Closing the AI Hype Gap

Gary Hamel
Visiting Professor, Saïd Business School, Oxford University

Michele Zanini
Director, The Management Lab

July 2026

© G. Hamel and M. Zanini, 2026.

The proclamations have been relentless. Artificial intelligence will eliminate half the workforce within a decade. AGI is arriving by 2027. Productivity is about to be unleashed on a scale not seen since the Industrial Revolution. For the better part of three years, these claims have cascaded out of Silicon Valley boardrooms, splashed across front pages, and shaped policy conversations on Capitol Hill and in Brussels. They've provided the rationale for a capital spending boom that by the end of 2026 will have channeled more than $1 trillion into AI infrastructure.

There's little doubt that AI will have a pervasive impact—likely on the scale of the internet, broadband communications, and smartphones. But when you look closely at the boldest claims—that AI will massively displace workers, that AGI is imminent, and that AI will generate unprecedented productivity gains—the evidence either fails to support the hype or actively contradicts it.

The Jobs Apocalypse That Isn't

No AI narrative has generated more alarm than the prospect of mass unemployment. The images are vivid: white-collar professionals replaced by chatbots, knowledge workers automated out of existence, and a generation of young workers arriving in a labor market that no longer needs them. A 2023 McKinsey Global Institute study estimated AI could automate 60 to 70 percent of the work currently done by human beings,[1] and a 2024 IMF report calculated that 40 percent of jobs globally would face meaningful AI exposure.[2] In May 2025, Dario Amodei, co-founder and CEO of Anthropic, warned that AI could eliminate half of all white-collar jobs and spike unemployment to 10 to 20 percent within the next five years,[3] a prediction echoed a month later by Ford Motor Company CEO, Jim Farley.[4]

Such apocalyptic forecasts, stripped of context, are alarming. But context is everything, and the context that keeps getting missed is the distinction between job exposure and job replacement. In 2016, to cite a typical blunder, the cognitive scientist and Nobel laureate, Geoffrey Hinton, often referred to as the "father of AI," predicted that rapid advances in image recognition would soon make radiologists obsolete. "People," he said, "should stop training radiologists now."[5] Yet a decade later, in 2026, the American College of Radiology described workforce shortages as the profession's biggest challenge.[6] Turns out, spotting anomalies on a scan is only part of a radiologist's job. They must also interpret ambiguous results, plan treatment options, coordinate with other members of the care team, communicate with the patient, and exercise legal responsibility.

<u>What Does "Exposure" Actually Mean?</u>

The foundational academic work on this question comes from Tyna Eloundou, Sam Manning, Pamela Mishkin, and Daniel Rock, whose 2024 paper in *Science* estimated that roughly 80 percent of U.S. workers have at least 10 percent of their tasks exposed to AI, while 19 percent have more than half their tasks exposed.[7] These figures are frequently cited in media coverage. What is less frequently cited is what the authors themselves stressed: that

exposure measures the technical feasibility of task automation, not the inevitability of job destruction.

A lawyer whose legal research can be assisted by AI tools is "exposed" to AI, but that same lawyer may find that AI enhances her productivity, increases her billable hours, and makes her job more rewarding, not less. The exposure figures, in other words, are not unemployment forecasts, and were never designed to be.

What the Unemployment Data Shows

When you shift from potential exposure to actual labor market impact, the picture changes dramatically. Researchers at the Economic Innovation Group conducted a detailed analysis of unemployment rates across workers ranked by their degree of AI exposure, spanning the period from 2022 through early 2025. The headline finding: the effect of AI on employment is essentially invisible. Between 2022 and early 2025, the unemployment rate for the quintile of workers most exposed to AI actually increased less than for the quintile least exposed. The most exposed group saw a 0.3 percentage point increase in unemployment, while the least exposed cohort saw a 0.94 percentage point rise.[8] Whatever AI has done to the labor market thus far, it hasn't hammered the most "exposed" workers.

