack scale, which is what breaks the value-cascade defense mechanically. Nvidia claims the rack-scale Rubin delivers up to 10x higher inference throughput per watt than Blackwell at one-tenth the cost per token. When performance-per-watt improves that fast, an older GPU does not merely become slower — it becomes uneconomic to power and cool relative to the revenue it can generate, regardless of whether it still physically functions. Two process-node progressions roughly halve the energy per FP16/BF16 matrix operation, which is why the industry converges on an **18–36 month TCO obsolescence window** for frontier training. Nvidia has reportedly locked up roughly 60% of TSMC's advanced packaging capacity for 2026, and TSMC is committing on the order of $56B of capex to roughly double CoWoS capacity — a supply-side signal that the cadence will hold.

The rack-scale power trajectory is the bridge to the stranded-asset argument. The VR200 NVL72 rack draws roughly **190–230 kW** versus 120–130 kW for Blackwell and ~40 kW for Hopper, with the 2027 Rubin Ultra "Kyber" rack specified at approximately **600 kW**. First-generation AI data-center shells were designed for 5–10 kW/rack. The physical obsolescence of the shell is baked into the roadmap before the first Rubin ships.

SECTION · V

Recent Signals from Q2 2026 Earnings

–$5.9B

ALPHABET Q2 2026 FREE CASH FLOW — FIRST NEGATIVE SINCE 2004 IPO

The July 22–30 window put all four US hyperscalers on the tape within nine days, and in aggregate the reports validated rather than complicated the depreciation thesis. Alphabet posted its first-ever negative quarterly free cash flow and raised capex guidance again. Microsoft's CFO acknowledged on a live earnings call that roughly two-thirds of the company's capex is going to short-lived assets. Meta beat revenue but missed EPS by 14% and raised the low end of its capex guidance a second time. Amazon hiked full-year capex from $200B to $220B mid-cycle. Every hyperscaler that reported after Alphabet either raised capex further or absorbed an earnings miss. None walked back the AI infrastructure commitment.

Alphabet — The Leading Canary

Alphabet's Q2 2026 report, released July 22, 2026, is the most important single data point in this section because it was the first to show the book-versus-economic gap surfacing in the financials of the most disciplined hyperscaler. The headline numbers were operationally excellent: revenue up 24% year-over-year to **$119.8B**, operating income up 30% to $40.8B at a 34% margin, Google Cloud revenue up 82% to $24.8B with operating margin expanding from 20.7% to 35.6%, and cloud backlog swelling more than $50B sequentially to **$514B**. But the cash flow statement told the other story. Capital expenditures roughly doubled year-over-year to $44.9B in the quarter, outrunning $39.1B of operating cash flow and producing negative free cash flow of **–$5.9B** — Alphabet's first negative quarterly free cash flow since its 2004 IPO. Full-year 2026 capex guidance was raised to **$195–205B** from a prior $180–190B, with management explicitly warning that 2027 will "increase significantly." CFO Anat Ashkenazi described the "supply-constrained environment" and said free cash flow will "remain under pressure." The stock fell 4–7% on the release.

The depreciation signals in the release were unusually explicit. Other Cost of Revenues rose 22%, "driven by increases in depreciation, inventory costs, primarily from the sales of TPU systems to customers, and content acquisition costs." More tellingly, R&D expenses rose 32%, "driven by compensation from investment in AI talent, as well as depreciation" — depreciation appearing as a named driver in the R&D line is unusual and signals how pervasively the capital charge is now bleeding through the P&L. Alphabet also disclosed, for the first time, its capex composition: **approximately 60% servers and 40% data centers and networking**. Notably, Alphabet made no useful-life change — the depreciation pressure is arriving from capex volume alone, before any policy tightening.

