Scientia Moralitas International Journal of Multidisciplinary Research
ISSN 2472-5331 (Print) | ISSN 2472-5358 (Online) | Vol. 11, No. 1, 2026
DOI:10.5281/zenodo.21499401

Big Tech Companies' Massive Capex Spending on AI Investment Plans Doesn't Add Up, and the Math Could Break the Entire Tech Industry

Kas Mayanga
Marymount University, Arlington, VA, USA
mayanga@verizon.net

ABSTRACT: The U.S. big tech companies, namely the "magnificent seven" cohort of Amazon, Microsoft, Alphabet-Google, Meta, Nvidia, Apple, and OpenAI, appear firmly positioned at the center of this new AI era and are collectively investing hundreds of billions of dollars annually in artificial intelligence infrastructure, with capex spending expected to approach $700 billion by 2026. This paper examines the sustainability of these investments, their environmental impact, vulnerabilities in the semiconductor supply chain, and societal effects. Drawing on 90 sources, including peer-reviewed studies, industry reports, and economic analyses, the research highlights a notable gap between capital spent and returns achieved, as well as rising risks of overinvestment, market dominance, and environmental damage, which often affect vulnerable communities most. The paper proposes a strategic investment framework and calls for coordinated efforts among tech companies, policymakers, and stakeholders to ensure that AI infrastructure growth remains economically sound, environmentally responsible, and fairly distributed, while addressing economic disparities.

KEYWORDS: Artificial Intelligence Infrastructure, Capital Expenditure, Big Tech Investment, AI Bubble, Semiconductor Supply Chain, Environmental Sustainability, Market Concentration

1. Introduction

The surge in artificial intelligence has spurred a marked increase in capital expenditure among major U.S. technology companies, significantly reshaping the global tech industry's competitive landscape. Between November 2022 and 2025, top American tech firms, including Microsoft, Google, Amazon, OpenAI, and Meta, invested hundreds of billions of dollars in AI infrastructure, marking a historic era of substantial capital investment (Naughton, 2024). This spending includes building hyperscale data centers, acquiring advanced semiconductor chips, developing proprietary AI models, and expanding energy infrastructure to handle increasingly demanding computational workloads (Hao, 2024).

This level of investment underscores a strategic priority for tech leaders, who have long viewed AI as a key driver of competitive advantage. Microsoft's partnership with OpenAI and the integration of AI features into its products illustrate how established companies leverage their financial resources to stay ahead in AI (Roose, 2023). Similarly, Google's significant investment in Tensor Processing Units (TPUs) and ongoing AI research demonstrate its commitment to leading AI innovation, despite increasing competition from nimble startups and well-funded competitors (Metz, 2023).

This infrastructure investment spans the entire AI value chain, not just computational resources. Companies are prioritizing talent recruitment and retention, establishing specialized AI research labs, designing custom silicon chips tailored for machine learning, and building energy-efficient facilities to meet the substantial power demands of large language models and other AI systems (Clark & Hadfield-Menell, 2020). These investments have far-reaching implications, affecting regional economic development, energy grid planning, semiconductor supply chains, and global technology competition (Khan & Mann, 2020).

As AI technology advances and its commercial use expands across industries, it becomes crucial for policymakers, investors, competitors, and society to understand the scale, distribution, and strategic rationale behind technology companies' infrastructure investments. This conceptual paper offers a detailed overview of current spending patterns, identifies key challenges in scaling AI infrastructure, and guides on building that infrastructure sustainably and responsibly (Brynjolfsson & McAfee, 2017).

1.1 Problem Statement

Investment spending on artificial intelligence is on the rise. The huge gusher of money is rushing in at a rate that is stunning Wall Street financial experts and economists. Despite AI's transformative potential and major U.S. tech companies' substantial investments, current funding trends for AI infrastructure face significant challenges related to economic sustainability, efficiency, and societal impact (West, 2023). A key concern is financial stability and the uncertainty surrounding investment returns.

Big tech companies, including Amazon, Meta, Google, OpenAI, and Microsoft, invested more than $400 billion in AI in 2025 (CNBC, 2025a). CNBC (2026b) forecasts that these companies will spend about $700 billion in 2026 alone on new data centers and energy infrastructure to support their advanced AI models. According to McKinsey (2025), data centers are expected to require about $7 trillion in capital expenditures by 2030 to meet rising demand for AI and computing power. However, the lack of clear timelines for returns and profitability creates a persistent tension between strategic investment and fiduciary duties to shareholders and investors (Woo, 2024).

OpenAI, the creator of ChatGPT, entered the AI race in November 2022 and announced plans to invest about $1.4 trillion in long-term data center agreements with partners including NVIDIA, Broadcom, Oracle, and Alphabet-Google over the next eight years. CEO Sam Altman reported that the company's annual revenue is approaching $20 billion. This growth depends heavily on increasing subscriptions from individuals and businesses using its AI services (Yahoo Finance, 2025a; CNBC, 2025b). Given current revenue and investment rates, it appears unlikely that major tech companies will recoup their investments within the next decade, given an ROI of just 3% (CNBC, 2025b).

The problem landscape is further complicated by market concentration, competitive pressures, and operational inefficiencies (Ahmed & Waheed, 2020). The substantial capital required for competitive AI infrastructure creates high barriers to entry, concentrating AI capabilities among a few well-capitalized tech giants and potentially stifling innovation by smaller companies, academic researchers, and startups (Crawford, 2021). Compounding this, many organizations struggle to achieve optimal utilization of their AI infrastructure investments because of mismatches between peak capacity needs and typical usage patterns, demand fluctuations, and the rapid obsolescence of hardware (Shwartz et al., 2020).

Beyond market dynamics, AI infrastructure faces acute supply-chain vulnerabilities and geopolitical risks. The sector relies heavily on semiconductor supply chains concentrated in politically sensitive regions, leaving it vulnerable to significant disruption (Khan & Mann, 2020). Persistent shortages of advanced chips, particularly NVIDIA's GPUs and Taiwan Semiconductor Manufacturing Company (TSMC)'s cutting-edge nodes, create serious bottlenecks even when capital is available. Geopolitical tensions, especially between the United States and China, further complicate supply-chain management and long-term planning (Nellis & Dave, 2023).

These challenges are compounded by mounting environmental and energy pressures. Training large AI models requires vast amounts of electricity, straining power grids and putting AI companies in direct conflict with their sustainability commitments (Strubell et al., 2019). It's estimated that training a single large language model can consume enough electricity in a year to power hundreds of American households, substantially expanding the sector's carbon footprint and raising urgent questions about the long-term environmental cost of the AI infrastructure boom (Patterson et al., 2021).

1.2 Significant Statement

The literature review identifies a notable gap in understanding why Big Tech's substantial capital expenditure-to-revenue ratios in AI haven't yet matched their infrastructure investments (IoT Analytics, 2026). Although these companies are heavily investing in data centers, direct AI revenue remains nascent, creating a disconnect between spending and immediate returns (Yahoo Finance, 2025a). This issue is compounded by investor concerns about slow AI monetization, high hardware costs, including chips and data centers, and the risk of overinvesting in infrastructure (Investopedia, 2026). Additionally, investors worry about the expense of "circular" deals in large tech partnerships, where hardware providers fund AI firms to buy their own products, patterns reminiscent of the bubble era, and casting doubt on the true demand for AI (NPR, 2025).

This conceptual paper contributes to existing knowledge by examining research inefficiencies, new AI funding trends, challenges, and societal impacts (West, 2023). It aims to improve understanding of the financial aspects of AI investments, including risks such as over-investment, economic effects, and uncertain returns, especially given big tech's uncertain future (The Wall Street Journal, 2025a; MIT NANDA, 2025). The author also emphasizes making this research paper more accessible to AI experts, policymakers, practitioners, and the public so that it can benefit more communities. Ultimately, solving these issues requires collaboration among tech companies, policymakers, investors, and others to ensure AI investments are sustainable and equitable (Acemoglu, 2021).

1.3 Methodology of Literature Review

Types of Review: This literature review used a systematic approach to identify, evaluate, and synthesize relevant scholarly research and sources on large-scale AI investments by tech companies.
Keywords: "Big tech + AI investment +Strategic rationale", "Capital expenditure + Artificial intelligence + Technology companies", Generative AI + Infrastructure Spending + Market dynamics", "AI competition +Platform strategy + Investment justification".
Boolean Operator: Search strategies and filters include criteria for selecting or excluding peer-reviewed sources, which helped identify eligible publications. These publications were then manually reviewed to extract relevant evidence-based information for analysis.
Total Articles Reviewed: The review process was thorough, involving the manual examination of more than 100 sources. Of these, 90 were relevant to the topic and included peer-reviewed dissertations, research articles, published papers, and studies from 2020 to 2026. These encompassed industry reports from McKinsey & Company, Deloitte, S&P Global, and Accenture, as well as research from Gartner, Forrester, and IDC. Academic contributions from Stanford HAI and MIT Technology Review, along with insights from private equity and venture capital entities such as Goldman Sachs, Menlo Ventures, Ropes & Gray, NVIDIA, and Microsoft, were included. Additionally, major news outlets such as Bloomberg, BBC, Business Insider, CNBC, The Wall Street Journal, The New York Times, Reuters, Yahoo Finance, Fortune, NPR, The Washington Post, The Atlantic Council, Associated Press, and Decoder Podcast provided reports. Government data sources from the

U.S. Census Bureau, U.S. Department of Commerce, and the Federal Reserve Bank, as well as academic databases such as ProQuest, EBSCO, ERIC, and Google Scholar, were used. The primary goal was to identify key issues surrounding U.S. tech companies' significant AI investment plans, emphasizing sources relevant to the research problem statement and insights into big tech firms' AI strategies in the U.S.