Goldman Sachs Research, synthesizing the most recent data, estimated that if current AI use cases expanded across the economy, and reduced employment in proportion to the efficiency gains, roughly 2.5 percent of U.S. employees would be at risk of AI-related job losses. That's a far cry from the hyperbolic figures that have dominated public discourse. Goldman's economists further projected that AI-related unemployment would increase the U.S. jobless rate by just half a percentage point during a transition period, and that the displacement would dissipate within two years, mirroring the pattern of other technology rollouts.[9]

The last 30 years have seen waves of automation across manufacturing, logistics, financial services, and retailing, yet technology-related job losses as a share of total employment have trended downward during that period. There is no particular reason to expect AI to break from that historical pattern, at least not in the near term. Even the more bearish projections show net job growth. The World Economic Forum's *Future of Jobs Report 2025*, surveying over 1,000 employers representing more than 14 million workers, projected 92 million roles displaced by 2030 alongside 170 million new ones created, for a net gain of 78 million jobs.[10]

Fewer Entry-Level Jobs: AI or Something Else

There's one cohort whose job prospects have undoubtedly worsened in recent years: young graduates entering the job market. An analysis of labor market data by Stanford University researchers Erik Brynjolfsson, Bharat Chandar and Ruyu Chen revealed that the number of early-career workers (ages 22 to 25) in AI-exposed occupations (software engineering and customer service) declined 16 percent between 2022 and 2025, while employment for experienced workers remained stable. The authors attribute the decline in starter jobs to the impact of AI—ChatGPT was launched in November 2022—since, it is argued, entry-level tasks are more susceptible to AI automation.[11]

Yet a 2026 study led by University of Pittsburgh professor Morgan Frank found that hiring conditions for enry-level AI-exposed jobs had started to weaken in early 2022, months before the release of ChatGPT, and that by June 2023, nearly half of the 16 percent decline had already materialized. Frank and his co-authors believe it's unlikely that companies would have been able to automate hundreds of thousands of jobs in a matter of months. It's more likely, they argue, that the job losses were triggered by aggressive monetary tightening and Big Tech downsizing after a pandemic-era hiring surge.[12] A parallel study by researchers at the Economic Innovation Group came to a similar conclusion.[13]

However you cut the data, it's hard to sustain the argument that AI is, or will be, a large-scale job killer. When the International Center for Law and Economics reviewed the empirical literature in 2026, it concluded that most datasets find little evidence of economy-wide AI-triggered job losses or wage declines. While predictions of mass job displacement may generate clicks, such forecasts aren't backed by the data.[14]

The Long Road to AGI

Of all the AI claims, the most dramatic is the assertion that Artificial General Intelligence—a machine capable of outperforming humans on virtually every intellectual task—is imminent. When Elon Musk was asked in April 2024 about the timeline for AGI, he replied, "I think it's probably next year, within two years."[15] OpenAI CEO Sam Altman wrote in early 2025 that he was "confident" the company already knew how to build AGI as traditionally understood.[16] These declarations have generated urgent calls for regulatory action, and ominous predictions about the future of humankind. But once again, there's a chasm between the claims and the evidence.

What Benchmarks Tell Us (and Don't)

The primary evidence cited for rapid AGI progress consists of AI benchmark scores. Between 2023 and 2026, AI software coding systems went from solving 4.4 percent of problems on SWE-Bench (a software engineering benchmark) to solving 71.7 percent. Models have achieved 90-plus percent scores on the Massive Multitask Language Understanding (MMLU) test which covers 57 academic subjects, have completed bar exam simulations in the top decile, and have solved competition-level math problems.[17]

These are genuine achievements, but there is a growing body of evidence that suggests benchmark performance has become decoupled from generalizable capability, and that the progress in benchmark scores increasingly reflects optimization for those specific tests rather than the emergence of broad intelligence.

A grade school math benchmark, GSM8K, illustrates the problem. When it launched in 2021, GPT-3 scored around 35 percent. By 2024, GPT-40, Claude 3.5, and Gemini 1.5 all exceeded 90 percent. The benchmark has since saturated completely for frontier models. MMLU has similarly saturated above 88 percent for leading models. Andrej Karpathy, co-founder of OpenAI, and one of the most respected researchers in the field, wrote in his year-end 2025 review that he had developed "general apathy and loss of trust in benchmarks," observing

that "benchmarks are almost by construction verifiable environments and are therefore immediately susceptible" to target-focused training. "Training on the test set," he argued, has become "a new art form."[18]

Researchers identify a hard benchmark as a measure of progress toward general intelligence, labs optimize for it, models achieve near-perfect scores, a new, harder benchmark is introduced, and the cycle repeats. Humanity's Last Exam (HLE), introduced by researchers at the Center for AI Safety, was designed explicitly as a benchmark sufficiently difficult that it could not be quickly saturated.[19] The test includes 2,500 highly specific questions across a range of academic disciplines. As of April 2026, Anthropic's Claude Mythos had achieved the highest test score, correctly answering 64.7 percent of the questions[20]—an impressive performance, but less than the 90 percent-plus success rate of human experts. This gap will likely narrow over time, but HLE creators have warned that while high accuracy on their benchmark would demonstrate expert-level performance on structured academic questions, it wouldn't demonstrate AI's autonomous research capabilities or AGI.