Microsoft — The CFO's Concession

Microsoft's Q4 FY26 print (July 29, covering April–June 2026) was operationally clean. Azure crossed **$100B in annualized revenue** for the first time in the company's history; Microsoft 365 Copilot surpassed 30 million paid seats; adjusted EPS came in at $4.74 against a $4.24 consensus; the stock added roughly $260B in market value in a single after-hours session. But the analytically important disclosure was in CFO Amy Hood's prepared remarks on capital expenditure. Q4 capex reached a record **$41B** — and, per Hood: *"roughly two-thirds of our capex was for short-lived assets, primarily CPUs and GPUs as customers increasingly build solutions that leverage both AI and non-AI infrastructure."*

That single sentence — from the CFO of arguably the most disciplined hyperscaler, on a live earnings call, unprompted — is the tape's clearest concession to the depreciation debate. It concedes what Burry, footnotebrief, and Cembalest have been arguing: that the marginal dollar of hyperscaler capex is flowing into assets whose economic life is closer to two to four years than to the five- and six-year lives on the books. Two-thirds of $41B annualizes to roughly **$110B per year of short-life assets** from Microsoft alone. Full-year FY26 capex reached **$115.95B**, up nearly 80% year-over-year. Free cash flow after capex fell to $19.64B for the quarter, down from $25.57B YoY — a 23% decline even as operating cash flow grew 30%. The gap between the two is depreciation, deferred. Management guided FY27 capex to **$255–260B** (a 35% step up from FY26), framed as pre-contracted against the backlog now exceeding twice annual revenue. One caveat: part of the FY26 outlook comparison is affected by an accounting-classification change taking effect in FY27, which investors should note when tracking the trajectory.

Meta — Depreciation Reaches the P&L

Meta's Q2 2026 print (July 29) was where the depreciation thesis moved from tape signal to income-statement reality. Revenue beat consensus at **$60.8B** against $60.3B expected, with growth holding at 28% year-over-year. But EPS **missed by 14%**: $6.18 against a $7.22 consensus, with EPS down 13% year-over-year. The revenue side is intact; the earnings side is showing the depreciation and infrastructure-cost pressure the book-versus-economic gap would predict. Meta raised the low end of full-year 2026 capex guidance to **$130–145B** from $125–145B — the second increase this year — and simultaneously raised the low end of full-year expense guidance to **$165–169B** from $162–169B.

The implied second-half capex trajectory shows just how quickly this can spread. Meta guided full-year capex at $130–145B, but spent only $19.84B in Q1 (well below the $27.6B Street estimate). To hit the midpoint of $137.5B, Meta must now spend **$39–47B per quarter** across Q2, Q3, and Q4 — an upper bound that exceeds Alphabet's Q2 record of $44.9B. The IOU footnotebrief flagged has grown, not shrunk. And crucially, unlike Amazon in February 2025, Meta made no useful-life change; the earnings pressure is arriving from capex volume alone.

Amazon — The Mid-Cycle Raise

Amazon's Q2 2026 report (July 30) confirmed peer contagion is engaged. AWS revenue came in at **$42.2B** against $40.5B consensus, with Andy Jassy describing the segment as "booming"; Amazon's AI services and Trainium/Inferentia units each cleared a **$25B annual revenue run rate**. In the same release, Amazon raised its full-year 2026 capex guidance from $200B to **$220B** — a $20B mid-year increase, echoing precisely the move Alphabet made the week prior. Amazon's capex trajectory has now expanded roughly 4x from $52.7B in FY23 to over $220B guided for FY26. Amazon is also the only hyperscaler in the group that has already taken a useful-life shortening (February 2025). The combination of that precedent with a mid-cycle capex hike shows the thesis playing out end-to-end on a single company's tape.

The Aggregate Picture

Combining the four disclosures, the calendar-2026 capex totals now sit near **$740–770B** across the group — Alphabet at $195–205B, Microsoft near $190B (with FY27 stepping to $255–260B), Meta at $130–145B, Amazon at $220B — roughly double the ~$394B combined capex the group ran in FY25. Of that total, roughly **60–67% runs to short-lived compute** (GPUs, CPUs, networking, memory) and **33–40% to long-lived civil and power infrastructure** (shells, substations, cooling) — the depreciation debate cuts against the former, not the latter. Silicon Analysts' framework for the gap remains the cleanest single expression of the debate: roughly **$549B of deferred cost** across the group, combined free cash flow projected to fall 91% to ~$16B against combined net income of ~$506B. That divergence between reported net income and cash flow is the depreciation debate expressed as a single ratio. None of the four hyperscalers changed a useful-life assumption in these prints. The gap is widening from capex volume alone, before any accounting adjustment.