2. Literature Review Section

Current Challenges in AI Capital Investments

2.1 Massive Capital Expenditure Scale and Financial Pressure

Big Tech's massive capital expenditure (capex) on AI infrastructure is often portrayed in tech analyses as a high-stakes "gold rush," with four of the biggest names in technology, including Amazon, Microsoft, Google, and Meta (The Wall Street Journal, 2026d), investing to stay competitive, even at the risk of short-term profit margins. Additionally, a bank survey indicates that 35% of fund managers view this as overinvestment. Nonetheless, some investors remain optimistic that Big Tech's large free cash flow will enable them to handle the debt.

The scale of AI infrastructure investment has reached unprecedented levels, creating significant financial pressure even for the largest technology companies. In 2023, the combined capital expenditure of Amazon, Microsoft, Google, and Meta exceeded $150 billion, a substantial increase over historical spending levels (Woo, 2024). The four companies are projected to increase spending by 60% in 2026 compared to 2025, totaling approximately $700 billion. Amazon has indicated it will spend $200 billion in capital expenditures this year (Bloomberg, 2026a; CNBC, 2026b), while Google anticipates spending between $175 billion and $185 billion (Bloomberg, 2026a). Meta forecasts its 2026 capital expenditure in the range of $115 billion to $135 billion (Fortune, 2026), and analysts estimate Microsoft's capex for the same period at approximately $114 billion (Reuters, 2026a), though analysts' 2026 forecasts were significantly lower than actual outcomes for the other companies (Sherwood, 2026).

This extraordinary level of spending creates multiple interconnected challenges. It pressures operating margins and cash flow generation, particularly when AI-generated revenue growth has not yet matched infrastructure investment (West, 2023) and is expected to reduce free cash flow across these companies (CNBC, 2026a; Reuters, 2026a). It also creates significant opportunity costs, as capital allocated to AI infrastructure cannot be deployed to other strategic initiatives, shareholder returns, or innovation in adjacent areas. Perhaps most consequential, it establishes strong expectations among investors and stakeholders that these investments will generate substantial returns, creating reputational and strategic risks if the expected benefits fail to materialize (Naughton, 2024).

If the capital expenditure-to-revenue gap remains unaddressed by the second half of 2026, analysts anticipate a shift in investor preference from growth stocks to high-dividend and stable profit-margin stocks. References to the "AI bubble" increased by 64% from the previous quarter, accounting for 2.3% of earnings calls (IoT Analytics, 2026), and several CEOs have noted that current valuations exceed those seen at the peak of the dot-com era on a price-to-sales basis, a warning sign that warrants serious attention.

The competitive dynamics further compound these pressures. No major technology company can afford to underinvest in AI infrastructure relative to its peers without risking a competitive disadvantage, creating a collective action problem in which individually rational decisions may lead to industrywide overinvestment (Brynjolfsson & McAfee, 2017). This pattern closely mirrors historical cycles of technology investment, in which competitive pressure drove capital deployment beyond economically optimal levels, ultimately leading to market corrections and industry consolidation. This precedent raises urgent questions about the sustainability of the current AI investment boom.

2.2 Financial Structure of AI Investment Deals and Economic Concerns. Economic Concerns and Financial Structure of AI Investment Deals

Three years after ChatGPT, the AI industry remains a focal point, with many economists and financial analysts frequently invoking terms like "AI bubble" and "AI hype" to explain why AI has yet to generate substantial returns beyond a handful of tech companies. Despite these doubts, businesses remain committed to AI, driving record investments and expanding adoption, as ongoing research indicates significant productivity gains.

The scale of financial investment in AI is remarkable. In 2024, private AI funding in the U.S. reached $109.1 billion, nearly 12 times China's $9.3 billion and 24 times the U.K.'s $4.5 billion (Stanford HAI, 2025). Generative AI alone secured $33.9 billion globally, an 18.7% rise from 2023. Corporate AI spending increased from $1.7 billion in 2023 to $37 billion in 2025, making up 6% of the worldwide SaaS market (Menlo Ventures, 2025). Investment in data centers is also surging, with ongoing large inflows of capital that impress financial experts and underscore AI's expanding strategic role.

Stanford University's latest AI scorecard indicates that China has nearly caught up with the United States in AI model performance. While the U.S. continues to develop more top-tier large language models, China now surpasses the U.S. in AI research publications, underscoring the growing global competition that partly fuels increased Western investment (Stanford HAI, 2025).

The urgency is evident in the scale of infrastructure investments. Big tech firms such as Meta, Amazon, Google, and OpenAI spent more than $400 billion on AI data centers in 2025. This figure is expected to reach $700 billion by 2026, driven by the growing need to develop and operate advanced AI models. OpenAI alone aims to invest $1.4 trillion in data centers over the next eight years, with support from partners including Nvidia, Broadcom, Oracle, and Alphabet-Google (The Wall Street Journal, 2025b; Investor Place, 2026; The New York Times, 2025a).

The concentration of AI control among a few wealthy big tech companies raises concerns about economic disparity, potentially widening wealth gaps and limiting funding and opportunities for academic researchers and nonprofits engaged in high-impact work. This could exacerbate the digital divide in AI. Additionally, as financial disparities between big tech firms and smaller startups widen, AI may create new economic hurdles and barriers to entry for smaller startups (Britton, 2025).

2.2.1 Current Financial Structure of Data Center Investments

An emerging narrative from the early AI revolution suggests that big tech giants such as Google, Microsoft, Meta, Amazon, Nvidia, Apple, and OpenAI are central to this new AI era. This implies that the world's wealthiest and most influential companies are likely to grow even richer and more powerful. Although new entrants such as Perplexity and Anthropic have emerged, they are often backed by established giants; for example, Microsoft is a key investor in OpenAI, Amazon in Anthropic, and Nvidia in Perplexity (Britton, 2025).

Big tech companies like Meta and Oracle use sophisticated financial tactics to fund data center expansion without significantly affecting their balance sheets. They often use special-purpose vehicles (SPVs), primarily financed by outside investors. The SPV borrows to purchase equipment, allowing large tech companies to expand capacity while keeping debt off their balance sheets. For instance, Meta and Blue Owl Capital recently financed a $27 billion data center in Richland Parish, Louisiana, using this structure. Blue Owl secured a loan backed by Meta's lease payments, effectively a mortgage. While Meta owns part of the SPV, it receives all the data center's computing power. If the project fails, Meta will owe Blue Owl a substantial amount, making this a risky financial strategy (Reuters, 2025; NPR, 2025). Financial arrangements such as Special Purpose Vehicles (SPVs) have a complex history, especially after Enron's collapse in 2001, when the term came to public attention. Unlike in the past, companies now openly acknowledge these structures, but they shouldn't depend on them for future growth (NPR, 2025).

The current AI investment climate shows unusual trends, especially in large deals such as the $100 billion agreement between NVIDIA and OpenAI (NVIDIA, 2025). In this deal, Nvidia funds data centers to support OpenAI's operations, while OpenAI uses Nvidia's chips. NVIDIA has also suggested that it might invest another $100 billion with partners such as Microsoft to develop large-scale AI data centers powered by NVIDIA technology. Some analysts worry that this setup, in which NVIDIA supports its major customer, could artificially inflate AI demand, similar to bubble patterns seen in the dot-com era. Essentially, NVIDIA is encouraging OpenAI to buy more of its chips by funding the necessary infrastructure, a practice common on smaller scales but rare at such large levels (CNBC, 2025b). Goldman Sachs analysts note that hyperscale companies, big tech firms with extensive cloud and computing resources, have amassed $121 billion in debt over the past year, a more than 300% increase compared to typical industry debt levels (NPR, 2025).

2.2.2 Industry Perspectives and Market Behavior

NVIDIA Corporation, Chief Executive Officer Jensen Huang, defended current AI infrastructure spending as appropriate and sustainable amid investor concerns about tech companies' capital expenditures. Huang predicted that the build-out of artificial intelligence infrastructure would continue for seven to eight years, driven by extremely high demand for AI technology. Huang described any eventual investment as huge but said it would not reach that level. The remarks come as investors question the circular nature of AI partnerships, in which major tech firms fund customers who also purchase their hardware. NVIDIA shares were little changed after the comments (Yahoo Finance, 2025a).

