HB Wealth

The Great AI Buildout: Boom, Bubble, or Both?

By: Josh Rowe, CFA, PhD, Managing Director of Research & Family Office Investment Strategy, HB Wealth

Executive Summary

*The rush to build out AI infrastructure is changing how big companies spend money, and it's having a real impact on the economy and the global energy landscape. Creative financing has caused some to worry about parallels to the turn-of-the-millennium fiber-optic boom and crash, which led to numerous bankruptcies.**

The explosion in generative AI (artificial intelligence) tools has triggered a historic wave of capital expenditure, with so-called hyperscalers like Microsoft, Google, and Meta on track to spend more than half a trillion dollars on AI infrastructure by 2029. Included in this enormous bill are data centers, specialized chips, and power grid upgrades. The energy consumption of this infrastructure is immense, impacting local power bills and potentially reshaping the domestic energy industry. While this spending has become a major driver of U.S. gross domestic product (GDP) growth in 2025, its sustainability is uncertain.

Ultimately, only end-user demand will generate the revenue to support all this investment. Today, revenue is heavily concentrated among chipmakers like Nvidia, while cloud providers operate on thin margins and software applications are still in early monetization stages. A complex web of circular investments—where tech giants fund startups that then spend on their platforms—raises concerns about inflated numbers.

Historical parallels to the fiber-optic land rush amid the dot-com boom suggest that while infrastructure overbuilds can pay off for society, there will be casualties unless downstream demand accelerates. Market valuations for chipmakers, cloud services providers, and venture-backed model builders all depend on the capex boom generating a real return in the next few years.

Key Points:

- Historic AI Investment: Hyperscalers are spending nearly $500B annually on AI Infrastructure, driving economic growth but raising sustainability questions.

- Economic Dependence: AI-related construction is now a primary driver of U.S. GDP, offsetting slower consumer activity.
- Profit Imbalance: Chipmakers like Nvidia capture most gains, while cloud and software margins remain thin.
- Circular Spending: Round-trip investments between tech giants and startups inflate revenue optics and obscure real demand.
- Short Timelines, Big Risks: With GPUs depreciating in 2-4 years and power cost surging, returns must come quickly.
- Crucial Year Ahead: The next 12-18 months will reveal whether this AI buildout is lasting infrastructure or overbuilt optimism.

Who pays for the AI spending boom?

While the stock market celebrates the arrival of generative AI (artificial intelligence) as big business, shrugging off tariffs, deficits, and government shutdowns, the bill to construct these massive data and computing centers keeps rising. Chipmakers sign contracts with ten zeroes, gleaming datacenters fly up where warehouses once stood, and model-building startups raise private rounds at mega-cap valuations. Some have rightly wondered if this massive tab will ever be paid off by downstream users of this technology, if generative AI-native applications can be so embedded in business and society to justify the infrastructure expense? And if so, how long will it take? Perhaps there is a historical precedent: the enormous outlay that went into the fiber-optic cables in the late 1990s-early 2000s paved the way for ubiquitous e-commerce and much of the mobile internet decades later. In doing so, it created an amount of economic value far outstripping its initial expense. But the payoff was more than a decade down the road and led to the bankruptcies, reorganizations, or stock market wipeouts for many who were building the state-of-the-art hardware. Others have highlighted 19th-century railroads as an episode of competitive capital destruction that eventually precipitated widespread economic flourishing. Will this era's CapEx splurge prove sustainable? Do the web of crisscrossing, circular deals between the stock market's darlings mean that investors are double or triple-counting value?

At bottom, the fundamental question of who pays for it all—and how—remains murky. The uncomfortable reality is that much of today's AI boom rests on a circular flow of money between tech giants, chip manufacturers, and venture-backed startups, with genuine end-user revenue still catching up to the extraordinary scale of investment.

How big is the infrastructure CapEx build?

Consider the sheer magnitude of spending by the "hyperscalers" alone. Microsoft alone has committed roughly $80 billion to capital expenditure this fiscal year, while Google recently increased its plans from $75 billion to $85 billion. Facebook—sorry, Meta—expects to spend between $66 and $72 billion. Add it all up, and we are talking about serious money. Hyperscaler CapEx in Q2 alone was estimated at $127 billion, implying an annualized pace close to $500 billion. This is not counting billion-dollar signing bonuses for AI researchers—this is hard *stuff*: data centers, specialized chips, and the electrical grid capacity to power them. Independent trackers report that global data center capital expenditure has surged 30-40% year-over-year, reaching record highs.