SECTION · VI

Data Center Shell and the Stranded-Asset Cliff

35,000

SQUARE FEET STRANDED IN A MODELED 50 MW RETROFIT TO 132 KW RACKS

The third pillar under the reported spread assumes the data center itself — shell, power, cooling — outlives the compute inside it. That is where hyperscaler accounting breaks down. GAAP (ASC 360) permits component depreciation, and a data center is not one asset but a stack of assets with wildly different lives: GPUs at 2–4 years, racks and networking at 6–10 years, cooling at 15–25 years, and the shell itself at 25–40 years. Hyperscalers report blended rates that mathematically smooth this mismatch — a GPU depreciating at 30–35% salvage over 3.5–4.5 years sits inside a reported "server and networking" pool that also contains 6–8 year chassis, power, and cooling components. The blend hides the stranding.

The rack-scale power trajectory documented in Section IV makes the stranding physical rather than merely accounting-driven. First-generation AI data-center shells were designed and permitted for 5–10 kW/rack; Blackwell-class racks draw 120–130 kW; Rubin-class racks draw 190–230 kW; the 2027 Rubin Ultra "Kyber" rack is specified at approximately 600 kW. A modeled retrofit of a legacy 50 MW shell designed for 132 kW-class racks to accommodate the newer power density strands roughly **35,000 square feet** of usable floor space — capacity that was fully productive under the prior generation and is now either unused or requires capital investment in power and cooling infrastructure to reclaim. That stranded square footage sits on the books at its original cost, depreciating over a 25–40 year shell life that assumed continuous productive use.

The renewal cliff this points toward lands squarely in 2026–2027. The 2022–2023 H100 cohort — the vintage that powered the initial ChatGPT-driven capex wave — crosses the 3–4 year economic-life threshold precisely as the Rubin generation ships and rack power requirements roughly double again. Hyperscalers face a simultaneous compute refresh and a power/cooling refresh, on assets whose book lives assumed neither would be necessary this soon.

SECTION · VII

Power Infrastructure Timeline vs. Compute Curve

9–15 yr

SMR DESIGN + LICENSE + CONSTRUCTION STACK — NO US COMMERCIAL LICENSE ISSUED

The fourth assumption — that power will be available to run the compute being deployed — is contradicted by the physical timeline of every clean-baseload solution the hyperscalers have announced. Start with small modular reactors, the technology the industry has bet most heavily on. The regulatory-plus-construction stack runs roughly **9–15 years**: design certification (3–5 years), site licensing (3–5 years), and construction (3–5 years). As of early 2026, no US SMR construction license had been issued for a commercial data-center project on the accelerated timelines the AI narrative implies, and the marquee projects all land in the early-to-mid 2030s. TerraPower's Natrium plant in Kemmerer, Wyoming broke ground in June 2024 targeting commercial operation in the early 2030s. NuScale, the first to receive US design certification, saw its UAMPS Carbon Free Power Project in Utah cancelled in November 2023 on cost overruns; its ENTRA1 partnership targets a first plant around 2030 (72 modules, 6 GW plan). The most optimistic North American case, Ontario's BWRX-300 at Darlington, targets 2029. The hyperscaler-specific deals are 2030-and-beyond: Oklo–Meta's ~1.2 GW Ohio campus (first phase ~2030), and Google–Kairos Power's first reactor around 2030 with a full 500 MW by 2035.

The reactor timelines are the optimistic version. Even if the plants were ready sooner, the fuel wouldn't be. Nine of ten advanced reactor designs require high-assay low-enriched uranium (HALEU), yet current US production is minimal — measured in hundreds of kilograms against DOE needs of **40+ metric tons by 2030** and **50+ metric tons by 2035**, with EIA-cited projections of ~50 metric tons per year by 2035, and the only established commercial producer at scale is Russia (the primary pre-2022 source). The DOE reissued a $900M funding tender in March 2025 to accelerate domestic SMR and fuel development, but until domestic enrichment scales, fuel availability caps SMR deployment regardless of reactor progress.

