

AI Bubble or Platform Shift?

An Evidence-Based Assessment of the LLM Economy

Prepared on March 14, 2026

Executive Summary

My conclusion is that AI, especially the LLM-centered part of the industry, does contain a meaningful bubble, but it is not a fake-technology bubble. It is better understood as a real platform shift with a speculative financing and infrastructure layer built on top of it. The technology is delivering real demand: NVIDIA, AMD, SK hynix, TSMC, ASML, Microsoft, Amazon, Alphabet, Oracle, OpenAI, and Anthropic all report strong AI-linked revenue, backlog, usage, or capacity signals.^{[1][2][3][4][5][6][7][8][9][10][11][12]}

At the same time, the monetization path is less mature than the investment path. Hyperscalers are committing tens of billions of dollars of annual capex, model labs are raising capital at extraordinary valuations, and infrastructure vendors are booking growth before many end-use business models are fully proven.^{[6][7][8][13][14][15]} This mismatch is the core of the bubble thesis.

My base-case judgment, clearly an inference rather than a reported fact, is that roughly 25% to 40% of current AI ecosystem enthusiasm, valuation premium, and near-term capital deployment reflects bubble-like forward pricing rather than cash flows that are likely to be realized on schedule. That does not mean AI adoption is illusory. It means the market is pricing a valid long-term technology too aggressively in the near term.

The closest historical analogy is not a pure repeat of the 1999-2000 internet bubble. It looks more like a hybrid of the dot-com boom and the telecom/fiber overbuild cycle: a real general-purpose technology, genuine long-run winners, but too much capital chasing too few near-term monetization outlets.

Research Question and Method

This report asks three questions:

- Does the current AI buildout qualify as a bubble?
- If so, how large is the bubble component?
- Is it similar to the internet bubble around 2000?

The method is simple: separate the stack into upstream suppliers, cloud and data-center infrastructure, model/application companies, academic and scientific commentary, and bank research. I then compare claims about demand, revenue, backlog, financing, and productivity. Sources are current as of March 14, 2026 and are weighted toward official company disclosures and primary research.

What the Upstream Supply Chain Is Signaling

The upstream hardware chain is not behaving as if demand is fictional. NVIDIA reported fiscal 2026 revenue of \$215.9 billion and data-center revenue of \$193.7 billion, with management describing compute demand as growing exponentially.^[1] AMD reported record 2025 revenue of \$34.6 billion, with data-center revenue of \$16.6 billion and management emphasizing strong demand for high-performance and AI platforms.^[2]

Memory and manufacturing signals point in the same direction. SK hynix reported record 2025 results driven by AI memory and HBM.^[3] TSMC reported 2025 revenue up 31.6% year over year and very strong fourth-quarter profitability, reflecting sustained advanced-node demand.^[4] ASML reported a record 2025 and said customers had become more positive on the medium-term outlook because of the sustainability of AI-related demand.^[5]

These are not the financial fingerprints of a fake market. They are the fingerprints of a real capital cycle. The hard question is not whether demand exists. It is whether current demand is durable enough, broad enough, and profitable enough downstream to justify the speed and scale of investment upstream.

What Cloud and Data-Center Infrastructure Is Signaling

The hyperscalers are making an extraordinary industrial bet. Alphabet said 2026 capex is expected to be \$175 billion to \$185 billion, after already spending \$91.4 billion in 2025, with backlog reaching \$240 billion and cloud revenue growing 48% in the fourth quarter.^[6] Microsoft reported FY26 Q2 capital expenditures of \$37.5 billion in one quarter alone, said customer demand still exceeds supply, and disclosed that roughly 45% of its commercial RPO balance is from OpenAI.^[7]

Amazon's 2024 shareholder letter is especially revealing. Andy Jassy argues that AI will reinvent virtually every customer experience, says Amazon is building more than 1,000 GenAI applications internally, and openly acknowledges that the company is spending large amounts of capital up front on data centers and chips because AI demand is growing at triple-digit rates and already represents a multi-billion-dollar annual revenue run rate.^[8] Oracle, meanwhile, is positioning itself as an AI infrastructure landlord and operator for OpenAI-related capacity, including Stargate-linked sites and new campuses provisioned for OpenAI workloads.^[9]

This layer of evidence shows two things at once. First, the spend is not imaginary; it is backed by companies with real balance sheets. Second, the sector has become dependent on a very small number of counterparties and use cases. That concentration is a classic bubble risk amplifier.

What the Application Layer Is Signaling

The application layer is where the evidence becomes mixed rather than uniformly strong.