Architectural Constraints

A growing number of AI researchers believe current LLM architectures face constraints that won't be overcome through scaling alone. A 2025 survey conducted by the Association for the Advancement of Artificial Intelligence (AAAI) found that 76 percent of AI researchers believe that scaling up current AI approaches to achieve AGI is "unlikely" or "very unlikely" to succeed.[21]

The core technical critique centers on what Gary Marcus, emeritus professor at New York University, calls the "world model" problem. LLMs are fundamentally next-token predictors—auto-complete on steroids. They are trained to minimize prediction errors by identifying statistical patterns in text. They do not maintain persistent representations of entities in the world, their properties, or how those entities interact causally over time. Without a proper world model, Marcus argues, you cannot reliably reason about time, causality, spatial relationships, or counterfactual scenarios. In a 2026 paper co-authored with Walter Quattrociocchi and Valerio Capraro, Marcus argued that current AGI claims rest on a conceptual error: "conflating increasingly sophisticated statistical approximations with intelligence itself." By the original criteria for AGI—robustness across environments, reliable generalization under novelty, and autonomous goal-directed behavior—current models are highly limited. Despite impressive gains in competence and fluency, today's LLMs lack persistent goals, struggle with long-horizon reasoning, and depend extensively on human scaffolding for task formulation, evaluation, and correction.[22]

Multi-step reasoning is a particular challenge for LLMs. A 2025 report by a team of Apple researchers found that even the best "reasoning" models failed to solve simple logic puzzles when the task required multiple sequential steps.[23]

Like Marcus, Yann LeCun, a Turing Award winner and former Chief AI Scientist at Meta, is skeptical that LLMs are the route to AGI. "LLMs are not a path to superintelligence or even human-level intelligence," he told the New York Times in January 2026. "The entire industry

has been LLM-pilled."[24] Like Marcus, LeCun argues that true general intelligence requires the kind of physical, embodied understanding of the world that no amount of text training can instill—an ability to model the dynamics of reality, generate novel questions, and produce new knowledge, that current architectures simply do not possess.

Training Hits a Wall

Even if one sets aside architectural challenges, there is a more immediate technical constraint: the scaling laws that powered the rapid progress from GPT-2 through GPT-5 appear to be losing momentum. Thus far, the rapid improvement of foundational models has largely been driven by increases in compute, data, and model size. But leading AI labs have largely exhausted the accessible corpus of high-quality human-generated text. Synthetic data—training LLMs on AI-produced data—can supplement this in some domains, but it introduces the risk of circular improvements and has proven more effective for specific tasks like mathematics and coding than for general reasoning.

As a detailed technical analysis published in late 2024 put it: "We might well have reached a point where pretraining scaling laws are not exactly breaking down, but perhaps slowing down."[25] This was also the conclusion reached by Haosen Ge, Hamsa Bastani, and Osbert Bastani, who argued in a February 2026 paper that gains in LLM performance are now decelerating.[26]

AI labs have responded by shifting investment toward post-training techniques such as reinforcement learning with verifiable rewards (RLVR), and inference-time scaling. These methods have delivered improvements, particularly in structured tasks such as mathematics and competitive coding. They are, by construction, most effective where verifiable objective rewards exist. Unfortunately, such rewards are hard to define in the domains that matter most for AGI, such as open-ended reasoning, novel research, complex social judgment, and physical interaction with an uncertain world.