CoreWeave evolved from a crypto mining startup to a data center provider, focusing on the booming AI industry. It now supports major AI companies with training and deployment services. OpenAI, for example, has secured billion-dollar agreements for data center chip capacity, paying partly in stock. NVIDIA, meanwhile, which also owns part of CoreWeave, has a deal that guarantees it will absorb any unused data center capacity through 2032 (NPR, 2025).

Some leading AI tech leaders are recognizing that there might be too much hype surrounding AI investments. In August 2025, OpenAI CEO Sam Altman noted that investors appear overly enthusiastic about AI, despite acknowledging its substantial importance (Fortune, 2025a). Likewise, Google CEO Sundar Pichai noted some irrationality in the AI market and warned that if the bubble bursts, no company, including Google, will be immune (BBC, 2025). MIT economist Acemoglu warns that these large deals could be unstable, like a "house of cards"(NPR, 2025).

2.2.3 Financial Viability and Economic Sustainability Concerns

Some financial experts, including those from Wall Street firms like Goldman Sachs, a major global investment bank, have expressed doubts about AI investments in their 2024 research report, arguing that leading tech companies are spending too much and seeing too little benefit. Report indicates that although these companies have recently spent hundreds of billions on AI-related capital expenses, they have generated comparatively little direct AI revenue, raising concerns about short-term ROI (Goldman Sachs, 2024). Despite the large scale of AI investments, they still make up a smaller share of the annual revenue of big tech companies. However, this share is gradually increasing, as highlighted by Deloitte's analysis of technology budgets (Deloitte, 2025). Silicon Valley tech executives believe that the high investment is justified by anticipated returns over the next 1–10 years rather than immediate gains. For example, AI's role in search revenue is expanding; still, as of early 2026, it accounted for only a small share of total search revenue, according to analysis from the IO Fund (IO Fund, 2026).

IBM CEO Arvind Krishna strongly questioned the economic sustainability of the current AI infrastructure boom. During the Decoder podcast, Krishna shared what he called "napkin math," suggesting that Big Tech's large-scale data center investments might not be financially viable (Decoder Podcast, 2025). He pointed out that building a single 1-gigawatt AI data center costs about $80 billion. With plans for big tech companies to build roughly 100 gigawatts of capacity, total capital expenditure could reach around $8 trillion. To cover the interest on this investment, industry profits would need to total approximately $800 billion annually, which Krishna considers unrealistic, stating, "*There's 'no way' you're going to get a return on that*" (Fortune, 2025; Business Insider,2025; Yahoo Finance, 2025).

To provide perspective, the technology sector's total market value within the S&P 500 is about $22 trillion, or more than 36% of the index's total of over $65.69 trillion, as of early May 2026. This makes the technology sector the largest, contributing more than 36% of the S&P 500's overall value. Its high valuation is mainly driven by the "magnificent 8 companies," including Nvidia, Apple, Microsoft, Alphabet-Google, Broadcom, Meta, Tesla, and Amazon (S&P Global, 2026; Yahoo Finance, 2026).

While Krishna recognizes that AI can boost enterprise productivity by trillions, he differentiates between practical AI applications and the quest for artificial general intelligence (AGI). He estimates only a 0 to 1% chance of success for AGI with current technology, claiming that this pursuit inflates an infrastructure bubble that contradicts fundamental economic principles (Tech Spot, 2025).

Similarly, McWilliams, a distinguished economist, warns that the current AI-driven boom in the U.S. may soon end. He stresses that AI does not generate new jobs, and the large investments in GPUs are temporary, comparing them to "digital lettuce" that wilts quickly. Essentially, these investments involve perishable assets (GPUs/hardware) that quickly become obsolete. "Technological change suggests that if you buy a GPU today, the chip is going to be outdated next year," he noted (Fortune, 2025c).

The key credit factors in the AI ecosystem include the significant timing gap between funding frontier AI labs and the potential returns on those investments, as well as the long-term profitability prospects of big tech incumbents given their substantial investments. The broader economic impact will unfold over the years, with those able to successfully reimagine their core operations through AI gaining substantial productivity advantages, while laggards risk enduring competitive losses (S&P Global, 2026).

Furthermore, investors primarily focus on how quickly they can recoup their investments, because high capital costs require higher revenue for these investments to remain viable. Looking ahead, the long-term success of AI investments depends on expanding the market by increasing technology/IT spending as a share of GDP. Recently, AI adoption has been somewhat limited, centered on funding advanced large language models (LLMs) and primarily among early adopters and high-tech companies. Still, some AI experts anticipate a surge in AI applications driven by competitive pressures that boost efficiency, reduce costs, and accelerate adoption. To realize real benefits, businesses need clear productivity gains, and consumer adoption must grow significantly (S&P Global, 2026).

2.2.4 Risks of Overinvestment and Broader Economic Impact

The surge in AI investments has kept Wall Street abuzz and fueled the narrative that they are driving growth in the U.S. economy. Recently, President Donald J.Trump has used this argument to oppose state and federal regulations for the industry. In a November 2025 post on Truth Social, he wrote, "Investment in AI is helping to make the U.S. economy the 'HOTTEST' in the world, but overregulation by the states is threatening to undermine this growth engine. We must have one federal standard instead of a patchwork of 50 state regulatory regimes" (Truth Social, 2025). Several prominent economists have lent credibility to this narrative through their analyses. Jason Furman, a professor of economics at Harvard, noted that investments in information processing equipment and software contributed to 92% of GDP growth in the first half of 2025 (Furman, 2025; Yahoo Finance, 2025b). Apart from these technology-related categories, Furman noted in a September 27, 2025, post on X.com that GDP growth would have been only 0.1% on an annualized basis, indicating a near halt (Fortune, 2025b). This highlights the growing importance of high-tech infrastructure in shaping macroeconomic outcomes. Additionally, economists at the Federal Reserve Bank of St. Louis estimated that AI-related investments accounted for 39% of GDP growth in the third quarter of 2025 (Federal Reserve Bank of St. Louis, 2025).

The Morgan Stanley (2025) analysis indicates that big tech companies plan to invest approximately $3 trillion in AI infrastructure by 2028, with their own funds covering only half of that amount. Additionally, McKinsey predicts that global investments in data centers will reach $6.7 trillion by 2030 to meet rising demand for computing power. Specifically, AI-focused data centers may require $5.2 trillion, while traditional data centers will need about $1.5 trillion. In total, nearly $7 trillion in capital is expected to be needed by 2030, underscoring the scale of this investment (McKinsey, 2025). Despite President Trump's political backing for reducing regulations and claims that investments in AI boost U.S. economic growth, some Wall Street analysts and economists are beginning to question this view. "It was an appealing narrative," remarked Joseph Briggs, an analyst at Goldman Sachs, in an interview with The Washington Post. "That may have hindered or restricted the necessity to truly explore what was unfolding" (The Washington Post, 2026).

Goldman Sachs Chief Economist Jan Hatzius said in an interview with the Atlantic Council that AI investment spending has had "basically zero" contribution to U.S. GDP growth in 2025. He noted, "We don't actually view AI investment as strongly growth positive." Hatzius added that many reports on AI's impact on GDP growth are misleading, and the actual effect is much smaller than people think. One major reason for this is that a lot of the equipment needed for AI comes from other countries. Although U.S. companies are spending billions on importing chips and hardware, these imports affect GDP calculations. He explained, "A lot of the AI investment that we're seeing in the U.S. adds to Taiwanese GDP, and it adds to Korean GDP but not really that much to U.S. GDP."

Additionally, there is currently no reliable method to measure how businesses and consumers use AI, and how it contributes to economic growth. Many business leaders claim that AI hasn't significantly improved productivity (Atlantic Council, 2026). If AI market growth stabilizes, we might quickly overbuild capacity, rendering debt worthless and causing financial losses for banks. Drawing parallels to the dot-com bubble 25 years ago, when debt financed unnecessary fiber-optic cable, tech companies seem unbothered about repeating the mistake. Spending hundreds of billions on data centers that become surplus could trigger another financial crisis (NPR, 2025).

2.2.5 Skepticism on AI's Productivity and Market Hype

Many businesses report little to no productivity or employment impact from AI use despite wide adoption. A recent survey of nearly 6,000 executives in the U.S., Europe, and Australia found that, despite 70% of firms actively using AI, about 80% reported no impact on employment or productivity (NBER, 2026). Similarly, this aligns with some researchers and scholars who remain skeptical. This skepticism is justified, as growing research from sources such as McKinsey (2025) and MIT NANDA (2025) shows that most organizations haven't yet seen AI chatbots affect their profits. Additionally, Menlo Ventures (2025) reports that only 3% of consumers are willing to pay for AI.

