Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

BEYOND THE AI BUBBLE ANALOGY: VALUATION, PRODUCTIVITY, AND INFRASTRUCTURE IN THE GENERATIVE AI INVESTMENT BOOM

Sérgio Matos da Silva
Faculdade Presidente Antônio Carlos, Brazil
Corresponding author: sergiomatos570@gmail.com
Orcid ID: https://orcid.org/0009-0003-3465-9145

Abstract

The investment boom surrounding generative artificial intelligence has revived comparisons with the dot-com bubble. This article argues that the analogy is useful only when technological diffusion, financial valuation, market structure, and physical infrastructure are analyzed separately. A technology may create substantial social value while investors overpay for particular securities, and a profitable incumbent may still misallocate capital during a competitive investment race. Through a selective, mechanism-oriented integrative narrative review and comparative historical analysis, the study evaluates the AI cycle across adoption, task-level productivity, model economics, semiconductor and cloud concentration, capital expenditure, energy demand, and regulatory risk. The evidence supports a dual diagnosis. Generative AI has achieved rapid adoption and measurable productivity gains in bounded tasks and populations. Current causal evidence does not, however, establish a large economy-wide productivity effect. Monetization remains uneven; model convergence is task- and benchmark-dependent; and infrastructure commitments embed demanding assumptions about utilization, asset life, external customer demand, and willingness to pay. AI displays several characteristics associated with an emerging general-purpose technology, but definitive classification remains premature. The dot-com comparison therefore neither proves an imminent collapse nor validates current prices. The more defensible interpretation is technological transformation accompanied by localized speculative excess and capital-allocation risk. The article proposes an author-developed monitoring framework based on revenue quality, contribution margin, inference economics, economic obsolescence, utilization, enterprise retention, grid execution, and complementary organizational investment.

Keywords: artificial intelligence; generative AI; speculative bubbles; dot-com bubble; productivity; data centers; capital expenditure; energy demand; technological diffusion

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

1. Introduction

Since the public release of widely accessible generative AI systems in late 2022, artificial intelligence has shifted from a specialized research field to a major object of corporate strategy, public policy, and capital-market attention. Firms have invested in advanced semiconductors, cloud capacity, data centers, foundation models, enterprise applications, and energy procurement. The speed and scale of this buildout have encouraged a familiar question: is AI the foundation of a durable technological transformation, or the center of a speculative bubble comparable to the internet boom of the late 1990s (Stanford Institute for Human-Centered Artificial Intelligence [Stanford HAI], 2025; International Energy Agency [IEA], 2025)?

The question is often framed too simply. The commercial internet transformed economic life even though internet equities experienced extreme overvaluation and many firms failed. Technological validity and investment validity are not the same proposition. A technology can diffuse broadly while shareholders in particular companies earn disappointing returns; conversely, a period of overinvestment can create infrastructure that later supports productivity and new business formation. Research on technology bubbles therefore distinguishes the social value of innovation from the private value captured by securities purchased at a particular price (Lansing, 2008; Ofek & Richardson, 2003; Shiller, 2015).

The current AI cycle also differs internally. It includes profitable hyperscalers funding infrastructure from operating cash flow, semiconductor suppliers earning current revenue, startups dependent on external finance, enterprise software firms adding AI features, and application companies whose differentiation may be thin. Treating these participants as one market obscures differences in capital intensity, bargaining power, margins, and exposure to commoditization. The relevant unit of analysis is therefore not “AI” as a single asset class, but the layered value chain through which compute, models, data, distribution, and complementary organizational capabilities produce economic value (Bresnahan & Trajtenberg, 1995; Stanford HAI, 2025).

This study asks three questions. First, which structural similarities between the dot-com era and the contemporary AI boom are economically meaningful? Second, what evidence demonstrates that AI adoption is producing value rather than attention alone? Third, which indicators would reveal that infrastructure and valuations are moving beyond plausible demand? The central thesis is that AI displays several characteristics associated with an emerging general-purpose technology (GPT) while creating conditions in which localized overvaluation and capital misallocation are possible. Definitive GPT classification is usually clearer retrospectively, after widespread diffusion, complementary investment, and persistent cross-sector productivity effects. The strongest present evidence supports

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

bounded task-level utility and rapid diffusion; the weakest link remains conversion of large, durable infrastructure commitments into sustained, high-margin external revenue.

The article contributes a disciplined comparison that resists both technological determinism and bubble rhetoric. It separates capability from adoption, adoption from productivity, productivity from revenue, and revenue from investor return. It also expands the analysis beyond software to electricity, grid capacity, water, permitting, and semiconductor supply. These layers matter because an AI service can be technically valuable but uneconomic at prevailing inference costs, or commercially attractive but constrained by power and interconnection delays (IEA, 2025; Noy & Zhang, 2023). Its original contribution is an author-developed evidence ladder that requires each claim to pass sequentially from capability to adoption, productivity, commercialization, capital-market return, and infrastructure deliverability. Failure at one stage does not negate earlier stages, but it prevents inference to the next.

Economic terminology requires separation. Asset-price overvaluation means that a security price exceeds a defensible estimate of fundamentals. A speculative bubble is a broader price dynamic sustained materially by expectations of resale, narrative reinforcement, or extrapolative beliefs rather than realizable cash flows. Infrastructure overinvestment occurs when capacity is built faster or at greater cost than monetizable demand can absorb. Strategic overcapacity may be created deliberately to preserve optionality, resilience, or competitive position even when near-term utilization is low. Capital misallocation describes resources committed to uses whose risk-adjusted economic return is inferior to feasible alternatives. These conditions can overlap, but none alone proves the others.

Table 1 presents this six-layer author-developed analytical framework. It is a conceptual synthesis for controlling inference across capability, adoption, productivity, commercialization, capital markets, and infrastructure; it is not a formally validated economic index, investment rating, or external standard.

2. Methodology

2.1 Review design and analytical purpose

This study employs an integrative narrative literature review combined with comparative historical analysis. The design was selected because the research question spans financial economics, innovation studies, labor economics, computer science, industrial organization, and energy systems, while the evidence base remains heterogeneous in units of analysis, outcome measures, and publication types. The purpose is therefore not to estimate a pooled effect or claim exhaustive systematic coverage. It is to compare causal mechanisms, identify boundary conditions, and assess how confidently evidence

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

from one analytical layer can be transferred to another (Grant & Booth, 2009; Snyder, 2019).