OpenAI announced \$40 billion of funding at a \$300 billion post-money valuation in March 2025.^[10] In January 2025 it announced Stargate, a plan to deploy \$500 billion over four years for AI infrastructure, with SoftBank, OpenAI, Oracle, and MGX as equity funders and Arm, Microsoft, NVIDIA, Oracle, and OpenAI as key initial technology partners.^[11] In February 2026 OpenAI announced another \$110 billion in new investment at a \$730 billion pre-money valuation, including \$30 billion from NVIDIA and \$50 billion from Amazon, alongside a strategic partnership with Amazon and next-generation inference compute secured from NVIDIA.^[12]

Anthropic shows the same pattern of real adoption mixed with extraordinary financing. Amazon and Anthropic deepened their strategic collaboration with AWS as Anthropic's primary cloud and training partner.^[13] Anthropic's own disclosures show run-rate revenue rising above \$5 billion by August 2025 and then to \$14 billion by early 2026, alongside very rapid enterprise customer growth and funding rounds at very high valuations.^{[14][15]}

These are serious businesses with real users and real revenue. But the corporate structure matters. A growing share of the ecosystem is built on interlocking agreements in which cloud providers invest in model labs, model labs commit their training and inference workloads back to those same clouds, and chip suppliers both

sell into and financially support the labs that drive GPU demand.^{[11][12][13]} That is not necessarily fraudulent or even irrational. It may be an efficient way to synchronize capital, supply, and demand in a compute-constrained market. But it does reduce the independence of market signals. Revenue, financing, and capacity commitments can start validating one another in a loop.

That is the strongest basis for the user's suspicion of a "left hand to right hand" circularity. The evidence supports the existence of strategic financing loops. It does not prove that the demand is fake. It does show that some demand is being accelerated, financed, and institutionally reinforced by the same small cluster of actors.

What Academic and Scientific Voices Suggest

The academic and scientific evidence cuts against both extreme narratives.

On the bullish side, Erik Brynjolfsson, Daniel Rock, and Chad Syverson's "Productivity J-Curve" framework is directly relevant.^[16] Their core point is that general-purpose technologies often require large complementary investments before productivity shows up cleanly in measured output. That means weak near-term productivity data does not, by itself, prove the investment boom is irrational.

Anthropic's Economic Index and related labor-market work also show that AI is already being used in a meaningful, measurable way across real tasks, especially software and knowledge work, with augmentation still slightly ahead of automation.^{[17][18]} That matters because it suggests adoption is not merely a valuation story; it is a workflow story.

On the skeptical side, Yann LeCun argues that the current LLM paradigm has a short shelf life and may be displaced by new architectures within a few years.^[19] If he is even partly right, then parts of today's valuation stack are exposed to architecture risk: not just the risk that AI disappoints, but that current leaders and current infrastructure mixes are not the final winners.

Taken together, the academic and scientific view is not "AI is a fraud" and not "all spending is rational." It is closer to this: AI is real, important, and increasingly deployed, but the timing, measurement, and eventual architecture of value capture remain highly uncertain.

What the Banks Are Saying

Goldman Sachs offers the cleanest statement of the cautious case. Its "Gen AI: too much spend, too little benefit?" report argues that around \$1 trillion of capex may be committed in coming years while the realized benefits remain limited so far.^[20] At the same time, Goldman still believes AI can eventually raise GDP and productivity and continues to publish constructive long-run scenarios for adoption, power demand, and AI-enabled software growth.^{[21][22][23]}

Morgan Stanley is similarly bifurcated. On one hand, it warns that the AI capex boom is testing credit markets and has noted pressure on free cash flow and valuation quality among some AI spenders.^{[24][25]} On the other hand, it projects GenAI revenue could exceed \$1 trillion by 2028, sees measurable productivity gains among adopters, and argues that broad AI adoption could lift corporate value materially over time.^{[26][27][28]}

This matters because the more serious bank research is not describing a simple binary choice between bubble and no bubble. The emerging consensus is that a genuine technological shift is underway, but investment and valuation are running ahead of proven monetization.

So Is There a Bubble?

Yes, but it is layered.

There is little evidence that the entire AI complex is a classic empty bubble. The upstream revenue, backlog, and utilization data are too strong for that claim. The core suppliers and hyperscalers are also dramatically stronger businesses than many late-1990s internet companies were. NVIDIA, Microsoft, Alphabet, Amazon, Oracle, TSMC, ASML, and SK hynix are not speculative shells. They are highly profitable incumbents funding AI expansion from substantial operating businesses.^{[1][3][5][6][7][8]}

The bubble is concentrated in four places:

First, in timing. Capital spending is being pulled forward much faster than broad end-market monetization is being proven.