The Agents Problem

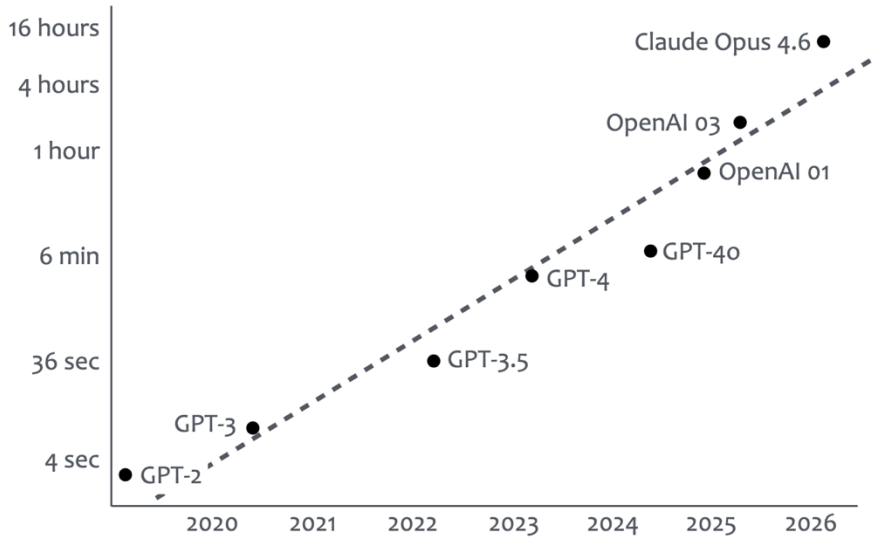
The latest iteration of AGI development centers on AI agents—systems that can complete burdensome tasks with minimal human supervision. Coding agents in particular have shown real progress. The most compelling data comes from Model Evaluation and Threat Research's (METR) Long Tasks benchmark. METR calculates the time it would take a human coder to complete various tasks, and then compares that to the time required for an AI agent to do the same work with a success rate of at least 50 percent. In late 2022, the longest task a coding agent could complete took human beings a minute or two. By May 2026, the time horizon was 16 hours. The METR chart portraying this exponential increase has been widely circulated as evidence that genuine AGI is just around the corner. (See Figure 1).

Yet three measurement choices flatter the chart. First, METR scored its tasks on sixteen "messiness factors" meant to capture how real engineering work differs from basic coding. Unsurprisingly, AI performance is sharply lower on messier tasks.[27] Second, the human task baselines came from a small group of outside engineers paid by the hour. METR's own staff

later completed the same tasks 5 to 18 times faster, suggesting agents look much less capable when measured against experienced engineers who actually know the codebase.[28] Third, the performance bar for the AI agents—50 percent probability of completing the tasks correctly—is far lower than what would be acceptable performance for human coders.

Karpathy, who joined Anthropic in May 2026, calls today's agents "intern entities." Agents, like interns, are prone to making mistakes[29] as Kiro, Amazon's internal AI coding agent, did in December 2025 when it autonomously deleted a critical AWS application.[30] Each agentic step inherits the error rate of the step before, and even a 1 percent error rate per step puts the chance of a hundred-step chain holding together at roughly 37 percent. As LeCun put it, "Mistakes pile up like cars after a collision on a highway."[31]

Figure 1
Time horizon of software tasks LLMs can successfully complete 50 percent of the time.



Source: Model Evaluation and Threat Research

It's also worth noting that METR's benchmarks are focused on software coding, a task for which agentic AI is particularly well-suited. AI's effective time horizon in other domains, METR reports, is between 40 and 100 times shorter.[32] The conclusion: while the growing ability of coding agents to tackle time-intensive tasks is impressive, it's a poor proxy for genuine AGI.

Defining "Intelligence"

In the end, how one judges progress towards AGI depends mostly on one's definition of "intelligence." If, like Elon Musk, you define AGI as a system that is "smarter than the smartest human being,"[33] you may believe AGI is already here. But that's not saying much. Wikipedia is also smarter than the most intelligent human being, as it encompasses a breadth of knowledge no human mind can match.

A better definition of AGI was offered in 2027 by Ben Goertzel and Cassio Pennachin, editors of the first book on artificial general intelligence. They defined AGI as a system that "has the ability to solve a variety of complex problems in a variety of contexts, and to learn to solve new problems they didn't know about at the time of their creation."

This is a demanding, but appropriate test, and one which LLMs are unlikely to pass. Real-world problems, such as reducing manufacturing defects, resuscitating a moribund brand, or improving grade school reading scores, require causal analysis under uncertainty, broad-spectrum contextual knowledge, a deep understanding of stakeholders, creative hypothesis generation, astutely designed interventions, and subsequent adaptation and iteration. While AI is rapidly becoming an indispensable adjunct to this sort of problem solving, there's little chance that in contexts such as those just mentioned it will replace human reasoning, ingenuity, and social cognition. LLMs excel when there's a clearly defined solution space, an objective scoring function, and abundant training data. That brings a large number of tasks within AI's purview, but there's a vast, and arguably more consequential, array of problems that defy the sort of statistical pattern-matching that underpins current frontier models. If AGI implies a capacity to solve the problems that matter most to individuals and institutions, genuine AGI won't arrive this year, next year, or for many years to come.