2.2 Information sources and search strategy

The final search was conducted on September 6, 2026. Searches covered Google Scholar; the National Bureau of Economic Research (NBER) repository; Science; ACM Digital Library; Federal Reserve publications; Stanford HAI; the International Energy Agency (IEA); the U.S. Census Bureau; the Federal Trade Commission (FTC); NIST; and publisher platforms, including Science/AAAS, Wiley Online Library, Oxford Academic, Elsevier/ScienceDirect, Princeton University Press, and proceedings platforms used for exact-title retrieval and bibliographic verification. Corporate sources were limited to identifiable earnings disclosures, investor reports, contractual or partnership disclosures, and technical documentation relevant to capital expenditure, capacity, product pricing, or architecture. Core combinations included ("generative AI" OR "foundation model") AND (productivity OR adoption OR investment OR valuation); ("artificial intelligence" AND "data center") AND (capital expenditure OR utilization OR electricity); (dot-com OR telecommunications boom) AND (bubble OR overinvestment OR capacity); and (inference cost OR scaling law) AND (economics OR efficiency). Backward citation tracing located seminal work on general-purpose technologies, productivity lags, speculative innovation, and telecommunications overbuilding.

2.3 Eligibility, exclusion, and source roles

Eligible analytical sources had to contribute directly to at least one study question through empirical measurement, historical evidence, an explicit economic mechanism, or a documented institutional assessment. Peer-reviewed articles and transparently identified working papers were prioritized for causal and theoretical claims. Official statistical, regulatory, competition, and energy reports were retained for data or institutional analysis unavailable in journals. Eligible corporate materials included named earnings releases, annual or investor reports, technical documentation, and disclosed partnership or capacity terms; they were treated as primary evidence of what the issuing organization reported, not as independent proof of market-wide conditions. Sources were excluded when they relied on anonymous market commentary, promotional vendor surveys without transparent methods, unsourced numerical claims, purely rhetorical bubble comparisons, or Wikipedia.

2.4 Selection and corpus composition

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

Selection was performed by the sole author through title and abstract relevance screening, full-text assessment against eligibility rules, and claim-level verification during synthesis. The archived worksheet supports 21 sources in the original corpus and one targeted recent macroeconomic source added during revision, producing a final corpus of 22 sources. A reliable total of every initially displayed record and duplicate across adaptive Google Scholar, publisher, institutional, and repository searches was not preserved; these unavailable counts are not retrospectively invented. Principal exclusions were purely rhetorical comparisons, promotional evidence without transparent methods, duplicate treatment of the same claim, absence of an identifiable evidentiary contribution, and unverifiable numerical assertions. The final corpus is selective and mechanism-oriented rather than exhaustive. Because several publications contribute to more than one domain, categories indicate primary roles rather than mutually exclusive subjects.

2.5 Data extraction and thematic coding

For each retained source, the author recorded publication type, unit of analysis, evidentiary role, principal outcome, time horizon, and limitations relevant to transferability. Evidence was then coded across eight dimensions: financial conditions; investment composition; technological capability; adoption; task and organizational productivity; revenue and cost economics; market structure; and physical infrastructure. The comparative synthesis separated observed outcomes from forecasts and distinguished task-level causal evidence from firm-level, sector-level, and economy-wide inference. This prevented benchmark gains or user counts from being treated automatically as evidence of profitability or justified valuation.

2.6 Evidence appraisal and synthesis rules

Evidence strength was appraised using four questions: whether the design supported causal or descriptive inference; whether measurement was transparent; whether the setting resembled sustained production use; and whether the conclusion was corroborated by an independent source type. Controlled experiments and field studies were rated as relatively strong for the tasks and populations observed but limited for aggregate extrapolation. Official reports were treated as strong for disclosed statistics and scenarios, but forecasts were interpreted conditionally. Historical studies informed mechanisms rather than one-to-one equivalence. A conclusion was considered robust only when its wording remained within the level of inference supported by the underlying evidence.

2.7 Reviewer procedure and reproducibility limits

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

All searching, eligibility decisions, extraction, coding, and interpretation were conducted by one reviewer. Inter-rater reliability statistics such as Cohen's kappa were therefore unavailable and are not presented as applicable. Classifications such as moderate or moderate-to-strong necessarily include interpretive judgment. Auditability is improved through the final search date, named source platforms, selection rules, coding dimensions, evidence-appraisal criteria, and the source-coding matrix in Supplementary Table S1. These measures do not eliminate confirmation bias or inconsistent judgment, which remain explicit limitations.

2.8 Methodological limitations

The review is selective, mechanism-oriented, and explanatory rather than comprehensive or systematic. It is bounded by rapid technological and market change, uneven corporate disclosure, a relatively small cross-disciplinary corpus, and the absence of standardized measures for AI revenue, inference cost, utilization, and organizational productivity. Adoption surveys may conflate experimentation with sustained integration. Private-company economics and partnership terms are frequently unavailable. Historical analogy cannot establish causal identity across periods, and English-language searching may underrepresent non-U.S. evidence. Accordingly, the study offers a structured synthesis and monitoring framework, not a prediction of a security price, market correction date, or investment return.

3. Conceptual Framework: Four Propositions That Must Be Separated

3.1 Technological capability is not economic value

Benchmark improvement establishes that systems can perform more tasks under test conditions, but economic value depends on reliability, latency, cost, integration, liability, and human oversight. A model that produces a high-quality draft can save time, yet errors may require verification that offsets part of the gain. The value of capability is therefore conditional on workflow design and error tolerance. Experimental studies have found meaningful productivity improvements in writing and customer support, but effects vary across workers and tasks and do not establish universal substitution for labor (Brynjolfsson et al., 2023; Noy & Zhang, 2023).

Scaling-law research explains why investors have expected additional compute and data to improve models. Empirical relationships between loss, model size, data, and computation provided a rationale for large training investments. Later work showed that compute-optimal training requires balancing parameters and data, demonstrating that performance does not follow from scale alone. These results support continued

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

improvement while also implying rising engineering complexity and the possibility that marginal gains become expensive (Hoffmann et al., 2022; Kaplan et al., 2020).

3.2 Adoption is not productive integration

Generative AI spread rapidly because access required little fixed investment from users: a web interface could expose sophisticated capability without local hardware or lengthy installation. Nationally representative U.S. survey evidence found substantial workplace use within two years of public diffusion, a speed that compares favorably with earlier information technologies. Rapid trial, however, is easier than redesigning processes, establishing controls, and measuring value. Adoption statistics should therefore distinguish registration, trial use, occasional use, regular use, workflow integration, enterprise renewal, and attributable economic contribution (Bick et al., 2024).

General-purpose technologies often require complementary investment in skills, processes, data, management, and organizational capital before aggregate productivity becomes visible. The historical productivity paradox of information technology showed that computer expenditure can rise before institutions reorganize around it. AI may follow the same pattern: short-run experimentation can coexist with modest measured gains, while later redesign creates larger effects. This lag is a reason for patience, but it is not evidence that every investment will pay (Brynjolfsson, 1993; Brynjolfsson & Hitt, 2000).