Furthermore, Bloomberg (2026a) reported that OpenAI failed to meet key targets for user growth and revenue. The Wall Street Journal also reported that OpenAI's CFO, Sarah Friar, told staff that the company could have difficulty paying for future computing contracts unless revenues increase. This situation could conflict with CEO Sam Altman's goal of continuing to buy computing resources as the company nears its IPO (The Wall Street Journal, 2026c; LinkedIn News, 2026). For years, Altman has sought to secure as much data-center capacity as possible, arguing that computing shortages were the primary constraint on OpenAI's growth. In 2025, he entered numerous deals committing OpenAI to about $600 billion in future expenses, tying much of the tech sector's success to OpenAI's performance (The Wall Street Journal, 2025c).

Some economists, such as Daron Acemoglu of MIT, the 2024 Nobel Memorial Prize in Economic Sciences winner, have cautioned that current AI investment levels exceed justified expectations. He noted that "AI models are being hyped up, and we're investing more than we should" (Business Insider, 2024). Yet some financial experts still argue that calling the money rushing into AI right now a bubble is "a crazy concept," and they declare that "we are on the precipice of a major, major revolution in the way companies operate." AI isn't a bubble but rather an opportunity, JPMorgan's Mary Callahan Erdoes says (CNBC, 2025c). However, some experts note that AI technology is very useful, but the pace of its improvement has more or less ground to a halt. "So, the notion that the revolution continues with the same drumbeat for the next five years is sadly mistaken," noted Kedrosky (NPR, 2025).

2.3 Energy Consumption and Environmental Sustainability

Across the country, many residents in multiple states oppose data center projects in their neighborhoods. Opposition is particularly strong in Virginia (Newport News, 2025), which leads the nation in data center construction (CBRE, 2025), and in Nevada, where data centers strain the state's limited water resources. Cooling water needs are also significant, especially for evaporative systems that draw millions of gallons annually, stressing local water supplies, notably in drought-prone areas like Nevada (Nevada Current, 2025; MIT Technology Review, 2025), and in Michigan, Arizona, and Georgia (The Guardian, 2025; KJJZ, 2025; MIT Technology Review, 2026). Interestingly, this surge in data center development was initially supported by Georgia's political leaders, who in 2018 passed legislation offering tax breaks for data centers. However, the situation has since diverged from those early expectations (MIT Technology Review, 2026).

In South Dakota, citizens of Sioux Falls expressed strong disapproval after a city council vote to allow a data center north of the city (KXLG, 2026). This backlash mainly stems from the rapid, large-scale construction of data centers, which strains power grids, raises energy costs for local consumers, and generates continuous noise pollution, a persistent, low hum from machinery that never stops, often compared to a perpetually busy highway, affecting nearby communities such as Great Oak outside Manassas in Virginia (Fauquier Prince William Times, 2025; KFYR TV, 2023).

Water consumption for data center cooling presents an additional environmental challenge. Many facilities use evaporative cooling systems that consume millions of gallons of water annually, creating pressure on local water resources, particularly in regions experiencing drought or water scarcity (Li et al., 2023). The concentration of data centers in specific geographical regions can strain local water infrastructure and create environmental justice concerns when technology companies' resource consumption competes with community needs.

The environmental and social consequences of data center expansion are unevenly distributed, with the heaviest burdens falling on the most vulnerable communities. Data center construction exposes class dynamics that drive environmental injustice, as communities least responsible for pollution often suffer most (Tennessee Lookout, 2025). Elon Musk's xAI established methane-powered generators at its Memphis data center, operating 33 methane turbines in a low-income, predominantly Black neighborhood, more than double the permitted 15 (Democracy Now, 2025). The facility consumes electricity comparable to that used by 100,000 homes and contributes significantly to local pollution, including nitrogen oxides and formaldehyde, which are linked to serious health issues (Tennessee Lookout, 2024). Memphis is increasingly seen as a "*sacrifice zone*" where primarily Black communities face higher exposure to hazardous pollutants and increased cancer rates, similar to other well-known sites such as Louisiana's Cancer Alley and Flint, Michigan. The reduction in environmental protections and the weakening of regulatory agencies have allowed unchecked pollution by powerful entities, while the top 10% of wealthiest Americans remain responsible for at least 40% of the nation's emissions (Scientific American, 2021).

The extensive energy and water needs of AI infrastructure pose significant environmental challenges for the tech industry. Training large language models like GPT-3 emits substantial carbon emissions, comparable to the lifetime of multiple automobiles (Strubell et al., 2019). A single hyperscale data center can consume 100 megawatts or more, enough to power about 80,000 homes, and the growing AI sector is increasing demand for more facilities (Jones, 2018). By 2025, AI data centers will be estimated to use 29.6 gigawatts of power, enough to run all of New York State; the U.S. leads globally with 5,427 data centers, compared to China's 449 (Stanford HAI, 2025). Wyoming's planned Meta AI data center will require more electricity than all households in the state combined (Associated Press, 2025; NPR, 2026).

In response, major tech firms are increasingly adopting nuclear power. Microsoft plans to invest $1.6 billion to upgrade the Three Mile Island reactor for AI projects (The Hill, 2026; Stanford HAI, 2025). Google is collaborating with Kairos Power to deploy 500 MW of advanced small modular reactors (SMRs) by 2035 (Kairos Power, 2025). Amazon has invested over $500 million in X-energy to develop SMRs, securing a 1.9-gigawatt agreement with Talen Energy (X-energy, 2024; Talen Energy, 2025). However, energy grid limitations challenge this approach, as many regions lack sufficient capacity for additional hyperscale facilities. The variability of renewable energy adds complexity, as balancing the inconsistent supply with the constant power needs of AI workloads remains difficult (Hao, 2024). These issues threaten corporate sustainability goals, like Microsoft's aim to be carbon negative by 2030, which clash with the rapid expansion of AI infrastructure (Microsoft Corporation, 2024).

2.4 Semiconductor Supply Chain Constraints

The U.S. remains the leader in developing top AI models, but China is quickly closing the performance gap. In 2024, U.S.-based institutions developed 40 notable large language models, while China produced 15 and Europe 3. Although the U.S. still leads in quantity, Chinese models have rapidly enhanced their quality, with differences on key benchmarks like MMLU and Human Eval shrinking from double digits in 2023 to nearly parity in 2024. China also continues to dominate in AI publications and patents. Model development is becoming more global, with significant launches from the Middle East, Latin America, and Southeast Asia (Stanford HAI, 2025). The AI ecosystem is shifting toward intense competition within the technology stack, moving away from earlier cooperation. Some experts predict AI labs will aim for self-sufficiency in chips and data infrastructure, while hyperscalers develop proprietary chips to lessen reliance on external partners (S&P Global, 2026).

Access to advanced semiconductors is a critical bottleneck to AI infrastructure deployment. Specialized chips for AI workloads, particularly high-end GPUs from NVIDIA and custom accelerators such as Google's Tensor Processing Units (TPUs), have faced severe supply constraints despite substantial investments in manufacturing capacity (Nellis & Dave, 2023). Lead times for advanced chip orders have stretched to 12 months or more, forcing technology companies to plan infrastructure deployments years in advance and limiting their ability to respond dynamically to market opportunities. The concentration of advanced semiconductor manufacturing in Taiwan compounds this vulnerability, as TSMC produces most cutting-edge chips using the most advanced process nodes, creating a single point of failure for global AI infrastructure supply chains (Khan & Mann, 2020). The rapid pace of hardware innovation further complicates investment decisions, as chip architectures evolve quickly and technology companies must balance investments in current-generation hardware against the risk that imminent next-generation releases will render recent purchases comparatively obsolete (Schwartz et al., 2020).

Geopolitical tensions are reshaping the semiconductor landscape in profound ways. Earlier this year, in 2026, Anthropic CEO Dario Amodei strongly criticized President Trump's decision to allow the export of AI chips to China, comparing it to "*selling nuclear weapons to North Korea*" and warning that supplying China with advanced AI chips could help it close the technological gap with the U.S. (Axios, 2026; Bloomberg, 2026c). In response, President Trump imposed a 25% tariff on high-end chips such as Nvidia's H200 AI processors and AMD's MI325X, likely to be shipped to China (TechCrunch, 2026; Axios, 2026). U.S. government export controls on advanced AI chips have created additional complexity, affecting technology companies' ability to deploy global infrastructure, serve international customers, and optimize resource allocation across geographies, while requiring them to navigate an increasingly intricate regulatory landscape (Cheng, 2023).

China has retaliated on multiple fronts, escalating the broader geopolitical divide over AI. It blocked Meta from acquiring Manus, a China-linked startup based in Singapore that Meta had purchased in December 2025 for $2.5 billion and quickly integrated into its systems, with China's National Development and Reform Commission ordering the deal's rescission (Bloomberg, 2026b; The Wall Street Journal, 2026a). This move reflects a deeper structural shift in which technology, data, and platforms are increasingly viewed as vital national assets, strategic tools for independence, and components of future supply chain control, marking a transition from a global marketplace to one shaped primarily by geopolitical interests (The Wall Street Journal, 2026b).

