

When Will AI Be Both Powerful and Profitable?

April 2026

The scale of today's AI infrastructure buildout is extraordinary. Bloomberg estimates 2026 aggregate capex by the major U.S. hyperscalers¹ will total \$650 billion, roughly 2% of U.S. GDP. The Wall Street Journal tells us that "Big tech is becoming the new steel and railroads".² But gross spending is not the same as capital accumulation. The difference is depreciation.

In industries where assets remain productive for decades, gross investment serves as a reasonable proxy for capital deepening and shareholder value creation. Steel mills and railroad tracks depreciated over 40 to 45 years. AI infrastructure depreciates in about 5 years.³

If the economic life of AI hardware is shorter than its accounting life, reinvestment needs are higher than reported depreciation suggests. What appears to be capital deepening by hyperscalers is largely capital churn.

Depreciation of AI Hardware

Hyperscalers report server depreciation lives of five to six years.⁴ Economically, AI hardware ages much faster. Nvidia now operates on an annual product cadence.⁵ Each generation of new chips delivers materially more compute per watt than the last. Older chips rarely fail physically. They become obsolete economically. The relevant test is not whether a GPU still runs, but whether it earns a return sufficient to cover depreciation, operating costs, and the cost of capital.

Rental markets provide a real-time signal of economic value. Observed hourly rates for Nvidia H100 GPUs fell from scarcity-driven peaks near \$8 in 2024 to below \$3 by late 2025, and to below \$1 in early 2026.⁶ As new generations arrive, pricing converges toward the marginal cost of the most efficient hardware.

Table 1 translates those rental rates into per-GPU economics.⁷ During the first three years, returns on capital are substantial. In year four, revenue falls below total economic cost. The chip remains in service because it still covers marginal operating expenses. But it no longer earns a return on invested capital. Economically, the productive life of the asset is closer to three years than to the five-year accounting for depreciation.



AUTHORS



Chris Brightman, CFA*
Partner, Senior Advisor

*Corresponding author

Key Points

- Today's historic level of AI capex generates headlines but may conflate gross spending with lasting capital accumulation.
- Under plausible depreciation assumptions, more than half of projected 2026 hyperscaler capex replaces economically obsolete hardware rather than expanding productive capacity.
- History shows that extraordinary capex can sustain a competitive position without creating value for investors.
- Investors should distinguish between gross capex and net capital formation when assessing whether AI spending builds sustainable cash flow.

Table 1. AI Infrastructure Becomes Economically Obsolete After Three Years

Cohort Age	Per-GPU Unit Economics: NVIDIA H100 GPU					
	Rate (\$/hr)	Revenue	Cost	Profit	Book Val.	ROCE
2022 (Training)	7.62	40,051	13,100	26,951	32,500	83%
2023 (Training)	9.17	48,198	12,580	35,618	26,000	137%
2024 (Training)	6.26	32,903	12,060	20,843	19,500	107%
2025E (Inference)	1.35	7,096	11,540	-4,444	13,000	-34%
2026E (Deep Tail)	0.75	3,942	11,020	-7,078	6,500	-109%

Note: Annual revenue equals hourly rental rate × 8,760 hours × 60% utilization. Total cost equals straight-line depreciation (five-year life), operating expense, and an 8% capital charge applied to beginning-of-year book value. Economic profit equals revenue minus total cost; ROCE equals economic profit divided by beginning-of-year book value.

Source: Research Affiliates and H100 SXM cost \$32,500 (Clarifai/IntuitionLabs, Aug. 2025). Rental rates: Bloomberg SDH100RT (Years 1–3); Hyperstack and ThunderCompute (Years 4–5). Operating cost \$4,000 per year; 8% cost of capital; 60% utilization (Microsoft estimate).



Rental markets are imperfect signals. Spot pricing can overshoot in both directions, and hyperscalers may achieve higher utilization or lower financing costs than third-party renters. The point is not that a GPU fails after three years, but that its ability to earn excess returns may fade well before its accounting life ends. Economic obsolescence precedes physical obsolescence by several years.

“The point is not that a GPU fails after three years but that its ability to earn excess returns may fade well before its accounting life ends. Economic obsolescence precedes physical obsolescence by several years.”