The review does not find causal evidence sufficient to establish a large economy-wide productivity effect from generative AI at the current stage. Existing causal studies support bounded conclusions about specified tasks and worker populations. They do not by themselves identify general-equilibrium effects, reorganization costs, occupational substitution, task creation, wage distribution, sector-wide productivity, or national total-factor productivity (Acemoglu, 2024; Brynjolfsson et al., 2023; Noy & Zhang, 2023).

3.3 Revenue is not profit, and profit is not investor return

AI services can generate revenue while destroying value if compute, customer acquisition, model development, and support costs exceed willingness to pay. Gross revenue also does not identify where surplus is captured. Semiconductor suppliers, cloud platforms, model developers, application firms, and customers bargain over the same productivity gains. Competition and falling inference prices may transfer much of the benefit to users rather than providers, producing broad social value without persistent excess profit for every layer (Agrawal et al., 2019; Bresnahan & Trajtenberg, 1995).

Even durable profit does not guarantee an attractive security return when valuation already assumes exceptional growth. Bubble analysis concerns expectations embedded

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

in prices, financing conditions, and capital allocation, not merely whether a product works. During the dot-com era, investors correctly anticipated the importance of the internet yet frequently misjudged timing, competitive survival, and the share of value captured by listed firms. The same distinction is essential when evaluating AI leaders whose market prices may discount years of rapid expansion (Ofek & Richardson, 2003; Shiller, 2015).

3.4 Infrastructure demand is not infrastructure scarcity value

Growing computation requires chips, networking, data centers, and electricity, but high demand does not make every project economically sound. Project value depends on location, interconnection, utilization, financing cost, customer concentration, equipment obsolescence, and contract quality. A facility completed late or without firm power may underperform even if global AI demand expands. Infrastructure analysis must therefore connect forecasts to specific capacity, timelines, and customers rather than infer profitability from aggregate growth (IEA, 2025).

4. The Dot-Com Episode: Useful Mechanisms, Misleading Shortcuts

4.1 Financial conditions and narrative amplification

The late-1990s internet boom combined a transformative technology with optimistic narratives, abundant capital, rapid initial public offerings, and valuation methods that often displaced conventional measures of earnings and cash flow. Investor attention and limits to arbitrage allowed prices to diverge from fundamentals for extended periods. The subsequent decline did not show that the internet lacked value; it showed that competitive entry, financing dependence, and unrealistic growth assumptions can overwhelm technological promise (Ofek & Richardson, 2003; Shiller, 2015).

Innovation narratives are powerful because future markets are difficult to measure. Investors may rationally pay for upside while disagreeing about probabilities, but social reinforcement can produce extrapolation from recent returns. When prices rise, success stories attract capital and talent, which temporarily validate the narrative through hiring, advertising, and infrastructure demand. This feedback can reverse when financing tightens or disappointing results change expectations, even if the underlying technology continues to improve (Lansing, 2008; Shiller, 2015).

4.2 Telecommunications overinvestment and timing mismatch

The telecommunications boom illustrates how correct long-run demand can coexist with poor near-term investment economics. Deregulation, technological change, and optimistic traffic forecasts stimulated network construction and debt. Capacity expanded faster than profitable demand, competition compressed prices, and many carriers were

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

unable to service obligations. The physical network later supported digital growth, but the social usefulness of the asset did not protect the original capital structure or every investor (Couper et al., 2003).

This mechanism is more relevant to AI than a superficial comparison of stock charts, but the analogy is mechanistic rather than deterministic. Data-center and accelerator projects have long lead times while model efficiency and demand can change quickly. Competitive firms may build simultaneously because being short of capacity is strategically costly, creating collective overcapacity even when individual projects appeared defensible. AI infrastructure nevertheless differs from late-1990s long-haul fiber in financing structures, regulatory exposure, modularity, demand responsiveness, and especially equipment obsolescence. The lesson concerns coordination, financing, utilization, and timing; it does not predict an identical outcome (Couper et al., 2003; IEA, 2025).

4.3 What the analogy misses

Contemporary AI investment is led in large part by incumbent firms with cloud platforms, established enterprise relationships, global distribution, and substantial cash generation. This differs from a cycle dominated by newly listed companies dependent on continual external finance. Incumbent balance sheets reduce immediate funding fragility and allow AI infrastructure to serve existing cloud workloads. Yet internal funding does not eliminate overinvestment; it changes the transmission mechanism from bankruptcy risk toward lower returns on capital and pressure on free cash flow (Stanford HAI, 2025).

The digital substrate is also more mature. Broadband, smartphones, cloud services, payment systems, developer tools, and enterprise software enable rapid distribution of AI functionality. Customers do not need to wait for a new consumer network before testing a model. This supports faster demand discovery than in the early commercial internet. At the same time, AI introduces a new physical constraint because large-scale training and inference depend on scarce accelerators, high-density facilities, cooling, and reliable electricity (IEA, 2025).

5. Evidence for a Real Technological Transition

5.1 Rapid diffusion and broad experimentation

Survey evidence indicates that generative AI entered household and workplace use unusually quickly. Bick, Blandin, and Deming reported that by late 2024 a substantial share of U.S. adults had used generative AI and that workplace use was already meaningful across occupations. The pattern matters because it demonstrates demand beyond investors and developers. It also reveals inequality in use by education, occupation, and age, suggesting that diffusion is broad but not uniform (Bick et al., 2024).

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

Organizational surveys similarly report rising AI use, although measurement requires caution because respondents may classify pilots, embedded features, and production deployments together. Stanford HAI documented a marked increase in reported organizational use and generative AI investment in 2024. These data support a transition from curiosity to experimentation, but the appropriate economic question is whether use persists after subsidies, pilot budgets, and novelty decline (Stanford HAI, 2025).

5.2 Causal evidence on task-level productivity

A field study of customer-support agents found that access to a generative AI assistant increased issues resolved per hour on average, with larger benefits for less experienced and lower-skilled workers. The system appeared to diffuse practices associated with stronger performers, while effects for the most experienced workers were smaller. This finding is important because it identifies a mechanism through which AI can compress knowledge gaps rather than merely automate a complete occupation (Brynjolfsson et al., 2023).

In a preregistered experiment involving professional writing tasks, Noy and Zhang found that generative AI reduced completion time and improved assessed output quality. Benefits were not identical across participants, and the controlled tasks did not include every organizational constraint. Nevertheless, the study provides causal evidence that language models can create value in bounded knowledge work, contradicting claims that the boom rests entirely on speculative storytelling (Noy & Zhang, 2023).