To put this in perspective, several economists now attribute a meaningful portion of recent U.S. economic growth to this AI-related construction boom. "Honey, AI CapEx is Eating the Economy" writes tech watcher Paul Kedrosky, which captures the scale of this transformation. Jason Furman calculated that without data center and information-processing investment, first-half 2025 GDP growth would have been nearly flat at 0.1%. Never bet against the U.S. consumer, we are told. Consumer typically accounts for 70% of GDP. And yet, according to Neil Dutta of Renaissance Macro Research, the dollar value of the AI data-center buildout has eclipsed consumer spending so far in 2025. This is not a normal year! Obviously, so much of the current economic momentum depends on whether this boom continues. Thus far, we have not experienced the transformative productivity boom that AI evangelists promise is coming; the economy and the stock market are running on the sugar rush of building the infrastructure itself.

Takeaway: AI-related construction is now a critical pillar of U.S. economic growth, masking weakness in traditional demand.

Where is the revenue?

Is new revenue synching up with this massive spending? Disentangling the various threads is no easy task, given that "AI companies" are making significant payments to other "AI companies" (and frequently back again). Let us go in sequence from closest to furthest from the server—hardware makers → lessors of "compute," → application software. This is also in order of relative scale.

At the top of the food chain sits Nvidia, which reported $35.6 billion in data center revenue in a single quarter, maintaining gross margins around 73%. Suppliers of inputs to Nvidia's value chain are similarly thriving: memory

manufacturer SK Hynix posted record profits and expects its high-bandwidth memory sales to double this year, while Micron reports that 2025 supply is essentially sold out.

One level down, cloud computing providers are also generating substantial AI-related revenue—Synergy Research estimates about $50 billion incremental quarterly sales across the major platforms. So, to configure and rent out this equipment (to model builders to train the next generation of AIs, to inference engines (running those models for end users), and to the applications and agents built on top), hyperscalers are annually spending about 2.5x what they recoup[1].

Here's where the economics get interesting: when Oracle's GPU rental margins became public, they revealed a sobering 14% gross margin on Nvidia-powered instances, with some operations actually running at a loss. Ignoring the purchase price, the *operating expense* of just renting and running the physical equipment is almost as high as the revenue. This is the reality of a land-grab where providers sacrifice profitability for market share.
The actual software layer—where AI supposedly delivers value to end users—fits a more modest narrative. Microsoft reports 20 million GitHub Copilot users, and various coding assistants collectively generate perhaps a few billion in annual recurring revenue. To be sure, the growth has been impressive; Anysphere's Cursor coding tool became the fastest-growing software company of all time, surpassing $100 million in its first 12 months of existence. Since this mark was reached in February, annualized revenue may have quintupled.

What Copilot is to professionals and coders, ChatGPT is to internet users. OpenAI's sales growth to approximately $12 billion leads all comers, with more than 700 million users logging in weekly. This is doubtless a staggering revenue uplift for a startup, even considering its recent half-a-trillion-dollar valuation. And yet it barely makes a dent in the industry's capital expenditure. Moreover, much of OpenAI's revenue flows right back to cloud providers for computing power. The revenue growth-valuation growth-CapEx growth merry-go-round continues.

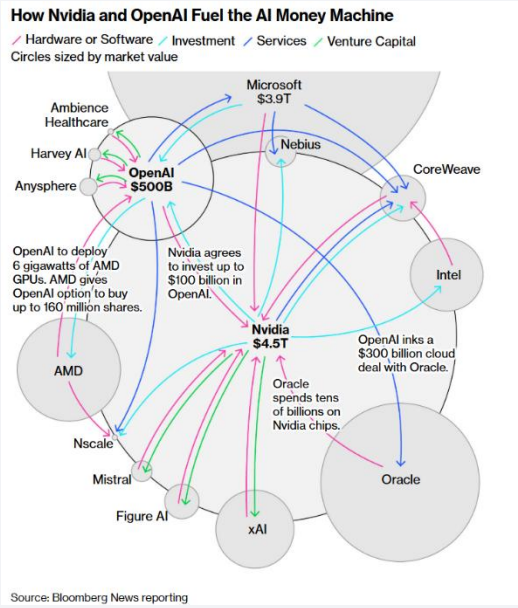
Takeaway: Revenue growth remains top-heavy, Nvidia wins big while software monetization lags far behind.

Musical chairs

What gives observers pause today is not just the scale of spending, but the intricate web of "round-trip" transactions that can blur the true source of demand. Tech investor Bill Gurley and others have highlighted how cloud providers and hardware vendors are increasingly engaging in deals that recycle capital and inflate headline revenue, sometimes without enough cash-paying third-party users to justify the optics. For example, Microsoft's

multi-billion-dollar investment in OpenAI included a substantial component of Azure cloud *credits*; OpenAI then spends those credits on Azure compute, which shows up as Microsoft's "AI revenue" growth. Amazon's $4 billion investment in Anthropic similarly comes with a commitment to make AWS the primary training partner, and the company has issued hundreds of millions in additional cloud credits to AI startups. Google, too, has invested up to $3 billion in Anthropic, locking in training and inference contracts that further entangle revenue recognition.