Second, in concentration. A small number of hyperscalers and frontier labs account for a disproportionate share of ecosystem demand.

Third, in financing structure. Strategic investments, long-term compute commitments, and cloud partnerships create feedback loops that can make demand look broader and more independently validated than it really is.

Fourth, in valuation. Private model-lab valuations and some public AI narratives appear to discount years of flawless scaling, adoption, and pricing power.

How Big Is the Bubble?

No source can answer this precisely, so any number here is an analytical estimate rather than a directly observed fact.

My base case is that the bubble component is material but minority-sized. A reasonable working range is 25% to 40% of current AI-sector excitement, valuation premium, and near-term capital deployment. In plain English: most of the AI story is real, but a substantial fraction of what the market is paying for today assumes monetization curves that are too smooth, too fast, and too concentrated in the hands of a few winners.

I would define the bubble component not as "all spending that will never earn a return," but as spending and valuation that is unlikely to earn the return implied by today's expectations on today's timetable.

That estimate is supported by three observations:

- Spending is industrial-scale and immediate, while broad productivity proof is still partial.^{[6][7][8][20]}
- Adoption is real, but concentrated in software, coding, and knowledge work rather than already diffused across the full economy.^{[17][18]}
- Financing loops now connect chip vendors, cloud providers, and model labs in ways that can amplify one another's numbers before final demand is fully diversified.^{[11][12][13]}

Is This Like the 2000 Internet Bubble?

Yes in psychology, no in underlying business quality, and only partly in market structure.

It is similar to 2000 in the following ways:

- Investors are extrapolating a real technological transformation into aggressive near-term valuation assumptions.

- Capital is racing ahead of measured productivity.
- Narratives around total addressable market and strategic necessity are overpowering short-term cash-flow discipline.
- Some cross-holdings, partnership structures, and promotional feedback loops blur independent price discovery.

It is different from 2000 in three important ways:

- Today's core AI leaders are mostly profitable, cash-generating incumbents rather than fragile startups.
- The technology is already in mainstream use, especially in software, search, coding, and enterprise productivity.
- The infrastructure bottlenecks are physical and measurable: power, chips, networking, memory, and data-center construction. This is more tangible than much of the late-dot-com application layer.

For that reason, the better analogy is a blend of dot-com euphoria and telecom overbuild. The internet did change everything, but many investors still overpaid and much infrastructure was built ahead of demand. AI can follow the same pattern: transformative in the long run, painful for some valuations in the medium run.

Bottom Line

The best answer is not "AI is a bubble" or "AI is not a bubble." The best answer is:

AI is a real general-purpose technology shift surrounded by a meaningful but partial bubble.

The technology is real. The demand is real. The infrastructure buildout is real. The financing circularity is also real. And the market is likely pricing too much too soon in parts of the stack.

If the next three years bring strong enterprise ROI, falling inference costs, and broad adoption beyond coding and copilots, much of today's spending will look rational in hindsight. If not, the likely outcome is not the death of AI, but a valuation reset, overcapacity in parts of the infrastructure layer, and a painful separation between durable winners and narrative-driven passengers.

Source Notes

[1] NVIDIA, "NVIDIA Announces Financial Results for Fourth Quarter and Fiscal 2026," February 25, 2026. <https://nvidianews.nvidia.com/news/nvidia-announces-financial-results-for-fourth-quarter-and-fiscal-2026>

[2] AMD, "AMD Reports Fourth Quarter and Full Year 2025 Financial Results," February 3, 2026. <https://www.amd.com/en/newsroom/press-releases/2026-2-3-amd-reports-fourth-quarter-and-full-year-2025-fina.html>

[3] SK hynix, "SK hynix Announces FY25 Financial Results," January 28, 2026. <https://news.skhynix.com/sk-hynix-announces-fy25-financial-results/>

[4] TSMC, "TSMC Reports Fourth Quarter EPS of NT\$19.50," January 15, 2026. <https://pr.tsmc.com/english/news/3281>

[5] ASML, "ASML reports EUR32.7 billion total net sales and EUR9.6 billion net income in 2025," January 28, 2026. <https://www.asml.com/en/news/press-releases/2026/q4-2025-financial-results>