Research on software development has also reported productivity gains under defined conditions, but results depend on task difficulty, developer experience, tool design, and measurement. Faster code generation may be partly offset by review, security, maintenance, or incorrect output. The general conclusion is therefore conditional: AI can accelerate components of knowledge work, particularly when outputs are verifiable, but implementation must account for quality assurance and downstream costs (Peng et al., 2023; National Institute of Standards and Technology [NIST], 2024).

5.3 Falling costs and expanding feasible uses

The economic frontier depends on the cost of producing a reliable unit of output, not only the cost of training a frontier model. Stanford HAI documented substantial declines in the inference cost associated with systems reaching comparable benchmark performance. Lower cost can broaden adoption by making high-volume or lower-value tasks feasible, increasing demand through a rebound mechanism. At the same time, falling prices can weaken provider margins and reduce the scarcity value of model access (Stanford HAI, 2025).

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

Efficiency is therefore ambiguous for investors. It can expand the total market while reducing revenue per query and the amount of hardware required for a fixed workload. If elasticity is high, cheaper inference generates enough new usage to increase aggregate compute; if elasticity is low, efficiency can leave capacity underutilized. The direction cannot be inferred from the existence of more efficient models alone. It must be measured through usage, price, workload mix, and utilization (Alcott, 2005; IEA, 2025).

5.4 Complementary organizational capital

AI becomes economically consequential when organizations redesign work around it. Data quality, process standardization, evaluation, security, employee training, accountability, and integration with existing systems determine whether a model produces repeatable value. These complements are costly and slow, which helps explain why aggregate effects may lag technical capability. They can also create defensible advantage for adopters whose proprietary workflows and data are difficult to copy (Brynjolfsson & Hitt, 2000; NIST, 2024).

This perspective changes the investment thesis. A large share of long-run value may accrue to firms outside the model layer: organizations that use increasingly commoditized AI to improve operations, design products, or lower search and coordination costs. If so, the technology can be transformative while pure model providers face intense competition. The internet similarly created enormous consumer surplus and new industries without ensuring that early entrants captured the majority of value (Bresnahan & Trajtenberg, 1995; Agrawal et al., 2019).

6. Evidence of Speculative Excess and Fragility

6.1 Capital expenditure and the monetization gap

Capital expenditure is not evidence of value by itself. Spending becomes economically defensible when it generates durable incremental revenue, contribution margin, and operating cash flow relative to financing cost and risk. Return on invested capital (ROIC) should be interpreted using explicit attribution and asset-life assumptions. AI hardware can experience economic obsolescence faster than scheduled accounting depreciation because newer accelerators alter performance per watt, price-performance, and customer preference. Accounting ROIC may therefore misstate project economics when depreciation lives diverge from economically competitive replacement cycles. The appropriate capital-conversion framework compares incremental capital expenditure with attributable external revenue, gross profit, contribution margin, operating cash flow, paid utilization, replacement cycles, and economic obsolescence.

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

Attribution is difficult because cloud infrastructure is shared across AI and non-AI workloads, while AI features may support retention or pricing across larger product bundles. This ambiguity can encourage narratives that count all data-center spending as AI commitment and all cloud growth as AI revenue. A more conservative approach triangulates disclosed AI revenue, customer commitments, utilization, depreciation, power contracts, and changes in free cash flow. Uncertainty should be represented explicitly rather than filled with optimistic assumptions.

6.2 Competitive convergence and weak application moats

Model performance is converging on some public tasks and benchmarks, but convergence is benchmark- and task-dependent. Material differences remain in latency, context length, multimodality, reliability, deployment mode, safety controls, price, and ecosystem integration. Similar benchmark scores therefore do not establish full commercial substitutability among frontier models. Where genuine convergence occurs, switching becomes easier and competition can compress prices, transferring value toward cloud platforms, applications, or adopters. Model developers may respond through proprietary data, distribution, tool integration, specialized reliability, or lower cost.

Some thin application layers face a related problem. Products that mainly wrap third-party models with limited workflow integration can be imitated as model access becomes broadly available. This does not imply that the entire application layer lacks defensibility. Durable advantages can arise from proprietary workflow data, regulatory certification, customer integration, distribution, switching costs, domain-specific evaluation, trusted support, and network effects. The relevant question is therefore which applications convert model access into complementary assets that competitors cannot reproduce cheaply.

Commoditization can benefit users while damaging projected provider margins. When multiple models achieve adequate quality, procurement shifts toward price, latency, privacy, and integration. Falling inference prices then expand use but challenge valuations based on persistent scarcity. This is a central reason the AI sector should be analyzed layer by layer: competitive pressure at the model layer can coexist with continued demand for chips, cloud services, or domain-specific applications, and the opposite can also occur.

6.3 Concentration and circular demand

A concentrated ecosystem creates another complication: demand can arise through economically distinct arrangements. Independently generated external end-customer revenue should be separated from capacity precommitments, strategic financing, cloud

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

credits, equity-linked demand, and transactions among related or ecosystem counterparties. These arrangements are not artificial merely because participants are strategically linked; they can finance learning, entry, and capacity coordination. The analytical question is whether demand persists after credits or financing support ends, whether an independent customer ultimately pays, and whether disclosed revenue represents recurring external willingness to pay. Analysts should map counterparties and avoid counting the same economic commitment repeatedly (FTC, 2025).

Concentration also changes competitive risk. A model developer may depend on the same cloud partner that distributes its service, supplies compute, and competes through an internal model. Regulators have examined these partnerships because they can affect access to inputs and the direction of competition. The issue is not that interfirm commitments are fictitious, but that dependency can amplify shocks and blur the boundary between external revenue and ecosystem financing (FTC, 2025).

6.4 Reliability, liability, and governance costs

Generative systems can produce incorrect, fabricated, biased, insecure, or privacy-sensitive outputs. These risks do not negate productivity gains, but they increase the cost of deployment in high-stakes settings. Organizations need evaluation, access control, monitoring, human review, incident response, documentation, and vendor governance. The difference between a compelling demonstration and a dependable production system is partly the cost of these controls (NIST, 2024).

Regulation and litigation may alter both costs and market structure. Compliance obligations can favor large firms that possess legal and technical resources, while uncertainty can delay adoption in regulated sectors. Intellectual-property disputes and data-governance restrictions can change training inputs or product design. Forecasts that extrapolate current growth without governance costs therefore risk overstating margins even when demand remains strong (NIST, 2024; Stanford HAI, 2025).

7. The Physical Economy of AI

7.1 Electricity demand and local bottlenecks

The IEA estimated that data centers consumed roughly 415 terawatt-hours of electricity in 2024, about 1.5% of global electricity consumption, and projected that demand could more than double by 2030 in its base case. AI-oriented accelerated servers are expected to account for a substantial share of the increase. Global percentages can understate operational stress because facilities cluster in regions where generation, substations, and transmission may already be constrained (IEA, 2025).