Compute Is Not Capital

Rising compute is often treated as evidence of capital accumulation. The two are not the same. Compute is a flow. It measures processing delivered per unit of time. Capital is a stock. It represents the accumulated value of deployed hardware.

AI capex increases compute faster than it builds capital stock. Each new chip generation delivers far more compute per watt than its predecessor. A data center constrained by power can materially increase compute simply by replacing last year’s hardware.

Table 2. Net Investment Declines with Economic Service Life

Scenario	Stock-Flow Accounting: Gross Investment, Depreciation, Net Formation, Ending Capital Stock (\$Billions)									
	2025 Estimate					2026 Estimate				
	Gross	Depr.	Net	Capital	Net/Gross	Gross	Depr.	Net	Capital	Net/Gross
2-Year Life	400	328	73	655	18%	650	525	125	1,050	19%
5-Year Life	400	219	181	1,096	45%	650	326	324	1,630	50%
7-Year Life	400	184	216	1,286	54%	650	264	386	1,846	59%

Source: Research Affiliates and Table 1 above. Net/Gross ratio = net capex / gross capex. Scenarios (2-Year, 5-Year, 7-Year) correspond to AI-specific, accounting disclosures, and traditional-IT depreciation assumptions respectively.



Table 2 puts the distinction in dollar terms. Under a two-year economic life, net capital accumulation in 2026 would be just \$125 billion, roughly one-fifth of the \$650 billion headline figure.

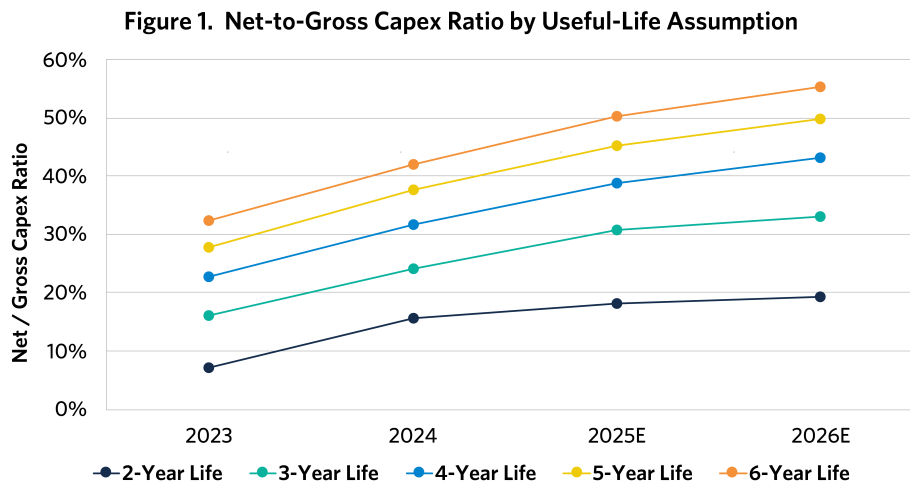
Reinvestment, Cash Flow, and Valuation

If the economic life of frontier AI hardware is closer to three years than five, steady-state reinvestment needs are structurally higher than financial statements report. With a three-year economic life, roughly one-third of the installed base must be replaced annually just to maintain capacity. Under those conditions, a large share of capex is not discretionary. It is maintenance.

“If the economic life of frontier AI hardware is closer to three years than five, steady-state reinvestment needs will be structurally higher than financial statements report.”

For valuation, this distinction matters. Free cash flow depends not on reported depreciation, but on the capital required to sustain competitive position. When economic depreciation exceeds accounting depreciation, much of the reported capex is more like an operating expense.