2.5 Talent Acquisition and Retention

Building and operating AI infrastructure requires specialized machine learning expertise across distributed systems, hardware engineering, data center operations, and energy management. Competition for this talent is intense, with technology companies, start-ups, academic institutions, and government agencies all seeking to attract individuals with scarce skills (Chen et al., 2021). Compensation for top AI talent has reached extraordinary levels, with some researchers and engineers commanding multi-million-dollar packages that strain traditional compensation structures and raise internal equity concerns.

The concentration of AI expertise in specific geographic regions, primarily the San Francisco Bay Area, Seattle, and a few other technology hubs, creates challenges for companies seeking to operate globally distributed infrastructure. Remote work has partially addressed this constraint, but hands-on hardware work, data center operations, and certain research activities still require physical presence, limiting access to the talent pool (Choudhury et al., 2020).

Academic partnerships and research collaborations face tensions as commercial AI development accelerates. Universities have long been sources of fundamental AI research and talent pipelines for technology companies. Yet the scale of industry's resources increasingly dwarfs academic capabilities, creating a brain drain from research institutions to commercial entities (Ahmed & Waheed, 2020). This shift may reduce the diversity of research directions and concentrate innovation within commercial frameworks oriented toward near-term applications rather than fundamental science.

2.6 Regulatory Uncertainty and Policy Fragmentation

The regulatory landscape for AI infrastructure remains fragmented and uncertain, creating planning challenges for technology companies. Environmental regulations governing data center energy use, water consumption, and emissions vary substantially across jurisdictions, requiring customized compliance strategies for each facility location (Hao, 2024). Some regions have imposed restrictions on new data center development due to concerns about resource consumption, limiting siting options, and increasing costs.

Data localization requirements and digital sovereignty policies in various countries, such as the EU General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA), mandate that certain data be stored and processed within specific geographic boundaries, requiring duplicative infrastructure investments that increase costs and reduce efficiency (Chander & Le, 2015). These requirements fragment the global AI infrastructure landscape and limit companies' ability to optimize resource allocation across geographies.

Emerging AI governance frameworks, such as the EU AI Act and the U.K. AI Act, and other proposed regulations on algorithmic accountability, transparency, and safety standards, create uncertainty about future compliance obligations. This may require infrastructure adjustments or operational changes (Calo, 2017). Technology companies need to invest in AI infrastructure while preparing for regulatory requirements still being developed, risking stranded assets or expensive retrofits if regulations impose unexpected constraints.

3. The Strategic AI Investment Model Section

3.1 The Infrastructure First Strategy

The 'picks and shovels' analogy from the gold rush applies to big tech companies building AI infrastructure, allowing them to gain value regardless of the success of specific AI products (Iansiti & Lakhani, 2020). Cloud services like Amazon AWS, Google Cloud (GCP), and Microsoft Azure earn revenue by providing AI computing resources. Proprietary chips, such as Tensor Processing Units (TPUs) and custom silicon, help lower costs but can lead to vendor lock-in. Additionally, foundation models serve as platforms for future applications, and developer ecosystems create network effects.

3.2 Platform Defense and Competitive Moats

The primary motivation is strategic defense. Big tech companies invest heavily not only for short-term AI profits but also to mitigate existential threats to their core business models posed by non-investment (Brynjolfsson & McAfee, 2014). Search engines risk disruption by AI-native competitors. Cloud providers compete to control AI workloads. Social media platforms use AI to boost engagement and improve ad targeting. Additionally, embedding AI into operating systems significantly increases switching costs.

3.3 Market Signaling and Capital Allocation

Big technology companies' substantial investment in technology signals to investors, employees, and partners that they remain leaders in innovation (Spence, 1973). This approach helps maintain high valuations, ensures continued access to capital, attracts top AI research talent, and fosters strategic partnerships. Moreover, demonstrating scale deters new entrants, reinforcing competitive advantages in the AI industry.

3.4 Option Value Theory

Big technology companies' substantial investments create various "real options," whose value stems from future flexibility rather than immediate profits (Dixit & Pindyck, 1994). The ability to adapt swiftly to advances in artificial intelligence, combined with first-mover benefits in emerging AI sectors, gives companies a strategic edge amid uncertain future conditions. Additionally, their negotiating leverage in AI deals and acquisitions, along with their ability to attract and retain top talent in competitive markets, further enhances the strategic significance of these investments.

3.5 The Productivity Multipliers Hypothesis

Internal AI deployment aims to significantly enhance productivity across an organization's functions, with cost savings often justifying the investments (Agrawal et al., 2019). Applications such as code generation (e.g., GitHub Copilot, Claude Code), large-scale content moderation, and automated customer service are delivering rapid operational improvements. Additionally, shorter R&D cycles and improved decision-making systems amplify these benefits, yielding compounded advantages from AI investments across the enterprise.

4. Recommendations

4.1 Foster Responsible Market Structure and Access

U.S. big tech firms should establish structured programs that provide subsidized or free access to AI infrastructure for academic researchers, nonprofits, and early-stage startups pursuing high-impact projects (Ahmed & Wahed, 2020). These programs can help mitigate concerns about market dominance while promoting innovation, talent development, and stronger relationships with regulators and civil society. Examples include Microsoft's AI for Good and Google's TensorFlow Research Cloud, both of which could be scaled significantly.

Additionally, major tech companies should support open-source AI models, tools, and infrastructure. A robust open-source ecosystem attracts skilled professionals, fosters positive network effects, and demonstrates responsible industry leadership, potentially avoiding tougher regulations (Crawford, 2021). Meta's release of the LLaMA models shows how companies can balance commercial goals with community benefits. Rather than resisting regulatory oversight, big tech should actively collaborate with policymakers to develop frameworks that address concerns about market concentration while encouraging innovation (Calo, 2017). Early participation in regulatory processes, through transparent information sharing, engagement in multi-stakeholder governance, and support for procompetition policies, generally leads to better outcomes than reactive responses to regulations developed without industry input.

4.2 Optimize Resource Utilization and Infrastructure Efficiency

U.S. big tech companies should invest in advanced orchestration platforms that efficiently distribute AI workloads across their infrastructure using real-time data on demand, energy costs, and the carbon footprint of electricity sources. Moving workloads to locations with abundant renewable energy or lower electricity prices can boost both financial performance and sustainability. Predictive analytics can further improve responsiveness by proactively allocating resources based on anticipated demand. In addition to smarter workload management, major tech firms should adopt integrated hardware-software strategies that optimize AI model architectures for their hardware capabilities and develop custom silicon tailored to expected workload patterns (Chen et al., 2021). Google's TPU development shows significant efficiency gains through this approach, but wider adoption requires closer collaboration among machine learning researchers, systems engineers, and hardware designers.

Furthermore, large technology companies should consider offering surplus infrastructure capacity to external clients, academic researchers, or smaller enterprises during periods of low demand (Ahmed & Wahed, 2020). While competitive concerns may curb enthusiasm for such arrangements, carefully designed programs could generate additional revenue, improve resource utilization, and foster goodwill among regulators and stakeholders worried about market concentration. Amazon Web Services has proven the commercial viability of infrastructure-as-a-service models, yet broader adoption for AI-specific capabilities remains limited.

4.3 Advance Environmental Sustainability and Energy Strategy

U.S. leading tech companies should move beyond purchasing renewable energy credits and instead invest directly in new renewable energy capacity located near or dedicated to AI infrastructure facilities (Patterson et al., 2021). This requires deploying substantial capital in solar, wind, and energy storage, possibly through collaborations with utilities or independent power producers, or by owning generation assets outright. Microsoft's renewable investments are one example, but the rapid expansion of AI infrastructure calls for even stronger strategies. At the same time, major tech companies should make energy efficiency a core factor in hardware purchasing decisions and AI model development (Strubell et al., 2019). This includes funding advanced, energy-efficient chip designs, enhancing cooling systems with methods such as liquid cooling and waste heat recovery, and developing training techniques that deliver similar results with less computation.

Research on model compression, knowledge distillation, and efficient architectures should be prioritized across organizations. Beyond individual efforts, big tech firms need to collaborate to establish industry standards for tracking and reporting the energy use and environmental impact of AI infrastructure (Hao, 2024). Transparent reporting systems would enable stakeholders to evaluate sustainability performance, share best practices, and hold companies accountable, with initiatives such as the MLPerf benchmarking effort serving as a model for joint standard-setting that could extend to environmental metrics.