This creates what might charitably be called a circular economy. Microsoft invests billions in OpenAI, which spends heavily on Microsoft's Azure cloud services. Both buy chips from Nvidia, which books the revenue and uses customer prepayments to fund expanded production. Cloud providers try to recover their costs through GPU rentals and software subscriptions, but early data suggest cloud GPU rental margins are thin.



The circularity extends to hardware vendors. Nvidia's $6.3 billion backstop agreement with CoreWeave obligates Nvidia to purchase any unsold cloud capacity through 2032, while also being a major supplier and equity holder. These arrangements can prop up reported demand and make it difficult to distinguish genuine customer pull from deal-engineering. While such practices remain GAAP-compliant, they blur the visibility into independent demand. Gurley cautions that such "round-tripping"—where investments and credits goose your own revenues—can inflate top-line optics without any external validation from independent demand drivers.

The accounting used by today's AI companies is generally above-board and GAAP-compliant, but it has more than a whiff of tricks used by companies caught up in corporate scandals at the beginning of the millennium. Regional telco Qwest and fiber operator Qwest traded contractual rights to use each other's networks, while using inflated revenues and valuations to bid for smaller rivals. When capital market enthusiasm dampened and regulators unraveled the spaghetti, Global Crossing was forced to file for Chapter 11. Bankrupt energy trader Enron made famous such "round-trip" contracts to goose its profits. The largest telecoms company of its day, Worldcom, was found by regulators to have overstated assets by $11 billion, partly as a result of selling and repurchasing rights to use the same network capacity with no net economic benefit added.

Today's behaviors are not necessarily signs of hide-the-ball; they are, however, symptoms of an ebullient market and the leeway afforded by markets willing to suspend disbelief and plow money into AI leaders. For investors, increased caution is required. Follow the cash, separate true third-party customer revenue from credit-funded or partner-financed usage, and be wary of double-counting when both sides of a circular trade are owned or reported by the same set of investors. The risk is that, if these practices become widespread, they could mimic past cycle blow-offs—where headline growth masks underlying fragility. As Gurley puts it, "classic round-tripping" is a warning sign that demands scrutiny, especially when the durability of AI revenue growth is still unproven.
The parallel to the fiber-optic boom of 1999-2002 is hard to ignore. Then, as now, massive infrastructure investment preceded clear demand. Then, as now, the bet was that "if you build it, they will come." But there are crucial differences. Unlike fiber networks that sat dark for years, today's AI infrastructure gets immediate use from model training—those massive computational runs that teach AI systems their capabilities. The question is whether ongoing inference demand (using the trained models) will justify keeping these expensive facilities running at capacity.

The depreciation timeline adds urgency. While fiber-optic cables last for decades, GPUs become obsolete in two to four years. This compressed equipment lifetime means companies must recover their investments much faster, requiring either higher prices or massive scale—preferably both. Perhaps most significantly, power infrastructure has become a binding constraint. The AI boom has pulled forward an estimated $1.1 trillion in U.S. utility capital expenditure through 2029. This shifts profits to a new set of beneficiaries: regulated utilities and grid equipment manufacturers who can earn steady returns on these investments regardless of whether AI fulfills its promises.

All of this takes place against a backdrop where investors have bet big on the merry-go-round continuing. The S&P 500 continues to bounce around record highs, while companies borrow at some of the lowest spreads since the pandemic. Like most self-recursive cycles, the loop of heavy CapEx spending, fast growth of proportionally small revenue, and breezy economic forecasts embeds a hidden fragility. Any slowdown in revenue growth (at any layer) could slow down the pace of spending, denting the economy and the capital markets. When there is a lot of debt involved—as there is here—the unwind would be accelerated. This is the scenario Goldman Sachs has recently warned about.

Takeaway: Circular investments inflate reported growth, making it hard to separate real demand from financial engineering.

Spending is universal. Revenue and profit are lopsided.

So, who is actually making money today? The answer is surprisingly concentrated. Chip manufacturers and memory suppliers are capturing most of the profits, while cloud providers accept thin margins in hopes of future pricing power. Consulting firms like Accenture report $5.9 billion in generative AI work—real money, but small compared to infrastructure spending. Enterprise software companies are betting on "seat licenses" priced at $20-60 per user per month, selling to knowledge workers, call centers, sales departments, and financial firms. There is a huge future here, and businesses are surely just scratching the surface. But the enormous CapEx buildout requires a return fairly soon. Surveys suggest only about 5% of firms report measurable value from AI so far.