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

A data center is not a purely digital asset. Its viable capacity depends on firm power, grid interconnection, backup systems, cooling, water availability where relevant, land, network access, equipment delivery, and community acceptance. Interconnection and transmission projects often take longer than server procurement, creating a mismatch between computing ambition and energy infrastructure. These constraints can support scarcity rents for well-sited assets while stranding projects that lack credible power pathways (IEA, 2025).

7.2 Efficiency and the rebound problem

Efficiency improvement does not determine total compute or energy use in a single direction. Lower cost per inference can reduce resource use for a fixed workload, but it can also enable new uses, more frequent invocation, larger contexts, and additional users. Whether total demand rises or falls depends on demand elasticity, deployment scale, model intensity, hardware efficiency, price, and the uses induced by lower cost. The article therefore treats Jevons-type rebound as an empirical possibility rather than an automatic law (Alcott, 2005; IEA, 2025).

The relevant denominator is also contested. Energy per query is easy to communicate but difficult to compare because queries differ in model size, input length, output length, batching, hardware utilization, and data-center efficiency. Life-cycle assessment should include manufacturing and construction where decision-relevant, while operational reporting should distinguish training from inference. Transparent measurement is necessary to prevent both alarmist and promotional estimates from dominating policy (IEA, 2025).

7.3 Environmental distribution and opportunity cost

Even when data centers represent a modest share of global electricity, they can compete locally with housing, industry, electrification, and public infrastructure for grid capacity. Costs may be distributed through utility investment, rate design, tax incentives, and land use, while benefits accrue to operators, customers, and local tax bases. Policy analysis must therefore consider who bears network upgrades and whether contracts protect other consumers from speculative or underutilized demand (IEA, 2025).

The emissions consequence depends on the marginal electricity supply, timing, and location rather than annual renewable procurement claims alone. The IEA expects renewables to supply a large portion of additional demand, but natural gas and other sources may also respond in constrained regions. Clean-energy contracting, storage, flexible computing, and transmission investment can improve outcomes, yet none removes the need for credible accounting and system-level planning (IEA, 2025).

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

8. A Layered Assessment of the AI Value Chain

Table 3 summarizes the permissible inference at each value-chain proposition and prevents capability, adoption, productivity, revenue, and infrastructure demand from being treated as equivalent evidence.

8.1 Semiconductors and manufacturing

Advanced accelerators benefit from technical barriers, software ecosystems, and rapidly expanding demand, but the segment is exposed to cyclicality, customer concentration, export controls, and architectural substitution. Orders can reflect genuine deployment, precautionary capacity reservation, or inventory accumulation. Long-term value depends on sustained workload growth and the ability to preserve margins as customers develop custom chips and competitors improve (Stanford HAI, 2025).

Fabrication capacity has different economics from chip design. Leading-edge manufacturing requires enormous fixed investment and long planning horizons, making utilization central. Geographic concentration creates strategic policy support but also execution and geopolitical risk. Public subsidies can improve resilience while encouraging simultaneous capacity additions that later pressure returns. Industrial policy objectives and private investment returns should therefore be evaluated separately.

8.2 Cloud and data-center platforms

Hyperscale platforms possess customer relationships, developer ecosystems, and the ability to spread fixed costs across workloads. These advantages make them structurally stronger than stand-alone infrastructure projects. They also face the largest absolute capital commitments and potential cannibalization: a customer may shift spending from conventional cloud services to AI rather than create entirely incremental demand. The correct measure is net wallet expansion and margin after infrastructure cost, not AI bookings alone.

Long-term take-or-pay contracts can reduce utilization risk, but they transfer credit and concentration risk to counterparties. Shorter contracts preserve pricing flexibility but expose facilities to rapid technological obsolescence. Project evaluation should examine power rights, customer quality, equipment ownership, residual value, network connectivity, and adaptability to non-AI workloads. A global shortage narrative cannot substitute for asset-level diligence (IEA, 2025).

8.3 Foundation models

Frontier developers may create substantial option value because model improvements enable new products and strengthen distribution partnerships. Their economics are

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

challenging because training, inference, research, safety, and customer acquisition are costly, while competing models can narrow quality gaps. Sustainable advantage may come from product ecosystems, exclusive data rights, enterprise trust, developer adoption, and learning from usage rather than benchmark leadership alone (Agrawal et al., 2019; Stanford HAI, 2025).

Open-weight models complicate scarcity. They can accelerate innovation, localization, and on-premises use while pressuring API prices. “Open” should be specified because weights, training data, code, and usage rights may have different availability. For financial analysis, the important consequence is that accessible alternatives can reduce switching costs and shift value toward deployment services, hardware, data, or applications (Stanford HAI, 2025).

8.4 Applications and adopters

Application firms operate closest to customer workflows and may capture value when they possess distribution, domain data, compliance expertise, and high switching costs. They are vulnerable when incumbents bundle similar features or foundation models commoditize functionality. Metrics should emphasize retention, expansion revenue, gross margin after inference, user frequency, and measurable customer outcomes rather than registrations or generated content volume.

Adopters may ultimately capture the largest share of surplus by using competitive AI inputs to lower cost or improve quality. This possibility weakens a simple investment thesis that equates technological importance with returns to AI-branded suppliers. General-purpose technologies frequently diffuse value across sectors through complementary innovation. Productivity measurement should therefore look beyond the technology industry to firms that reorganize effectively around the tools (Bresnahan & Trajtenberg, 1995; Brynjolfsson & Hitt, 2000).

9. Scenario Analysis

The four scenarios below are conditional pathways for organizing evidence. They are not forecasts of market timing, security prices, or the date or probability of a correction.

9.1 Productive diffusion

In a productive-diffusion scenario, inference costs decline, enterprise retention rises, and complementary investments convert experimentation into sustained workflow gains. Compute demand continues to grow because lower prices unlock new uses, while electricity and grid expansion proceed without severe delays. Competition compresses some model margins, but value is distributed across infrastructure, platforms,

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

applications, and adopters. Market corrections can still occur when individual valuations exceed realistic cash flows (Bick et al., 2024; IEA, 2025).

9.2 Infrastructure digestion

A second conditional scenario is infrastructure digestion. AI use continues to grow, but capacity and capital expenditure advance faster than monetizable external demand. Utilization and prices fall, weaker projects are delayed or written down, and firms absorb economic obsolescence faster than accounting schedules anticipated. The mechanism resembles telecommunications overbuilding, not its exact historical outcome: AI assets have different modularity, depreciation, regulation, and demand responsiveness. Evidence would include soft paid utilization, contract renegotiation, delayed interconnection, lower contribution margins, and capital expenditure persistently outrunning attributable cash flow. This scenario is an analytical pathway, not a forecast of timing or market prices.