4.4 Strengthen Supply Chain Resilience and Diversification

U.S. big tech companies can play a vital role in promoting a more geographically balanced semiconductor manufacturing industry by investing directly, entering into long-term supply agreements, and partnering with emerging fabrication facilities (Khan & Mann, 2020). The

U.S. Chips and Science Act, signed on August 9, 2022, offers approximately $53 billion in incentives for domestic chip manufacturing, including $39 billion in grants and a 25% tax credit, to restore semiconductor supply chains within the U.S., create jobs, support American innovation, and enhance national security by reducing dependence on foreign suppliers, especially in East Asia (U.S. Department of Commerce, 2024). Technology firms can access these opportunities through co-investment in new facilities. While it will take years and substantial capital to develop alternatives to Taiwan Semiconductor Manufacturing Company's (TSMC) advanced capabilities, diversifying sources for critical components can significantly reduce strategic vulnerabilities. Instead of designing systems tightly tailored to specific chip vendor architectures, large tech companies should develop abstraction layers and portability frameworks that allow workloads to run efficiently across various hardware platforms (Cheng, 2023).

This approach provides flexibility to switch vendors if necessary and promotes competition, encouraging better pricing and innovation among hardware suppliers. Although open-source frameworks like PyTorch and TensorFlow are already making progress toward this goal, ongoing investment in hardware-agnostic optimization remains crucial. Lastly, given the importance of AI infrastructure and ongoing supply constraints, big tech companies should maintain strategic reserves of key components, treating inventory as insurance against operational disruptions that could be far more costly (Nellis & Dave, 2023).

4.5 Enhance Coordination and Long-Term Planning

U.S. big tech companies, utilities, semiconductor manufacturers, government agencies, and policymakers should establish joint planning processes to coordinate AI infrastructure development with energy systems, semiconductor capacity, and workforce expansion (Hao, 2024). Given the large-scale capital investment and complex interdependencies, coordinated planning is preferable to purely market-based strategies. Industry consortia or government-led working groups can enable open information sharing while preserving each party's competitive independence. Beyond infrastructure coordination, big tech companies should significantly increase investment in foundational research to develop AI methods that match or exceed current performance with far lower computational requirements (Schwartz et al., 2020). The current trend of rapidly increasing model sizes and computational requirements is becoming environmentally and economically unsustainable, making breakthroughs in efficient learning, reasoning, and knowledge representation critical to a sustainable AI future. Additionally, big tech companies should publicly commit to clear, measurable goals for infrastructure efficiency, sustainability, and responsible deployment, along with regular progress reports (Patterson et al., 2021). Transparent accountability frameworks help build stakeholder trust, set performance standards, and demonstrate responsible corporate conduct, including independent third-party verification to ensure credibility and prevent greenwashing.

5. Conclusion

The extraordinary scale of AI infrastructure investment by major U.S. big tech companies represents both an unprecedented opportunity and a formidable set of challenges that will shape the technology landscape for decades to come. With combined annual capital expenditures exceeding $650 billion and projections suggesting continued acceleration, Microsoft, Google, Amazon, Meta, and other technology leaders are making bets of historic proportions on AI's transformative potential (Woo, 2024). These investments are driving remarkable advances in AI capabilities, enabling new applications across industries, and positioning the United States as the global leader in AI development and deployment.

However, the current trajectory of AI infrastructure spending faces significant headwinds that threaten its financial, environmental, and social sustainability. Energy consumption levels that conflict with climate commitments, semiconductor supply chain vulnerabilities that create strategic risks, capital requirements that concentrate capabilities among a small number of well-funded companies, and uncertain return-on-investment timelines that challenge financial justification all demand thoughtful responses from industry leaders, policymakers, and stakeholders (Strubell et al., 2019; Khan &Mann, 2020; West, 2023).

The recommendations outlined in this analysis, emphasizing resource optimization, environmental sustainability, supply chain resilience, responsible market structure, and coordinated long-term planning, provide a framework for addressing these challenges while preserving the dynamism of innovation that has characterized American technology leadership. Successfully navigating this complex landscape requires moving beyond the narrow optimization of individual corporate interests toward recognizing shared challenges that benefit from collective action and industry-wide cooperation (Brynjolfsson & McAfee, 2017). The stakes extend far beyond corporate balance sheets and competitive positioning. AI infrastructure decisions made today will influence energy systems, semiconductor supply chains, labor markets, innovation ecosystems, and international competitiveness for years to come. The environmental footprint of AI will affect climate trajectories, while decisions about infrastructure access will shape whether AI's benefits are broadly distributed or concentrated among a narrow set of actors (Crawford, 2021; Acemoglu, 2021). Big tech companies have shown great innovation, quick growth, and problem-solving skills. Using these strengths to address AI infrastructure challenges requires creativity, a long-term outlook, and a focus on results that benefit shareholders, workers, communities, and society. The next step for responsible tech leadership involves organizations that can balance strong AI investments with environmental care, supply chain stability, market access, and stakeholder trust. Such organizations are likely to become not just the top players in AI but also models for developing and using transformative technology responsibly.

As artificial intelligence (AI) capabilities continue to advance and applications proliferate across every sector of the economy, the infrastructure decisions made during this critical period will establish patterns and path dependencies that shape technological development for the foreseeable future. Ensuring these patterns promote broad human flourishing rather than narrow commercial interests necessitates sustained attention, meticulous governance, and a commitment to values extending beyond quarterly financial performance from technology leaders, policymakers, and civil society stakeholders alike.

References

Acemoglu, D. (2021). *Harms of AI* (NBER Working Paper Series, No. 29247). National Bureau of Economic Research. https://doi.org/10.3386/w29247

Agrawal, A., Gans, J., & Goldfarb, A. (Eds.). (2019). *The economics of artificial intelligence: An agenda*. University of Chicago Press.

Ahmed, N., & Wahed, M. (2020). *The de-democratization of AI: Deep learning and the compute divide in artificial intelligence research.* arXiv preprint. https://arxiv.org/abs/2010.15581

Associated Press. (2025, July 28). *Cheyenne to host massive AI data center using more electricity than all Wyoming homes combined.* https://apnews.com/article/ai-artificial-intelligence-data-center-electricity-wyoming-cheyenne-44da7974e2d942acd8bf003ebe2e855a

Atlantic Council. (2026). *AI, Supply Chains, and Trade Resets: The Global Economy in 2026*. [Video]. YouTube. https://www.youtube.com/watch?v=zZHN0-ZNe_4

Axios. (2026, January 20). *"This is crazy": Anthropic CEO blasts AI Chip Sales to China.* https://www.axios.com/2026/01/20/anthropic-ceo-admodei-nvidia-chips-china-trump

BBC News. (2025). *Google Boss Says Trillion Dollar AI Investment Boom has Elements of Irrationality*. https://www.bbc.com/news/articles/cwy7vrd8k4eo

Bloomberg. (2026a, February 6). *Big tech to spend $650 billion this year as AI race intensifies.* https://www.bloomberg.com/news/articles/2026-02-06/how-much-is-big-tech-spending-on-ai-computing-a-staggering-650-billion-in-2026

Bloomberg. (2026b, April 27). *China blocks Meta's $2 billion acquisition of AI firm Manus.* https://www.bloomberg.com/news/articles/2026-04-27/china-blocks-meta-s-2-billion-acquisition-of-ai-startup-manus

Bloomberg (2026c, January 20) *Anthropic's Amodei on AI: Power and Risk.* https://www.bloomberg.com/news/videos/2026-01-20/anthropic-ceo-dario-amodei-on-the-future-of-ai-video

Britton, M. (2025). *Generation AI: Why Generation Alpha and the Age of AI Will Change Everything*. Wiley.

Brynjolfsson, E., & McAfee, A. (2017). *The business of artificial intelligence: What it can—and cannot—do for your organization*. Harvard Business Review Press.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton & Company.

Business Insider. (2025, December 2). *IBM CEO Says There's 'no way' spending trillions on AI Data Centers will pay off at today's infrastructure costs.* https://www.businessinsider.com/ibm-ceo-big-tech-ai-capex-data-center-spending-2025-12

Business Insider. (2024, October 2). *A lot of money is going to get wasted: MIT economist sounds the alarm on AI hype.* https://www.businessinsider.com/ai-impact-overblown-investment-tech-stocks-billions-daron-acemoglu-mit-2024-10

Calo, R. (2017). Artificial intelligence policy: A primer and roadmap. *University of Bologna Law Review,* 3(2), 180–218. https://doi.org/10.6092/issn.2531-6133/7755

CBRE. (2025, September 8). *North America data center trends H1 2025*. https://www.cbre.com/insights/reports/north-america-data-center-trends-h1-2025

Chander, A., & Lê, U. P. (2015). Data nationalism. *Emory Law Journal*, *64*(3), 677–739.

Chen, T., Xu, B., Zhang, C., & Guestrin, C. (2021). *Training deep nets with sublinear memory cost*. arXiv preprint. https://arxiv.org/abs/1604.06174

Cheng, E. (2023, October 17). *U.S. tightens AI chip export controls, Nvidia and AMD shares drop.* CNBC. https://www.cnbc.com/2023/10/17/us-tightens-ai-chip-export-controls-nvidia-and-amd-shares-drop.html

Choudhury, P., Foroughi, C., & Larson, B. (2020). Work-from-anywhere: The productivity effects of geographic flexibility. *Strategic Management Journal*, *42*(4), 655–683. https://doi.org/10.1002/smj.3251

Clark, J., & Hadfield-Menell, D. (2020). *AI policy levers: A review of the U.S. government's tools to influence AI research, development, and deployment*. Center for the Governance of AI, Future of Humanity Institute, University of Oxford.