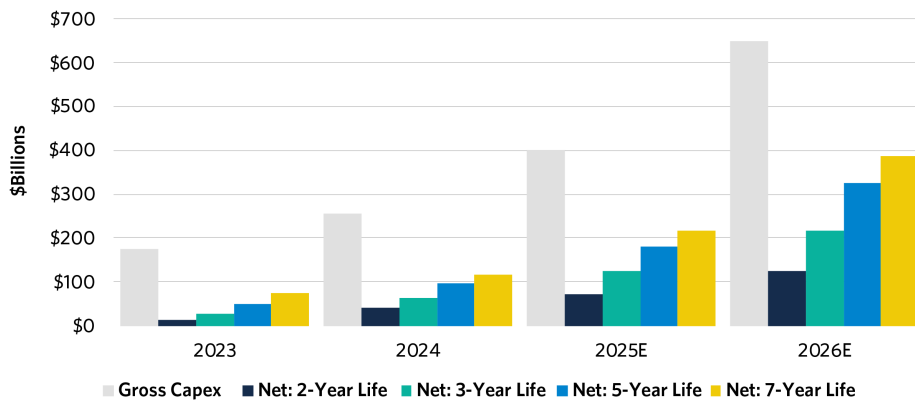
Higher reinvestment needs reduce free cash flow. Terminal value assumptions become more sensitive to reinvestment required to remain competitive. Even modest changes in economic life can meaningfully alter sustainable cash flow.



Source: Tables 1–2. Scenarios assume straight-line depreciation over the stated useful lives.

Figure 1 illustrates how sensitive AI capital deepening is to the useful life of the hardware. Under a two-year economic life, the net-to-gross ratio never exceeds 20%; more than four-fifths of every dollar spent is replacement rather than deepening. Even under a seven-year life, the ratio remains below 60%. Where the true economic life falls within this range determines how much headline capex adds to a hyperscaler’s enduring capital and how much is a treadmill of spending to maintain the existing stock.

Figure 2. Gross vs. Net Capex by Depreciation Scenario



Source: Authors' calculations based on Alphabet, Amazon, Meta, and Microsoft 10-K filings (SEC EDGAR) and Bloomberg consensus estimates for 2025–2026. Scenarios assume straight-line depreciation over the stated useful lives. Highlighted scenarios bracket the plausible range from AI-specific obsolescence to traditional IT assumptions.



Figure 2 translates the ratios into dollar terms. The gray bars show gross capex rose from \$175 billion in 2023 to a projected \$650 billion in 2026. The colored bars show what remains as net addition to capital under each depreciation assumption. Even with the most optimistic depreciation assumption, the gap between gross and net is surprisingly wide. If the chips generate most of their economic return in the first three years, net capital formation in 2026 will be approximately \$215 billion, only a third of the headline figure of \$650B.

A Physical Limit on AI Growth

Innovation determines how fast hardware becomes obsolete. Power increasingly determines how much hardware can be deployed. Electricity demand from data centers is rising faster than grid expansion. Interconnection timelines in many regions stretch five to seven years.⁸ When megawatts become the binding constraint, growth is no longer purely a function of capital spending.

Under fixed power, newer chips that deliver more compute per watt create an incentive to substitute rather than add. Consider a stylized example. A 50-megawatt data center operating at capacity cannot draw additional power from the grid. If a new generation of chips delivers twice the compute per watt, replacing the installed base could double compute without increasing electricity usage. Gross capital spending rises, and aggregate compute surges, but economically the firm has largely substituted new hardware for old rather than expanded its power-constrained footprint.

In a power-constrained environment, an older, less efficient chip becomes a liability because it consumes megawatts that could power a far more productive modern equivalent. When every firm pursues the same strategy, capital intensity rises without proportional expansion of economically valuable capacity. That is capital churn.

“When every firm pursues the same strategy, capital intensity rises without proportional expansion of economically valuable capacity. That is capital churn.”

Productivity J-curve

Even if infrastructure providers become far more capital intensive with lower returns, the rapid growth of compute may allow the broader economy to realize substantial productivity growth, but with a lag. Technology waves often exhibit long lags between infrastructure investment and measurable productivity gains (the J-curve). Firms must reorganize to exploit new capabilities. That process takes time.

AI hardware appears to generate most of its economic return within roughly three years. Organizational transformation in the broad economy may take much longer. If the hardware's high-return window passes before productivity gains materialize, customers may not be able to pay enough to cover the full cost of that compute.

AI users today pay less than the full economic cost of the compute they consume. Industry analyses show that AI compute is already a major drag on margins, and leading AI firms themselves project multibillion-dollar operating losses.² Hyperscalers appear willing to absorb these losses to accelerate adoption, defend market share, and lock users into their platforms.