9.3 Capability plateau or trust shock

A more adverse scenario combines slower capability improvement with high verification cost, legal restrictions, security incidents, or weak willingness to pay. Enterprises retain limited uses but postpone broad redesign, causing demand forecasts to be revised. The sector would then face both lower terminal revenue and excess infrastructure. Evidence for this scenario would include plateauing retention, poor unit economics, persistent failure in high-value tasks, and declining incremental returns to compute (NIST, 2024).

9.4 Concentrated success

AI may also succeed economically while returns concentrate in a small number of platforms, chip suppliers, or adopters. Network effects, scale economies, and access to capital can reinforce leadership. This outcome validates neither every startup nor every infrastructure project and may intensify regulatory concern. Diversified social benefits can coexist with concentrated private profit, just as broad technological diffusion can coexist with disappointing returns for late investors (FTC, 2025; Ofek & Richardson, 2003).

10. Discussion

The cross-domain synthesis yields an asymmetric conclusion. Evidence maturity is highest for capability and selected task-level productivity, moderate for adoption and infrastructure demand, and lowest for long-horizon profit capture and valuation adequacy. This gradient explains why confident claims about technological usefulness can coexist with cautious claims about financial returns. Table 2 makes those boundaries explicit and

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

prevents a result observed at one level from being promoted into a stronger conclusion at another.

The evidence rejects a binary choice between “bubble” and “revolution.” Generative AI has demonstrated rapid adoption and causal productivity gains in selected tasks. These facts distinguish the cycle from episodes based almost entirely on imagined demand. They do not establish the size, durability, or distribution of future profits. The strongest conclusion concerns usefulness; the weaker and more price-sensitive conclusion concerns whether current investment will earn adequate returns (Brynjolfsson et al., 2023; Noy & Zhang, 2023).

The dot-com analogy is most valuable as a warning about timing and capital structure. Telecommunications investors financed assets whose later social value exceeded the near-term cash flows available to service their obligations. AI infrastructure could experience a similar digestion if synchronized investment creates capacity faster than enterprises reorganize. The risk is magnified by fast equipment cycles and slow power infrastructure, which can make a facility late technologically even when it is delayed physically (Couper et al., 2003; IEA, 2025).

The analogy is least useful when it implies that incumbent financing, mature distribution, and observable demand are irrelevant. Today’s leading technology companies can deploy AI through existing cloud, productivity, advertising, and consumer platforms. They can also reuse some infrastructure across workloads. These advantages lower financing fragility and shorten the path from capability to users. Nevertheless, strategic competition can cause even profitable firms to invest beyond stand-alone expected value because the perceived cost of falling behind is high.

Market concentration produces a paradox. It improves coordination and financing capacity but increases dependency and can make demand appear more diversified than it is. A limited number of firms may finance, supply, host, and purchase services from one another. Analysts should map contractual relationships and distinguish end-customer revenue from ecosystem commitments. Regulators should protect contestability without assuming that scale investment is itself anticompetitive (FTC, 2025).

Energy is a binding consideration but should not be treated as the only bottleneck. Power constraints can delay projects and increase costs, while organizational change, data quality, trust, and evaluation may limit demand even when electricity is available. The phrase “AI infrastructure” should therefore include human and institutional complements as well as chips and data centers. Investment can fail because the grid is slow, because organizations are slow, or because unit economics do not support the intended workload (Brynjolfsson & Hitt, 2000; IEA, 2025).

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

For policymakers, the appropriate response is neither blanket subsidy nor blanket restraint. Grid interconnection rules should discourage speculative capacity reservations, cost allocation should protect other users, and environmental reporting should use comparable operational measures. Competition policy should examine dependency across chips, cloud, models, and distribution. Workforce and education policy should support complementary skills while empirical evaluation distinguishes task assistance from occupation-wide displacement (FTC, 2025; IEA, 2025; NIST, 2024).

For researchers, the priority is measurement that connects technical and economic layers. Needed evidence includes representative longitudinal studies of enterprise use, contribution margins after inference, utilization of accelerated computing, the durability of worker learning, and geographically resolved energy impacts. Public benchmark gains are informative but insufficient when reliability, price, and workflow integration determine realized value.

11. Monitoring Framework and Boundary Conditions

Table 4 translates the monitoring framework into observable expansion-consistent and warning signals. It is used as a diagnostic structure rather than an investment recommendation.

A disciplined monitoring framework begins with revenue quality. Analysts should separate one-time pilots, cloud credits, related-party or partnership demand, committed contracts, and diversified recurring end-customer revenue. Retention and expansion provide stronger evidence than bookings alone. Revenue should then be compared with the full cost of service, including inference, support, depreciation, networking, and energy. This reveals whether growth is becoming economically self-sustaining.

The second category is capital conversion. Incremental capital expenditure should be compared with incremental external revenue, gross profit, contribution margin, and operating cash flow over a realistic lag. Paid utilization, delivery schedules, accounting depreciation, technological obsolescence, asset lives, and replacement rates should be disclosed where possible. A sustained divergence between spending and attributable cash generation would strengthen the overinvestment interpretation even if usage continues to increase.

The third category is technical-economic efficiency. Cost per reliable completed task is more informative than cost per token when tasks require different model sizes and verification. Full inference economics can include accelerators, electricity, cooling, storage, networking, idle capacity, maintenance, technical staff, software licenses, customer support, security, governance, human review, and correction of errors. Monitoring should include latency, reliability, human verification time, contribution margin,

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

and willingness to pay. Efficiency that lowers unit cost can expand adoption while reducing provider value if competition transfers the benefit to customers.

The fourth category is organizational depth. Production integrations, employee training, process redesign, governance controls, and measured outcome improvement indicate durable adoption. High registration with low frequent use, pilot cancellation, or weak renewal would suggest novelty-driven demand. Because general-purpose technologies require complements, the pace of organizational investment helps distinguish delayed value from absent value (Brynjolfsson & Hitt, 2000).

The fifth category is infrastructure execution. Firm power contracts, interconnection milestones, transmission upgrades, construction progress, customer quality, and local cost allocation should be monitored. Announced megawatts are not equivalent to deliverable capacity. Queue reform and financial deposits can improve signal quality by removing speculative requests, while regional concentration should be reported because global averages conceal local scarcity (IEA, 2025).

The thesis of a real transition accompanied by localized excess would weaken if task-level gains fail to survive in production, enterprise retention stalls, and AI-related revenue remains small while capital expenditure continues rising. It would also weaken if inference prices collapse without an elastic increase in paid usage, leaving model providers and facilities with inadequate margins. Conversely, sustained productivity, diversified revenue, high utilization, and credible energy delivery would weaken the case for a broad bubble even if some securities remain overvalued.