CNBC. (2025a, February 8). *Tech mega caps plan to spend more than $300 billion in 2025 as AI race intensifies.* https://www.cnbc.com/2025/02/08/tech-megacaps-to-spend-more-than-300-billion-in-2025-to-win-in-ai.html

CNBC. (2025b, October 31). *As big tech pushes AI spending to the max, you may be helping to pay for it.* https://www.cnbc.com/2025/10/31/big-tech-ai-spending-billions-microsoft-google-software-subscriptions.html

CNBC. (2025c, November 14) *AI isn't a bubble but rather an opportunity, JPMorgan's Erdoes says.* https://www.cnbc.com/2025/11/13/ai-isnt-a-bubble-but-rather-an-opportunity-jpmorgans-erdoes-says.html

CNBC. (2026a, January 26). *AI spending wasn't the biggest engine of U.S. economic growth in 2025, despite popular assumptions.* https://www.cnbc.com/2026/01/26/ai-wasnt-the-biggest-engine-of-us-gdp-growth-in-2025.html

CNBC. (2026b, February 6). *Tech AI spending may approach $700 billion this year, but the blow to cash raises red flags*. https://www.cnbc.com/2026/02/06/google-microsoft-meta-amazon-ai-cash.html

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Decoder Podcast. (2025, December 1). *IBM CEO Arvind Krishna says there is no AI bubble after all* [Podcast episode]. The Verge. https://www.theverge.com/podcast/829868/ibm-arvind-krishna-watson-llms-ai-bubble-quantum-computing

Deloitte. (2025). *Deloitte's 2025 Tech Value Survey: AI is capturing the digital dollar. What's left for the rest of the tech estate?* https://www.deloitte.com/us/en/insights/topics/digital-transformation/ai-tech-investment-roi.html

Democracy Now. (2025, April 25). *"Musk is scamming the city of Memphis": Meet two brothers fighting Colossus, Musk's xAI data center*. https://www.democracynow.org/2025/4/25/elon_musk_xai_memphis_tennessee

Dixit, A. K., & Pindyck, R. S. (1994). *Investment under uncertainty.* Princeton University Press.

Fauquier Prince William Times. (2025). *Data center noise at Great Oak community* [Video]. YouTube. https://www.youtube.com/watch?v=9HJe_SDFNPY

Federal Reserve Bank of St. Louis. (2025, January 12). *Tracking AI's contribution to GDP growth.* https://www.stlouisfed.org/on-the-economy/2026/jan/tracking-ai-contribution-gdp-growth

Fortune. (2025, December 3). *IBM CEO warns there's "no way" hyperscalers like Google and Amazon will be able to turn a profit at the rate of their data center spending.* https://fortune.com/2025/12/03/ibm-ceo-no-way-hyperscalers-google-amazon-turn-profit-data-center-spending/

Fortune. (2025a, August 19). *Wall Street isn't worried about an AI bubble. Sam Altman is.* https://fortune.com/2025/08/19/wall-street-ai-bubble-sam-altman/

Fortune. (2025b, October 7). *Without data centers, GDP growth was 0.1% in the first half of 2025, Harvard economist says.* https://fortune.com/2025/10/07/data-centers-gdp-growth-zero-first-half-2025-jason-furman-harvard-economist/

Fortune. (2025c, November 20). *One of the world's most popular economists on why AI is "undoubtedly going to crash."* https://fortune.com/2025/11/20/david-mcwilliams-interview-history-of-money-ai-bubble-crash-economy/

Fortune. (2026, February 6). *Big tech's $630 billion AI spree now rivals Sweden's economy, unsettling investors.* https://fortune.com/2026/02/06/what-is-a-data-center-capex-spending-630-billion-dollars-amazon-microsoft-google-meta/

Furman, J. (2025, September 27). *Investment in information processing equipment & software is 4% of GDP* [Post]. X. https://x.com/jasonfurman/status/1971995367202775284

Goldman Sachs. (2024, June 25). *Gen AI; Too Much Spend, Too Little Benefit?* https://www.goldmansachs.com/insights/top-of-mind/gen-ai-too-much-spend-too-little-benefit

Google & BBC Politics. (2025). *Full interview: Google CEO Sundar Pichai on the 'AI boom' and the future of AI* [Video]. YouTube. https://www.youtube.com/watch?v=BYx63PKKPvg

The Guardian. (2025, December 18). *Uniquely evil: Michigan residents fight against huge data center backed by top tycoons.* https://www.theguardian.com/us-news/2025/dec/18/michigan-data-center-fight

Hao, K. (2024, March 10). *The AI boom is here, and so are the costs*. The Washington Post. https://www.washingtonpost.com/technology/

The Hill. (2026). *Microsoft wants to resurrect Three Mile Island. It will never happen*. https://thehill.com/opinion/energy-environment/5667831-microsoft-constellation-nuclear-challenges/

Iansiti, M., & Lakhani, K. R. (2020). *Competing in the age of AI: Strategy and leadership when algorithms and networks run the world.* Harvard Business Review Press.

Investor Place. (2026). *AI's $700 Billion Question.* https://investorplace.com/2026/04/ais-700-billion-question/

Investopedia. (2026, February 17). *Investing Experts Sound Alarms On Big Tech's Massive AI Spending*. https://www.investopedia.com/investing-experts-sound-alarm-on-big-tech-massive-ai-spending-11907836

IO Fund. (2026). *The $530 billion AI Question: Which Big Tech Stock Is Winning?* https://io-fund.com/ai-stocks/big-tech-ai-stock-winner

IoT Analytics. (2026). *AI bubble references in earnings calls* [Research report]. https://iot-analytics.com/what-ceos-talked-about-in-q1-2026-iran-war-physical-agentic-ai-openclaw-and-the-saaspocalypse/

Jones, N. (2018). How to stop data centers from gobbling up the world's electricity. Nature*,* 561(7722), 163–166. https://doi.org/10.1038/d41586-018-06610-y

Kairos Power. (2024, October 14). *Google and Kairos Power partner to deploy 500 MW of clean electricity generation.* https://www.kairospower.com/updates/google-and-kairos-power-partner-to-deploy-500-mw-of-clean-electricity-generation

KFYR-TV. (2023, June 22). *Atlas Power data center noise pollution* [Video]. YouTube. https://www.youtube.com/watch?v=9HJe_SDFNPY

Khan, L. M., & Mann, K. (2020). *Competition and the future of computing: The semiconductor industry after Moore's Law*. Yale Law School.

KJZZ. (2025, December 9). *Rejected by Tucson and abandoned by Amazon, Project Blue data center moves forward.* https://www.kjzz.org/the-show/2025-12-09/rejected-by-tucson-and-abandoned-by-amazon-project-blue-data-center-moves-forward

KXLG. (2026, January 7). *Tensions run high at Sioux Falls City Council meeting*. https://www.mykxlg.com/news/state/tensions-run-high-at-sioux-falls-city-council-meeting/article_eb707777-46fc-4734-82a3-33cda5d214e3.html

Li, P., Yang, J., Islam, M. A., & Ren, S. (2023). *Making AI less "thirsty": Uncovering and addressing the secret water footprint of AI models.* arXiv preprint. https://arxiv.org/abs/2304.03271

LinkedIn News. (2026). *OpenAI rebuffs claims it missed key revenue target*. https://www.linkedin.com/news/story/openai-rebuffs-claims-it-missed-key-revenue-target-8734298/

Menlo Ventures. (2025). *2025: The state of generative AI in the enterprise*. https://menlovc.com/perspective/2025-the-of-generative-ai-in-the-enterprise/

Metz, C. (2023, April 20). *Google and the race to dominate artificial intelligence*. The New York Times. https://www.nytimes.com/2023/04/20/technology/

McKinsey& Company. (2025, April 28). *The cost of compute: A $7 trillion race to scale data centers*. https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/the-cost-of-compute-a-7-trillion-dollar-race-to-scale-data-centers#/

Microsoft Corporation. (2024). *Annual report pursuant to Section 13 or 15(d) of the Securities Exchange Act of 1934 for the fiscal year ended June 30, 2024*. U.S. Securities and Exchange Commission.