The Productivity Paradox, Revisited

The third pillar of AI boosterism rests on predictions of a soon-coming productivity bonanza. Consultancy forecasts routinely put the economic value of AI in the trillions of dollars. Goldman Sachs has projected a 7 percent increase in global GDP, worth roughly $7 trillion, over the next ten years,[34] while McKinsey anticipates gains of between $2.6 trillion and $4.4 trillion in annual output from generative AI.[35] These numbers have influenced everything from stock valuations to government investment programs, but they are not supported by the data.

Solow's Paradox

In 1987, Nobel laureate economist Robert Solow made a now-famous observation: "You can see the computer age everywhere but in the productivity statistics."[36] The paradox he identified—that the impact of a transformational technology was failing to show up in aggregate productivity numbers—proved to be a matter of timing. The productivity gains from computing eventually materialized, but not until the late 1990s, a decade or more after wide adoption. And even then, the gains were short-lived and limited to a few sectors such as retailing and wholesale trade.

Economists are now seeing the same phenomenon with AI. "You don't see AI in the employment data, productivity data, or inflation data," wrote Apollo chief economist Torsten Slok in a widely cited article.[37] It seems most executives would concur with that assessment. A 2025 survey asked 6,000 executives (from the U.S., Germany, the U.K., and Australia) to assess the productivity impact of AI on their companies over the preceding three years. Eighty-nine percent of the respondents reported "no impact." When asked to predict AI's pay-

off over the next three years, 12 percent said they expected a large positive impact, 24 percent predicted a small positive impact, and 60 percent anticipated no impact.[38]

The Penn Wharton Budget Model estimated that AI contributed a scant 0.01 percent to the growth of total factor productivity in 2025, an amount so small as to be irrelevant in macroeconomic terms. The productivity gains from AI will likely grow over the next decade, but will be slow to materialize given that most businesses have yet to deploy AI tools at scale.[39]

Acemoglu's Challenge to Trillion-Dollar Claims

The most rigorous challenge to the productivity optimists comes from MIT economist and Nobel laureate Daron Acemoglu, whose 2025 paper "The Simple Macroeconomics of AI" is, to date, the most thorough assessment of AI's likely macro-economic effects.[40]

Acemoglu starts with a study by Svanberg and colleagues that found that only 23 percent of AI-exposed computer-vision tasks are likely to be profitably automated within the next decade.[41] Combining this with estimates of average labor cost savings of 27 percent in automatable tasks, Acemoglu arrived at a projected total factor productivity gain of 0.53 to 0.66 percent over the next decade, or roughly 0.06 percent per year. He notes these modest gains may be overstated because early gains are likely to come from easy, well-defined tasks where AI has an advantage, while future applications will involve complex tasks where AI struggles.

The Meta-Analytic Evidence

Acemoglu's findings are consistent with a growing body of research. A sampling:

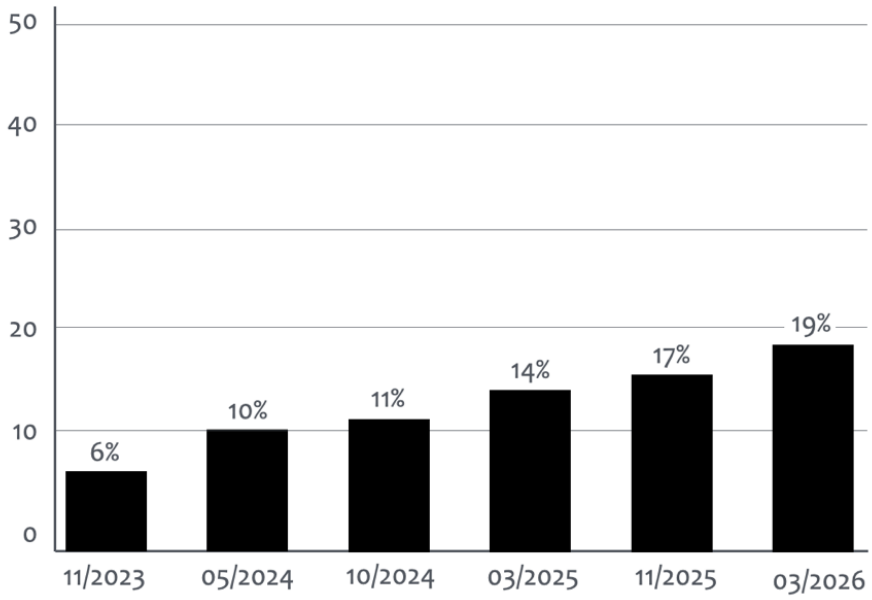
- A review of 37 studies examining the impact of LLMs on software development found that while AI accelerates basic coding tasks, this comes at the cost of quality, and increases the time senior developers must spend reviewing and correcting faulty code.[42]
- In an Upwork survey of 2,500 executives, salaried employees and freelancers, 77 percent of respondents said AI had increased their workload and reduced their productivity.[43]
- A 2025 meta-analysis of 371 studies on AI's impact on labor market outcomes, principally, employment and productivity, found that the mean effect was "close to zero and statistically insignificant."[44]