Meanwhile the compute is being deployed now. The mismatch is being papered over in two ways, each with hidden costs. First, "virtual" PPAs (financial contracts for renewable attributes) are being conflated in corporate messaging with physical, wired baseload power — but a virtual PPA does not put an electron on the data center's bus. Second, **transmission interconnection queues in congested markets now run 5+ years**, with FERC's queues holding more than 2,600 GW of proposed generation. The practical result: hyperscalers are bridging with on-site gas and fuel cells (Oracle contracted up to 2.8 GW of Bloom Energy fuel cells for Stargate, 1.2 GW through 2027) — expedient, but neither the clean-baseload nor the low-cost outcome the models assume. Stargate itself ($500B announced January 2025, ~10 GW target) is a national-infrastructure undertaking whose power needs would strain regional grids, financed substantially by Oracle debt (~$100B of planned borrowing) — which is why S&P downgraded Oracle from BBB to BBB–

in July 2026. The credit market had already been signaling the same concern for weeks: Oracle's five-year credit default swap spread, the market's own price on the risk of default, climbed to a multi-year high in July as bond investors began treating Oracle paper as a liquid proxy for AI-financing risk more broadly, not just company-specific concerns. That is a market-based echo of the same downgrade thesis, arrived at independently by a different set of investors.

SECTION · VIII

The Peer Contagion Scenario

1.5–2.5x

MODELED ALPHABET P/E TURN COMPRESSION ON PEER-CONTAGION SHORTENING

The useful-life coordination mechanism, having run in one direction from 2020–2023, can run in reverse under stress. Consider what happens if Alphabet — not Amazon, which has already moved — unilaterally shortens its GPU useful life. The effects would arrive through four channels: signaling, mechanical earnings, multiple compression, and credit.

The largest effect would be signaling, not mechanical. If Alphabet — regarded as one of the most disciplined and technically credible hyperscalers — concluded that its own utilization and obsolescence data no longer supported a six-year life, that alone would validate the bear thesis for the entire group. Peer contagion to Microsoft, Meta, and Oracle would follow the same fast-follow dynamic that drove the extensions, running in reverse. The direct earnings drag on Alphabet itself would be more modest: a mid-single-digit-billions annual operating income hit on a 6→5 move, scaling toward **$15–20B** on a 6→4 move as the affected fleet compounds. But the real damage would sit in the multiple: a **1.5–2.5 turn P/E compression** on Alphabet, roughly **$300–500B in enterprise value**, with proportionate hits across the peer set. And there is a second-order consequence beyond earnings and multiples: higher depreciation reduces reported income and can pressure debt covenants and buyback capacity precisely when the group is issuing record debt and equity to fund the build.

The cumulative suppressed depreciation of $80–200B across the four extending hyperscalers over 2026–2028 (footnotebrief's $228B, Burry's $176B) is the size of the earnings cushion that a coordinated reversal would remove. The point is not that reversal is imminent — Alphabet explicitly did not change its life in Q2 2026 — but that the option is live, the precedent exists (Amazon), and the market is not pricing the correlation. When one hyperscaler's auditor forces a change, the "identical technology, identical conclusion" logic that footnotebrief identifies makes it very hard for the others' auditors to hold the line.

SECTION · IX

EVA Sensitivity Matrix and Portfolio Implications

~7.7%

BEAR-CASE ROIC — NINE-TO-TEN-YEAR PATH BACK ABOVE 12% WACC

Once the depreciation-life and revenue axes are treated together, the output is a range, not a point. On the revenue side, InvestorPlace's decomposition is a useful spine: a mid-case of ~$7T annual AI revenue at 32% blended operating margins on ~$6.4T net invested capital yields **~28% ROIC** (generating ~$1.77T of annual profit, in the range of the greatest capital-allocation stories in US history); a bull case of ~$10T revenue at 40% margins implies **~49% ROIC** (Apple-tier); and a bear case of **~7.7% ROIC** that requires revenue at less than half the mid-case, crossing below a 12% cost of capital, with a nine-to-ten year path back to break-even. Layer onto that revenue axis the depreciation-life axis (3/4/5/6 years), plus residual-value and cascade-capture sensitivity (5–15% residual at year five moves ROIC by roughly 200–400 bps), and the matrix reveals a marginal-dollar spread that is positive-but-thin in the realistic middle of the distribution and negative in the bear corner.