Defending the Moat

Elevated capital spending does not necessarily signal poor decision making at the firm level. In industries shaped by scale and network effects, companies often invest not to maximize short-term return, but to protect their competitive position.

Heavy investment by hyperscalers may be entirely rational at the firm level. No major platform can afford to fall behind in AI capability. The strategic cost of under-investment may exceed the financial cost of over-investment.

For hyperscalers, advanced AI capability is increasingly part of maintaining market position. Amazon's cloud business, Microsoft's enterprise software licensing fees, Alphabet's advertising revenue, and Meta's dominant social media platform all depend on staying ahead of their rivals.

This rational behavior at the firm level can destroy returns at the industry level. When multiple dominant platforms invest simultaneously to defend share, the result is an arms race in capital intensity with no guarantee of pricing power. The pattern is familiar. During the telecom fiber buildout of the late 1990s, each operator expanded capacity to avoid strategic disadvantage. Industry capacity exploded, prices collapsed, and investors absorbed the losses.

AI infrastructure risks a similar fate. If hyperscalers continually reinvest to match one another's capabilities, aggregate compute may soar while returns on capital compress. In that case, extraordinary capex will protect competitive position but not necessarily generate extraordinary shareholder returns.

A Different Outcome?

The capital churn interpretation is not inevitable. For today's AI spending to compound rather than recycle, several demanding conditions must hold simultaneously. Demand for compute must grow faster than hardware becomes economically obsolete. Pricing must rise to reflect full economic cost rather than remain suppressed by competitive subsidy and share defense. Power constraints must ease, allowing expansion rather than substitution. And competitive intensity must moderate.

Each of these outcomes is possible. Together they set a high bar. If even one fails, a significant share of today's capex will function less as capital deepening and more as capital churn.

Implications for Investors

For investors, the question is not how much hyperscalers are spending, but what that spending earns. If frontier AI hardware generates most of its economic return within three years, then a large share of reported capex is maintenance, not expansion, and reported earnings may overstate economic profit.

AI may transform the economy. That does not guarantee that the firms building the infrastructure will capture the surplus. When capital turns over rapidly and competition forces continuous reinvestment, extraordinary spending can sustain competitive position without creating value for shareholders. The scale of investment is historic. Whether the returns will be is far less certain.

“AI may transform the economy. That does not guarantee that the firms building the infrastructure will capture the surplus.”

End Notes

1. Alphabet, Amazon, Meta, and Microsoft.
2. Alyssa Lukpat and Ben Glickman, "Big Tech Is Becoming the New Steel and Railroads," *The Wall Street Journal*, January 28, 2026.
3. BEA, "BEA Depreciation Estimates," (industrial/manufacturing structures: ~40–45 years; railroad structures: 30–60 years by component). Hyperscaler disclosed lives of five to six years: Alphabet, Amazon, Meta, Microsoft, and Oracle 10-K/10-Q filings (SEC EDGAR).
4. Alphabet, Amazon, Meta, Microsoft, and Oracle 10-K/10-Q filings (SEC EDGAR). For recent changes and earnings impacts, see Alex Usvyatsky, Deep Quarry, February and December 2025.
5. Nvidia investor presentations and Jensen Huang remarks at CES 2026 and Nvidia Q2 FY2026 earnings call, August 2025.
6. Silicon Data H100 Rental Index (SDH100RT) via Bloomberg Terminal. Hyperscaler on-demand and index rates remained in the \$2–\$3 range as of December 2025. The sub-\$1 figure refers to the open/spot market in 2026.
7. Annual revenue equals hourly rental rate \times 8,760 hours \times 60% utilization. Total cost equals straight-line depreciation (five-year life), operating expense, and an 8% capital charge applied to beginning-of-year book value. Economic profit equals revenue minus total cost; ROCE equals economic profit divided by beginning-of-year book value.
8. Electricity demand projections: S&P Global (November 2025); Goldman Sachs Global Investment Research (February 2025). Interconnection timelines: Lawrence Berkeley National Laboratory, "Queued Up: 2025 Edition" (emp.lbl.gov); Landgate (January 2026). Nadella quote on stranded GPUs: Bg2 Pod interview, reported by *Data Center Dynamics*, November 2025.
9. OpenAI financial projections: *The Wall Street Journal* and *Fortune*, November 2025 (citing documents shared with investors). Inference cost data: *The Information*, 2024–2025. Anthropic gross margin estimates: SaaStr, December 2025.