The risk when massive spending has been driven by a profits-later market-share landgrab, races against the ticking biological clock of technological obsolescence, and seeks returns on a debt-financed time schedule, is that there will be correspondingly massive disappointment if buyers fail to show up in their droves. Previous technology platform rollouts have gotten stuck on a few factors, leading to productivity lags: organizations require restructuring around the technology, strategic complements and interconnections favor incumbent technologies, and new generations of users need to get the hang of applying new tech to old (and new) business routines. Consequently, adoption tends to trickle before it floods. The enterprise computing revolution, which began in the 1950s, gave rise to the so-called Solow's paradox, after the Nobel laureate economist Robert Solow, who quipped in 1987, "we see computers everywhere but in the productivity statistics." AI use may suffer similar growing pains. If AI adoption follows a similar lag, valuations may be pricing in productivity gains that will take years to materialize.

Where we are cautious

Thus, the investment case for the hardware makers and the cloud builders ultimately rests on AI applications delivering genuine productivity gains across the economy. Proponents point to potential breakthroughs in customer service automation, coding productivity, and specialized applications in law, healthcare, and finance. For the purveyors of foundational models, monetization through advertising might be just in its infancy. The bearish case notes that we're funding this experiment in a very forgiving capital markets environment while waiting for end-user demand to materialize at scale. This latter phenomenon has, in the past, manifested over decades, not quarters.

Infrastructure spending may not necessarily define a "bubble"—circular flows of capital can and do resolve into sustainable businesses if external customers validate upfront CapEx. Per Gartner, enterprise IT spending surpassed $5 trillion globally in 2024. It has been growing consistently year after year. This is across all categories, but IT spending broadly seeks cost savings; generative AI applications and agents do just that. There is a big market out there for the taking. Today, however, margins for the infrastructure providers suggest that much spending may be aggressive and hopeful, perhaps even duplicative and economic wasteful. Nvidia and some fellow chipmakers are laughing now; it wasn't long ago that semiconductors were considered one of the most cyclical of industries. The early internet showed us that infrastructure overbuilds can eventually find justification, even if the timing and beneficiaries differ from initial expectations.

The next 12-18 months will likely determine whether this massive infrastructure bet pays off. Watch for evidence of sustained inference demand (inference-as-a-service); without consistent volume here, GPU utilization will fall short of estimates.[2] For the enterprise customer, whether enterprises are willing to pay for AI capabilities once the initial excitement fades are ultimately downstream of measured productivity improvements. Technology may be revolutionary, but the economics still need to prove themselves the old-fashioned way: through customers willing to pay more than it costs to serve them.

[1]$500 billion in annualized spend, covered by four times $50 billion in quarterly revenue.

[2]Inference-as-a-Service (Iaas) is a business model with lots of computing done at the "edge" (data near the source, rather than in a central cloud). The presumed use cases of IA-driven robotic process automation (RPA 2.0), automotive autonomy, and even medical diagnosis, fraud detection, and personalized advertising.

*Source: https://www.delloro.com/news/data-center-capex-to-grow-at-21-percent-cagr-through-2029/

If you have any questions, please reach out to your client service team, visit us at hbwealth.com, or call 404.264.1400.

Important Disclosures

All information is as of the date above unless otherwise disclosed. The information is provided for informational purposes only and should not be considered a recommendation to purchase or sell any financial instrument, product, or service sponsored by HB Wealth or its affiliates or agents. The information does not represent legal, tax, accounting, or investment advice; recipients should consult their respective advisors regarding such matters. This material may not be suitable for all investors. Neither HB Wealth nor any affiliates make any representation or warranty as to the accuracy or merit of this analysis for individual use. This information contains forward-looking statements, predictions, and forecasts ("forward-looking statements") concerning our belief and opinions in respect to the future. Forward-looking statements involve risks and uncertainties, and undue reliance should not be placed on them. There can be no assurance that forward-looking statements will prove to be accurate, and actual results and future events could differ materially from those anticipated in such statements. Certain information herein is based on third-party sources believed to be reliable, but which have not been independently verified. Past performance is not a guarantee or indicator of future results; inherent in any investment is the risk of loss. Specific investments described herein do not represent all investment decisions made by the above date. The reader should not assume that investment decisions identified and discussed were or will be profitable. Specific investment advice references provided herein are for illustrative purposes only and are not necessarily representative of investments that will be made in the future. Investors are advised to consult with their investment professional about their specific financial needs and goals before making any investment decision. HB may hold positions (long or short) in the companies mentioned in this paper.