Applied at the ticker level under a 3-year GPU life, the framework shows where each name's marginal-dollar spread lands. Illustrative of the dual-life decomposition, not published spot ROICs:

	SERVER LIFE	BOOK ROIC	3-YR ROIC	WACC	SPREAD
MSFT	6 yr	22.0%	12.5%	8.5%	**+400 bps**
GOOGL	6 yr	24.5%	13.0%	8.0%	**+500 bps**
META	5.5 yr	26.0%	11.0%	9.0%	**+200 bps**
AMZN	5 yr	18.0%	9.5%	8.5%	**+100 bps**
ORCL	5–6 yr	15.0%	5.5%	9.5%	**−400 bps**

Illustrative of the dual-life framework, not published spot values. Estimated at company level for AI-attributable invested capital; company-level ROICs differ from the aggregate 28.6% Convequity figure, which is computed on marginal AI capex only. See Section II for methodology.

What that means for positioning depends on where in the AI stack you are looking. First, the **infrastructure suppliers** (Nvidia and the packaging/memory/power complex) are paid regardless of whether the hyperscaler ROIC calculation resolves in the bull or base case — they collect the capex on the way in. Nvidia's own equity stakes in the neocloud layer — a disclosed position in CoreWeave, and a 9.3% beneficial-ownership stake in Nebius disclosed in July 2026 — mean the largest supplier is now also an equity holder in the names carrying the most leveraged

downside risk, which is worth watching as a circularity signal rather than a pure hedge. Second, the **hyperscalers themselves are differentiated** by their AI-share and by the degree to which their core businesses subsidize the build: Amazon (lower AI-share, ~50%, and already conservative on life) and Alphabet (highest-quality core, though now cash-flow-negative on the margin) screen differently from a pure-play. Third, the **neoclouds** are the leveraged play on exactly the assumptions under stress — CoreWeave, Nebius, and IREN carry the WACC, depreciation, and demand risk simultaneously and with none of the subsidizing core business, which is why the group has already sold off 40–60% from peaks (Nebius −37%, CoreWeave −61%, IREN −47% at the July 2026 trough). The correct posture, in Cembalest's framing, is *"not awe, but vigilance."*

SECTION · X

Information Asymmetry and the IU Digest Opportunity

$460B

DIRECT DEBT ACROSS SIX OPERATORS (MOODY'S); $1.2T LEASE COMMITMENTS

There is one more observation worth making, and it is about the research market itself. Retail investors are largely blind to the depreciation mechanism because it lives in property-and-equipment footnotes and useful-life disclosures that never appear in headline EPS. The sell-side has reasons of its own not to break the constructive frame on the largest index constituents and its biggest banking clients — even as Moody's warns that direct debt across Microsoft, Amazon, Alphabet, Meta, Oracle, and CoreWeave has reached roughly **$460B** (with off-balance-sheet lease commitments near **$1.2T**) and that capex now approaches or exceeds 100% of operating cash flow for some operators versus a 10-year average near 40%.

The analytical load on this question is being carried disproportionately by the boutique research space — Convequity, Deep Quarry, footnotebrief, The Forward View, Stanley Laman, Silicon Analysts, and Cembalest's *Eye on the Market* — precisely the reader set IU Digest serves. The opportunity is to be the venue that synthesizes the accounting, engineering, power, and real-estate threads into a single coherent EVA framework before the tape forces the question.

POSITIONING FRAMEWORK · §

Recommendations

The report's positioning framework is staged. Stage 1 is the near-term posture. Stage 2 tracks the four trigger variables that would move the recommendation. Stage 3 defines the benchmarks that would materially change the thesis in either direction.