The material contained in this document is for informational purposes only. It is not intended as an offer or a solicitation for the purchase and/or sale of any security, derivative, commodity, or financial instrument, nor is it advice or a recommendation to enter into any transaction. Research results relate only to a hypothetical model of past performance (i.e., a simulation) and not to actual results or historical data of any asset management product. Hypothetical investor accounts depicted are not representative of actual client accounts. No allowance has been made for trading costs or management fees, which would reduce investment performance. Actual investment results will differ. Simulated data may have under- or over- compensated for the impact, if any, of certain market factors. Simulated returns may not reflect the impact that material economic and market factors might have had on the advisor's decision-making if the advisor were actually managing clients' money. Simulated data is subject to the fact that it is designed with the benefit of hindsight. Simulated returns carry the risk that actual performance is not as depicted due to inaccurate predictive modeling. Simulated returns cannot predict how an investment strategy will perform in the future. Simulated returns should not be considered indicative of the skill of the advisor. Investors may experience loss of all or some of their investment. Index returns represent back tested performance based on rules used in the creation of the index, are not a guarantee of future performance, and are not indicative of any specific investment. Indexes are not managed investment products and cannot be invested in directly. This material is based on information that is considered to be reliable, but Research Affiliates, LLC ("RA") and its related entities (collectively "Research Affiliates") make this information available on an "as is" basis without a duty to update, make warranties, express or implied, regarding the accuracy of the information contained herein. Research Affiliates is not responsible for any errors or omissions or for results obtained from the use of this information.

Nothing contained in this material is intended to constitute legal, tax, securities, financial or investment advice, nor an opinion regarding the appropriateness of any investment. The information contained in this material should not be acted upon without obtaining advice from a registered professional. RA is an investment adviser registered under the Investment Advisers Act of 1940 with the U.S. Securities and Exchange Commission (SEC). Our registration as an investment adviser does not imply a certain level of skill or training. RA is not a broker-dealer and does not effect transactions in securities.

Investors should be aware of the risks associated with data sources and quantitative processes used to create the content contained herein or the investment management process. Errors may exist in data acquired from third party vendors, the construction or coding of indices or model portfolios, and the construction of the spreadsheets, results or information provided. Research Affiliates takes reasonable steps to eliminate or mitigate errors and to identify data and process errors, so as to minimize the potential impact of such errors; however, Research Affiliates cannot guarantee that such errors will not occur. Use of this material is conditioned upon, and evidence of, the user's full release of Research Affiliates from any liability or responsibility for any damages that may result from any errors herein.

The trademarks Fundamental Index™, RAFI™, Research Affiliates Equity™, RAE™, and the Research Affiliates™ trademark and corporate name and all related logos are the exclusive intellectual property of RA and in some cases are registered trademarks in the U.S. and other countries. Various features of the Fundamental Index methodology, including an accounting data-based non-capitalization data processing system and method for creating and weighting an index of securities, are protected by various patents of RA. (See applicable US Patents, Patent Publications and protected trademarks located at <https://www.researchaffiliates.com/legal/disclosures#patent-trademarks-and-copyrights>, which are fully incorporated herein.) Any use of these trademarks, logos, or patented methodologies without the prior written permission of RA is expressly prohibited. RA reserves the right to take any and all necessary action to preserve all of its rights, title, and interest in and to these marks and patents.

The views and opinions expressed are those of the author and not necessarily those of RA. The opinions are subject to change without notice.

©2026 Research Affiliates, LLC. All rights reserved. Duplication or dissemination prohibited without prior written permission.

AMERICAS

Research Affiliates, LLC
660 Newport Center Drive, Suite 300
Newport Beach, California 92660
USA

+1.949.325.8700
info@researchaffiliates.com

EUROPE

Research Affiliates Global Advisors (Europe) Ltd
78-79 Pall Mall
London SW1Y 5ES
United Kingdom

+44 (0) 20 3929 9882
uk@researchaffiliates.com