- [6] Alphabet, "2025 Q4 Earnings Call," February 2026.
https://abc.xyz/investor/events/event-details/2026/2025-Q4-Earnings-Call-2026-Dr_C033hS6/default.aspx
- [7] Microsoft, "FY26 Q2 Earnings Release" and "FY26 Q2 Earnings Conference Call," January 28, 2026.
<https://www.microsoft.com/en-us/Investor/earnings/FY-2026-Q2/press-release-webcast> and
<https://www.microsoft.com/en-us/investor/events/fy-2026/earnings-fy-2026-q2>
- [8] Amazon, "CEO Andy Jassy's 2024 Letter to Shareholders," 2025.
<https://www.aboutamazon.com/news/company-news/amazon-ceo-andy-jassy-2024-letter-to-shareholders>
- [9] Oracle, "Data Centers" and "Oracle is Set to Power on New Data Center in Michigan," 2025-2026.
<https://www.oracle.com/data-centers/> and <https://www.oracle.com/news/announcement/blog/oracle-is-set-to-power-on-new-data-center-in-michigan-2025-1018/>
- [10] OpenAI, "New funding to build towards AGI," March 31, 2025.
<https://openai.com/index/march-funding-updates/>
- [11] OpenAI, "Announcing The Stargate Project," January 21, 2025.
<https://openai.com/blog/announcing-the-stargate-project/>
- [12] OpenAI, "Scaling AI for everyone," February 27, 2026. <https://openai.com/index/scaling-ai-for-everyone//>
- [13] Amazon, "Amazon and Anthropic deepen strategic collaboration," 2024.
<https://www.aboutamazon.com/news/aws/amazon-invests-additional-4-billion-anthropic-ai>
- [14] Anthropic, "Anthropic expands global leadership in enterprise AI," September 26, 2025. <https://www.anthropic.com/news/anthropic-expands-global-leadership-in-enterprise-ai-naming-chris-ciauri-as-managing-director-of>
- [15] Anthropic, "Anthropic raises \$30 billion in Series G funding at \$380 billion post-money valuation," February 12, 2026.
<https://www.anthropic.com/news/anthropic-raises-30-billion-series-g-funding-380-billion-post-money-valuation>
- [16] Erik Brynjolfsson, Daniel Rock, and Chad Syverson, "The Productivity J-Curve: How Intangibles Complement General Purpose Technologies," NBER Working Paper 25148.
<https://www.nber.org/papers/w25148>
- [17] Anthropic, "The Anthropic Economic Index," February 10, 2025.
<https://www.anthropic.com/research/the-anthropic-economic-index>
- [18] Anthropic, "Labor market impacts of AI: A new measure and early evidence," March 5, 2026.
<https://www.anthropic.com/research/labor-market-impacts>
- [19] Newsweek, "Yann LeCun, Pioneer of AI, Thinks Today's LLMs Are Nearly Obsolete," April 2, 2025.
<https://www.newsweek.com/ai-impact-interview-yann-lecun-artificial-intelligence-2054237>
- [20] Goldman Sachs, "Gen AI: too much spend, too little benefit?" June 27, 2024.
<https://www.goldmansachs.com/insights/top-of-mind/gen-ai-too-much-spend-too-little-benefit>
- [21] Goldman Sachs, "AI is showing very positive signs of eventually boosting GDP and productivity," May 13, 2024. <https://www.goldmansachs.com/insights/articles/AI-is-showing-very-positive-signs-of-boosting-gdp>
- [22] Goldman Sachs, "AI to drive 165% increase in data center power demand by 2030," February 4, 2025. <https://www.goldmansachs.com/insights/articles/ai-to-drive-165-increase-in-data-center-power-demand-by-2030>

[23] Goldman Sachs, "Why AI Companies May Invest More than \$500 Billion in 2026," December 18, 2025. <https://www.goldmansachs.com/insights/articles/why-ai-companies-may-invest-more-than-500-billion-in-2026>

[24] Morgan Stanley, "AI Capex Boom Puts Credit Markets to the Test," November 21, 2025. <https://www.morganstanley.com/insights/podcasts/thoughts-on-the-market/ai-spending-credit-market-outlook-andrew-sheets>

[25] Morgan Stanley, "The Bull Market's Seventh Inning," October 1, 2025. <https://www.morganstanley.com/insights/articles/ai-spending-bull-market-2025>

[26] Morgan Stanley, "GenAI Revenue Could Surpass \$1 Trillion by 2028," March 4, 2025. <https://www.morganstanley.com/insights/articles/genai-revenue-growth-and-profitability>

[27] Morgan Stanley, "AI Could Affect 90% of Occupations," September 17, 2025. <https://www.morganstanley.com/insights/articles/ai-workplace-outlook-2H-2025>

[28] Morgan Stanley, "AI's Impact Accelerates," February 5, 2026. <https://www.morganstanley.com/insights/articles/ai-adoption-accelerates-survey-find>