STAGE ONE

Stage 1 — Now (before Q3 2026 earnings)

Treat the marginal capex dollar as a range, not a point — 28.6% book ROIC, 200–500 bps under engineering-realistic useful lives. Size positions to somewhere in the middle of that distribution, not the bull corner. Trim pure-play neocloud exposure (CoreWeave, Nebius, IREN) where the cost of capital, depreciation, and demand risks are stacked without a subsidizing core business — CoreWeave's cost of capital reversing from sub-6% back above 10% is the leading indicator that the fair-weather assumption on financing has broken.

STAGE TWO

Stage 2 — Trigger Variables to Track

Four variables would move the recommendation materially: (1) a second hyperscaler shortening useful life following Amazon's precedent; (2) GPU rental rates breaking back below the $1.70/hour October 2025 low rather than holding the SemiAnalysis-documented rebound; (3) CMBS collateral marks on data-center-backed debt deteriorating meaningfully below par; (4) two consecutive quarters of negative free cash flow at any of the four hyperscalers, which would confirm the crossover documented in Alphabet's Q2 2026 print is structural rather than lumpy.

STAGE THREE

Stage 3 — Benchmarks That Would Change the Thesis

A bullish resolution would show demand growth outrunning the depreciation wave — Pichai's "very early innings" framing proving correct, with revenue scaling toward the InvestorPlace bull case (~$10T, ~49% ROIC) rather than the mid-case. A bearish resolution would show a second and third hyperscaler following Amazon's useful-life shortening within the next two reporting cycles, confirming the peer-contagion scenario in Section VIII. Either outcome would resolve the current 200–500 bps ambiguity in one direction with much greater conviction than the data currently supports.

EDITORIAL DISCLOSURE · §

Caveats and Sourcing Notes

The observations below bound the analysis and apply to every claim in this report. Readers should weigh them against the strength of the individual arguments.

01 The strongest datapoints in the report are asymmetrically sourced. Alphabet's Q2 2026 figures, Amazon's and Meta's useful-life disclosures, Meta's Llama 3 failure study, Microsoft's Amy Hood commentary, and the Nvidia roadmap are all well-documented from primary or first-party sources. But several load-bearing claims in the bear case are weaker: the "1–3 year service life" attributed to an Alphabet architect is anonymous, third-hand, and explicitly not endorsed by Tom's Hardware, and should be read as directional rather than dispositive.

02 The value cascade is a genuine phenomenon with real supporting evidence (Azure's 7–9 year instance retirements, CoreWeave's 95% rebooking), and the recent SemiAnalysis rental-rate rebound (from $1.70 to $2.35/GPU-hour) cuts against the pure-obsolescence story — the economic-case ROIC compression is a scenario, not a forecast.

03 Many of the most dramatic figures in circulation — the $500B+ deferred-cost estimates, the CMBS "80 to 20 cents" collateral collapse — come from analytical models and secondary commentary, not realized events, and several carry visible authorial spin; they are presented here as sized risks, not established outcomes.

04 The hyperscalers' core businesses are extraordinarily profitable and self-funding, which means even the 7.7% bear-case ROIC is "survivable" for them in a way it is not for levered pure-plays — this is a margin-compression and multiple-compression thesis, not a solvency thesis, for the Big Four.

05 Several forward-looking figures (Bessemer's $6T revenue path, InvestorPlace's bull/bear ROIC cases, Oracle's $144B 2030 OCI target, sell-side FCF projections, Goldman's $7.6T cumulative-capex baseline) are explicitly projections using conditional language and should not be read as facts.

06 The entire debate could be defused relatively quickly if hyperscalers improve depreciation disclosure or if demand growth simply outruns the depreciation wave — Pichai's "very early innings" is not merely spin, and the bull case has repeatedly embarrassed the bears. Oracle's own fiscal Q4 print in early August, which sent the stock up nearly 9% on strong cloud growth even as its credit spreads stayed elevated, is a live example of that dynamic playing out in real time.

COLOPHON

IU DIGEST · INTERMARKET UNIVERSE