12. Limitations and Future Research

This article does not estimate fundamental values for individual securities and should not be read as investment advice. The 22-source corpus is selective and mechanism-oriented rather than an exhaustive census of financial bubbles, labor productivity, GPT theory, model scaling, cloud economics, semiconductors, regulation, and data-center energy. Several domains are necessarily represented by a small number of decision-relevant exemplars. Public disclosures do not allow precise attribution of shared infrastructure, and private-company economics are frequently unavailable. Forecasts are sensitive to model efficiency, price elasticity, regulation, economic obsolescence, and complementary organizational change. The narrative design permits cross-disciplinary synthesis but creates more researcher discretion than a preregistered systematic review (Snyder, 2019).

Future studies should combine technical telemetry with financial and organizational outcomes. Longitudinal firm-level research could measure whether early users deepen deployment and whether productivity gains persist after novelty and training effects.

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

Infrastructure research should compare announced capacity, energized capacity, utilization, and contract structure. Energy analysis should report marginal generation and local grid effects rather than rely solely on annual global totals (IEA, 2025).

The empirical base remains concentrated in the United States, and conclusions should not be interpreted as globally equivalent. Adoption, electricity prices, grid structures, financing costs, labor costs, data governance, language coverage, cloud availability, regulation, and access to advanced chips differ substantially across jurisdictions. An AI buildout that is economically rational for a U.S. hyperscaler may not be replicable in a smaller market or public institution. Comparative work should therefore examine geographically specific infrastructure and organizational conditions.

13. Conclusion

Generative AI is neither adequately described as a replay of the dot-com bubble nor protected from financial excess by demonstrated usefulness. Evidence shows rapid diffusion and meaningful gains in selected tasks, while falling inference costs and organizational interest broaden feasible use. These observations indicate characteristics associated with an emerging general-purpose technology; they do not yet establish a large economy-wide causal productivity effect, definitive GPT status, durable provider margins, or justified current valuations (Acemoglu, 2024; Bick et al., 2024; Brynjolfsson et al., 2023; Stanford HAI, 2025).

The principal financial vulnerability is a timing and conversion problem. Infrastructure can be socially useful but privately unprofitable when capacity arrives ahead of paid demand, competition lowers prices, equipment becomes obsolete, or financing assumes uninterrupted growth. The telecommunications boom provides a stronger warning than a simple collapse analogy because it shows how lasting infrastructure and investor losses can coexist (Couper et al., 2003).

The most defensible diagnosis is therefore “transformation with uneven speculation.” Incumbent balance sheets, mature distribution, and observable demand reduce some fragilities of the late 1990s, while concentrated partnerships, uncertain unit economics, enormous capital expenditure, and energy constraints create new ones. The decisive evidence will come from revenue quality, contribution margins, utilization, enterprise retention, and the successful integration of AI into organizational production.

The central analytical discipline is to keep five questions separate: Can the system perform the task? Will users adopt it repeatedly? Does adoption improve net productivity? Can providers earn sustainable profit after full costs? Are those profits sufficient relative to the capital and valuation committed? Only when these questions are answered

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

independently can the AI cycle be assessed without confusing technological excitement with economic proof.

Funding

The author received no financial support for the research, authorship, or publication of this article.

Conflicts of Interest

The author declares no conflict of interest.

Ethical Approval

Ethical approval was not required because this study analyzed publicly available literature and institutional documents and involved no human participants, identifiable personal data, animals, or clinical intervention.

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

Data Availability

No new dataset was generated or analyzed. The evidence used in this narrative review is contained in the cited publications and institutional documents. The source-coding matrix is provided as Supplementary Table S1.

Author Contributions

Sérgio Matos da Silva: conceptualization, methodology, investigation, evidence appraisal, writing - original draft, writing - review and editing, and final approval of the manuscript.

Declaration of Generative AI Use

Generative AI was used during revision for language editing, structural consistency checking, and assistance in organizing responses to editorial requests. The author independently verified all economic claims, terminology, source attributions, and references; made the final decisions regarding content; and accepts full responsibility for the manuscript.

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

Tables

Table 1. Author-developed evidence architecture of the review

Evidence layer	Primary question	Appropriate inference	Principal limitation
Capability	Can the system perform a defined task?	Performance at measured quality, cost, and conditions	Does not establish adoption or profit
Adoption	Do people or firms repeatedly use it?	Trial, frequency, retention, or workflow integration	Trial may be confused with durable use
Productivity	Does use improve net output or quality?	Task- or workflow-specific effect	Verification and redesign may offset gains
Commercialization	Will customers pay above full cost?	Revenue quality, contribution margin, and retention	Revenue does not prove sustainable profit
Capital markets	Do expected profits justify capital and price?	ROIC under explicit attribution and asset-life assumptions	Useful technology can still be overpriced
Infrastructure	Can compute and power be delivered economically?	Paid utilization, interconnection, contracts, and asset life	Aggregate demand does not validate every project

Note. Conceptual synthesis developed by the author. The six layers are not a validated economic index, investment rating, or external standard.

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

Table 2. Primary source roles and evidentiary maturity

Domain	Representative evidence	Maturity	Interpretive boundary
Task productivity	Brynjolfsson et al. (2023); Noy and Zhang (2023); Peng et al. (2023)	Moderate to strong	Strongest for bounded tasks; limited economy-wide transfer.
Adoption	Bick et al. (2024); Stanford HAI (2025)	Moderate	Use rates do not necessarily equal production integration.
Historical analogy	Couper et al. (2003); Ofek and Richardson (2003); Lansing (2008)	Moderate to strong	Identifies mechanisms, not identical historical outcomes.
Model economics	Kaplan et al. (2020); Hoffmann et al. (2022)	Moderate	Technical scaling does not establish willingness to pay.
Competition and governance	FTC (2025); NIST (2024)	Moderate	Institutional evidence; market conditions change rapidly.
Energy and infrastructure	IEA (2025)	Moderate	Forecasts remain sensitive to efficiency, location, and policy.

Note. Maturity labels are interpretive judgments from a single-reviewer narrative synthesis. They are not formal grades. Task-level causal evidence has the strongest bounded support; economy-wide transfer remains unestablished.

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

Table 3. Inference limits across the AI value chain

Proposition	What evidence can establish	What it cannot establish
Capability	A model performs a defined task at a measured quality and cost.	Broad adoption, profitability, or justified valuation.
Adoption	Users or firms are trying or repeatedly using AI.	Workflow redesign, retention, or net productivity.
Productivity	Output rises or time falls in a specified setting.	Provider profit or economy-wide gains at current prices.
Revenue	Customers pay for an AI-related product or service.	Positive margins, durable differentiation, or investor return.
Infrastructure demand	Compute and electricity requirements are expanding.	That every project, supplier, or valuation will succeed.