MIT NANDA. (2025, July). *The GenAI Divide: State of AI in Business 2025*. https://mlq.ai/media/quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf

MIT Technology Review. (2025, May 20). *The data center boom in the desert*. https://www.technologyreview.com/2025/05/20/1116287/ai-data-centers-nevada-water-reno-computing-environmental-impact/

MIT Technology Review. (2026, January 14). *Data centers are amazing. Everyone hates them*. https://www.technologyreview.com/2026/01/14/1131253/data-centers-are-amazing-everyone-hates-them/

Morgan Stanley. (2025, July 25). *Who will fund AI's $3 trillion ask?* [Podcast]. Thoughts on the Market. https://www.morganstanley.com/insights/podcasts/thoughts-on-the-market/ai-investing-credit-markets-andrew-sheets

Naughton, J. (2024, February 4). *The tech giants are pouring billions into AI—and not everyone is happy about it.* The Guardian. https://www.theguardian.com/technology/

NBER. (2026, February). *Firm data on AI (No. W34836; p. W34836*). National Bureau of Economic Research. https://doi.org/10.3386/w34836

Nellis, S., & Dave, P. (2023, August 9). *Nvidia, AMD face export restrictions to some Middle East countries*. Reuters. https://www.reuters.com/technology/

Nevada Current. (2025, January 10). *Effort to block Reno data center offers glimpse of increasingly thorny issue in Nevada.* https://nevadacurrent.com/2025/01/10/effort-to-block-reno-data-center-offers-glimpse-of-increasingly-thorny-issue-in-nevada/

Newport News. (2025, December 24). *Newport News residents rally against proposed data center near Fort Eustis.* https://www.13newsnow.com/article/news/local/mycity/newport-news/newport-news-residents-rally-against-proposed-data-center-near-fort-eustis/291-6069910e-753f-4770-8db3-056ab688120f

NPR. (2025, November 23). *Here's why concerns about an AI bubble are bigger than ever*. https://www.npr.org/2025/11/23/nx-s1-5615410/ai-bubble-nvidia-openai-revenue-bust-data-centers

NPR. (2026, January 2). *AI data centers use a lot of electricity. How it could affect your power bill*. https://www.npr.org/2026/01/02/nx-s1-5638587/ai-data-centers-use-a-lot-of-electricity-how-it-could-affect-your-power-bill

NVIDIA. (2025, September 22). *OpenAI and NVIDIA announce strategic partnership to deploy 10 gigawatts of NVIDIA systems* [Press release]. https://nvidianews.nvidia.com/news/openai-and-nvidia-announce-strategic-partnership-to-deploy-10gw-of-nvidia-systems

The New York Times. (2025a, September 23). *OpenAI to join tech giants in building 5 new data centers in U.S.* https://www.nytimes.com/2025/09/23/technology/openai-data-centers-united-states.html

Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L. M., Rothchild, D., So, D., Texier, M., & Dean, J. (2021). *Carbon emissions and large neural network training.* arXiv preprint. https://arxiv.org/abs/2104.10350

Reuters. (2025, October 21). *Meta in $27 billion financing deal with Blue Owl Capital for Louisiana data center.* https://www.reuters.com/technology/meta-forms-joint-venture-with-blue-owl-capital-louisiana-data-center-2025-10-21/

Reuters. (2026a, February 6). *AI spending may approach $700 billion this year, but the blow to cash raises red flags.* https://www.reuters.com/business/global-software-data-firms-slide-ai-disruption-fears-compound-jitters-over-600-2026-02-06/

Roose, K. (2023, March 21). *How Microsoft's bet on OpenAI is playing out*. The New York Times. https://www.nytimes.com/2023/03/21/technology/

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics, 87*(3), 355–374. https://doi.org/10.2307/1882010

S&P Global. (2026, January 21). *Where are AI investment risks hiding?* https://www.spglobal.com/ratings/en/regulatory/article/where-are-ai-investment-risks-hiding-s101665242

Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, *63*(12), 54–63. https://doi.org/10.1145/3381831

Scientific American. (2021, December 8). *Rich Americans have higher carbon footprints than other wealthy people.* https://www.scientificamerican.com/article/rich-americans-have-higher-carbon-footprints-than-other-wealthy-people/

Sherwood News. (2026). *Apple is the only big tech company whose capex declined last quarter*. https://sherwood.news/tech/apple-is-the-only-big-tech-company-whose-capex-declined-last-quarter/

Stanford University Human-Centered AI Institute (HAI). (2025). *The AI index 2025 annual report*. https://doi.org/10.48550/arXiv.2504.07139

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3645–3650. https://doi.org/10.18653/v1/P19-1355

Talen Energy. (2025, June 11). *Talen Energy expands nuclear energy relationship with Amazon*. https://ir.talenenergy.com/news-releases/news-release-details/talen-energy-expands-nuclear-energy-relationship-amazon

Tech Crunch. (2026, January 20th). *Anthropic's CEO stuns Davos with Nvidia criticism*.https://techcrunch.com/2026/01/20/anthropics-ceo-stuns-davos-with-nvidia-criticism/

Tech Spot. (December 5, 2025). *AMD and IBM's CEO doesn't see an AI bubble, just $8 trillion in data centers.*https://www.techspot.com/news/110508-amd-ibm-ceo-doesnt-see-ai-bubble-8.html

Tennessee Lookout. (2024, July 23). *Due diligence: Questions surround Musk's xAI plans*. https://tennesseelookout.com/2024/07/23/due-diligence-questions-surround-musks-xai-plans/

Tennessee Lookout. (2025, July 7). *A billionaire, an AI supercomputer, toxic emissions and a Memphis community that did nothing wrong.* https://tennesseelookout.com/2025/07/07/a-billionaire-an-ai-supercomputer-toxic-emissions-and-a-memphis-community-that-did-nothing-wrong/

Truth Social. (2025, November 18). *Investment in AI is helping to make the U.S. economy the "HOTTEST" in the world [Post].* Truth Social. https://truthsocial.com/@realDonaldTrump/posts/115572931492563128

U.S. Department of Commerce. (2024, August 9). *Two Years Later: Funding from CHIPS and Science Act Creating Quality Jobs, Growing Local Economies, and Bringing Semiconductor Manufacturing Back to America.* https://www.commerce.gov/news/blog/2024/08/two-years-later-funding-chips-and-science-act-creating-quality-jobs-growing-local

The Wall Street Journal. (2025a, February 25). *The AI data center boom is a job-creation bust.* https://www.wsj.com/tech/ai-data-center-job-creation-48038b67

The Wall Street Journal. (2025b, July 31). *Big tech's $400 billion AI spending spree just got Wall Street's blessing.* https://www.wsj.com/tech/ai/tech-ai-spending-company-valuations-7b92104b

The Wall Street Journal. (2025c, October 20). *How Sam Altman tied tech's players to OpenAI.* https://www.wsj.com/tech/ai/sam-altman-open-ai-nvidia-deals-d10a6525

The Wall Street Journal. (2026a, April 27). *China bans Meta's acquisition of Manus on national security grounds.* https://www.wsj.com/world/china/china-bans-metas-acquisition-of-manus-on-national-security-grounds-71e10c3f

The Wall Street Journal. (2026b, April 27). *Meta is preparing to have to undo its Manus acquisition after China ban.* https://www.wsj.com/tech/ai/meta-is-preparing-to-have-to-undo-its-manus-acquisition-after-china-ban-a4ffbefb

The Wall Street Journal. (2026c, April 28). *OpenAI misses key revenue, user targets in high-stakes sprint toward IPO.* https://www.wsj.com/tech/ai/openai-misses-key-revenue-user-targets-in-high-stakes-sprint-toward-ipo-94a95273

The Wall Street Journal. (2026d, April 29). *Big Tech Strikes Gold With AI, but at a Steep Cost*. https://www.wsj.com/tech/ai/big-tech-strikes-gold-with-ai-but-at-a-steep-cost-f6d82a22

The Washington Post. (2026, February 23). *This economic idea transfixed Wall Street and Washington. It may be a mirage.* https://www.washingtonpost.com/technology/2026/02/23/ai-economic-growth-gdp-mirage/

West, D. M. (2023). *The risks and realities of artificial intelligence*. Brookings Institution Press.

Woo, S. (2024, January 30). *Microsoft's AI spending reaches record levels*. The Wall Street Journal. https://www.wsj.com/tech/ai/

X-Energy. (2024, October 16). *Amazon invests in X-energy to support advanced small modular nuclear reactors and expand carbon-free power*. https://x-energy.com/media/news-releases/amazon-invests-in-x-energy-to-support-advanced-small-modular-nuclear-reactors-and-expand-carbon-free-power

Yahoo Finance. (December 12, 2025). *IBM CEO Arvind Krishna Breaks Down Why A 100-Gigawatt AGI Push Could Cost $8T, Says That's 'Today's Number.'*

Yahoo Finance. (2025a, August 1). *Big tech's AI investments set to spike to $364 billion in 2025 as bubble fears ease.* https://finance.yahoo.com/news/big-techs-ai-investments-set-to-spike-to-364-billion-in-2025-as-bubble-fears-ease-143203885

Yahoo Finance. (2025b, October 7). *Without data centers, GDP growth was 0.1% in the first half of 2025, Harvard economist says.* https://finance.yahoo.com/news/without-data-centers-gdp-growth-171546326.html

Yahoo Finance. (2026, May 01). *S&P 500Technology Sector; Industries in this Sector*. https://finance.yahoo.com/sectors/technology/