Note. Author synthesis based on the reviewed literature. Scenarios and indicators are conditional analytical tools, not forecasts or investment advice.

Table 4. Monitoring indicators for productive diffusion and excess

Indicator	Expansion-consistent signal	Warning signal
Utilization	High paid utilization with diverse customers.	Capacity additions outpace contracted and realized use.
Inference economics	Cost falls while revenue volume and contribution margin expand.	Price competition outruns usage growth and margins compress.
Enterprise adoption	Renewals, workflow integration, and measured productivity.	Pilots proliferate but production retention remains weak.
Capital conversion	Incremental cash generation tracks depreciation and capex.	Capex rises persistently faster than attributable cash flow.
Energy execution	Firm interconnection, realistic schedules, cost allocation.	Speculative queues, delayed power, and socialized upgrade risk.
Competition	Differentiation through data, distribution, and performance.	Interchangeable models and thin application layers.

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

Note. Author synthesis based on the reviewed literature. Scenarios and indicators are conditional analytical tools, not forecasts or investment advice.

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

Supplementary Table S1. Source-Coding and Evidence-Role Matrix

Source cluster	Primary role	Evidence layer	Design/type	Transfer limitation
Brynjolfsson et al. (2023); Noy & Zhang (2023); Peng et al. (2023)	Task productivity	Productivity	Field/experimental/working paper	Bounded tasks and populations; human verification and general-equilibrium effects remain
Bick et al. (2024); Stanford HAI (2025)	Adoption and diffusion	Adoption	Working paper/institutional report	Registration or use does not equal workflow integration or retention
Acemoglu (2024); Brynjolfsson (1993); Brynjolfsson & Hitt (2000)	Aggregate productivity and complements	Productivity	Working papers and economic synthesis	Aggregate effects depend on task exposure, reorganization, and time
Bresnahan & Trajtenberg (1995); Agrawal et al. (2019)	GPT and value distribution	Capability to commercialization	Economic theory	GPT status and rent capture cannot be inferred from technical usefulness
Ofek & Richardson (2003); Lansing (2008); Shiller (2015)	Bubble mechanisms	Capital markets	Historical/financial analysis	Mechanisms do not establish current overvaluation
Couper et al. (2003)	Telecommunications overbuilding	Infrastructure/capital markets	Historical institutional analysis	AI differs in modularity, depreciation, regulation, and demand response

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

Source cluster	Primary role	Evidence layer	Design/type	Transfer limitation
Kaplan et al. (2020); Hoffmann et al. (2022)	Scaling and compute efficiency	Capability/model economics	Working paper/conference paper	Technical scaling does not establish willingness to pay
IEA (2025)	Energy and data-center infrastructure	Infrastructure	Official report and scenarios	Forecasts depend on geography, efficiency, interconnection, and policy
FTC (2025)	AI partnerships and concentration	Commercialization/market structure	Official staff report	Institutional evidence changes with contracts and market structure
NIST (2024)	Risk and governance	Commercialization/organizational integration	Official framework	Normative framework, not evidence of productivity or profit
Review-method sources	Methodology	All layers	Methodological synthesis	Supports design, not substantive market outcomes

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

References

Agrawal, A., Gans, J. S., & Goldfarb, A. (2019). Economic policy for artificial intelligence. Innovation Policy and the Economy, 19(1), 139-159. https://doi.org/10.1086/699935

Alcott, B. (2005). Jevons' paradox. Ecological Economics, 54(1), 9-21. https://doi.org/10.1016/j.ecolecon.2005.03.020

Bick, A., Blandin, A., & Deming, D. J. (2024). The rapid adoption of generative AI (NBER Working Paper No. 32966). National Bureau of Economic Research. https://doi.org/10.3386/w32966

Bresnahan, T. F., & Trajtenberg, M. (1995). General purpose technologies: Engines of growth? Journal of Econometrics, 65(1), 83-108. https://doi.org/10.1016/0304-4076(94)01598-T

Brynjolfsson, E. (1993). The productivity paradox of information technology. Communications of the ACM, 36(12), 66-77. https://doi.org/10.1145/163298.163309

Brynjolfsson, E., & Hitt, L. M. (2000). Beyond computation: Information technology, organizational transformation and business performance. Journal of Economic Perspectives, 14(4), 23-48. https://doi.org/10.1257/jep.14.4.23

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). Generative AI at work (NBER Working Paper No. 31161). National Bureau of Economic Research. https://doi.org/10.3386/w31161

Couper, E. A., Hejkal, J. P., & Wolman, A. L. (2003). Boom and bust in telecommunications. Federal Reserve Bank of Richmond Economic Quarterly, 89(4), 1-24.

Federal Trade Commission. (2025). FTC staff report on AI partnerships & investments 6(b) study. https://www.ftc.gov/reports/ftc-staff-report-ai-partnerships-investments-6b-study

Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. Health Information and Libraries Journal, 26(2), 91-108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

Hoffmann, J., Borgeaud, S., Mensch, A., et al. (2022). Training compute-optimal large language models. Advances in Neural Information Processing Systems, 35, 30016-30030.

International Energy Agency. (2025). Energy and AI. https://www.iea.org/reports/energy-and-ai

Kaplan, J., McCandlish, S., Henighan, T., et al. (2020). Scaling laws for neural language models. arXiv. https://doi.org/10.48550/arXiv.2001.08361

Lansing, K. J. (2008). Speculative bubbles and overreaction to technological innovation. FRBSF Economic Letter, 2008-18. Federal Reserve Bank of San Francisco.

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 01.05 - October
DOI: 10.66104/gs6h3421
Pages: 1-32

National Institute of Standards and Technology. (2024). Artificial intelligence risk management framework: Generative artificial intelligence profile (NIST AI 600-1). https://doi.org/10.6028/NIST.AI.600-1

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Science, 381(6654), 187-192. https://doi.org/10.1126/science.adh2586

Ofek, E., & Richardson, M. (2003). dot-com mania: The rise and fall of internet stock prices. The Journal of Finance, 58(3), 1113-1137. https://doi.org/10.1111/1540-6261.00560

Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). The impact of AI on developer productivity: Evidence from GitHub Copilot. arXiv. https://doi.org/10.48550/arXiv.2302.06590

Shiller, R. J. (2015). Irrational exuberance (3rd ed.). Princeton University Press.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. Journal of Business Research, 104, 333-339. https://doi.org/10.1016/j.jbusres.2019.07.039

Stanford Institute for Human-Centered Artificial Intelligence. (2025). Artificial intelligence index report 2025. Stanford University. https://hai.stanford.edu/ai-index/2025-ai-index-report

Acemoglu, D. (2024). The simple macroeconomics of AI (NBER Working Paper No. 32487). National Bureau of Economic Research. https://doi.org/10.3386/w32487

CC BY