Given these realities, it's not surprising that companies have been slow to scale up AI. As of March 2026, only 19 percent of firms in UBS's biannual AI survey were using AI "in production at scale." A year earlier, 84 percent of the respondents had expected to reach that benchmark within 12 months.[45] (See Figure 2).

The most frequently mentioned obstacle to large-scale rollout was uncertainty about the pay-off. Tellingly, some companies that pressured employees to use AI (such as Amazon, Microsoft and Meta), have since reversed course, a response to soaring AI costs and meager efficiency gains. Unlike traditional software vendors, such as Salesforce and Workday, which charge enterprise customers a flat, per-user fee, OpenAI and Anthropic charge per token This has led

to widespread sticker shock. Uber, for example, burned through its entire 2026 AI budget in the first four months of the year, while another company reportedly ran up a $500 million AI bill in just 30 days when it failed to put a cap on AI licenses for employees.[46]

Figure 2
Percentage of surveyed firms with AI initiatives "in production at scale"



Source: UBS

The Jagged Frontier

In well-defined task environments, AI *can* produce genuine efficiency gains. For example:

- A 2023 study of 5,172 customer support agents at a Fortune 500 company revealed a 15 percent increase in issues resolved per hour, with gains of 36 percent for the least experienced agents.[47]
- A study of 758 consultants at Boston Consulting Group found that research tasks were completed 18 percent faster with generative AI.[48]
- A 2025 study at Procter & Gamble involving 776 experienced product professionals found that individuals assigned to use AI performed as well as teams of two working without it.[49]

The most comprehensive study of successful enterprise AI deployments, published in April 2026 by Stanford's Digital Economy Lab, examined 51 cases over five months. Where AI worked, it worked well, delivering a median 71 percent productivity gain in agentic implementations. The most impressive results were observed in a regional supermarket chain that had replaced its human procurement function with AI buying agents. The move doubled operating profit—a noteworthy result but for the fact the company's margins had long been half the industry average, reportedly due to haphazard buying decisions. Most of the success stories clustered in narrow domains such as field service, security operations, and customer support.[50] AI was, in other words, helpful, but not game changing.

Overall, the pattern that emerges from AI use cases is what Harvard Business School researcher Fabrizio Dell'Acqua calls a "jagged technological frontier:" AI performs brilliantly on structured tasks with clear success criteria, but dramatically underperforms on complex, ambiguous, multi-step problems requiring judgment and context.[51] There's little doubt that AI is a useful technology, but equally little evidence that it will transform large swathes of economic activity. As the San Francisco Federal Reserve summarized in a February 2026 letter: "Most macro-studies of productivity growth find limited evidence of a significant AI effect. Even firms that say it's useful find little evidence of transformative gains."[52]

In time, those gains may come. It took years for electrification and computerization to produce measurable, aggregate productivity gains. Princeton computer scientists Arvind Narayanan and Sayash Kapoor have argued that AI is subject to the same diffusion constraints that have faced other general-purpose technologies: tacit knowledge that doesn't compress into models, organizational inertia, regulatory uncertainty, and the need for real-world testing.[53] They advise patience, but patience is not what the current investment cycle is pricing in. The pre-market valuations of OpenAI and Anthropic, and the hundreds of billions of dollars being spent by hyperscalers such as Amazon, Microsoft, and Google, are premised on the assumption that AI will deliver unprecedented near-term productivity gains. This seems a textbook case of irrational exuberance.

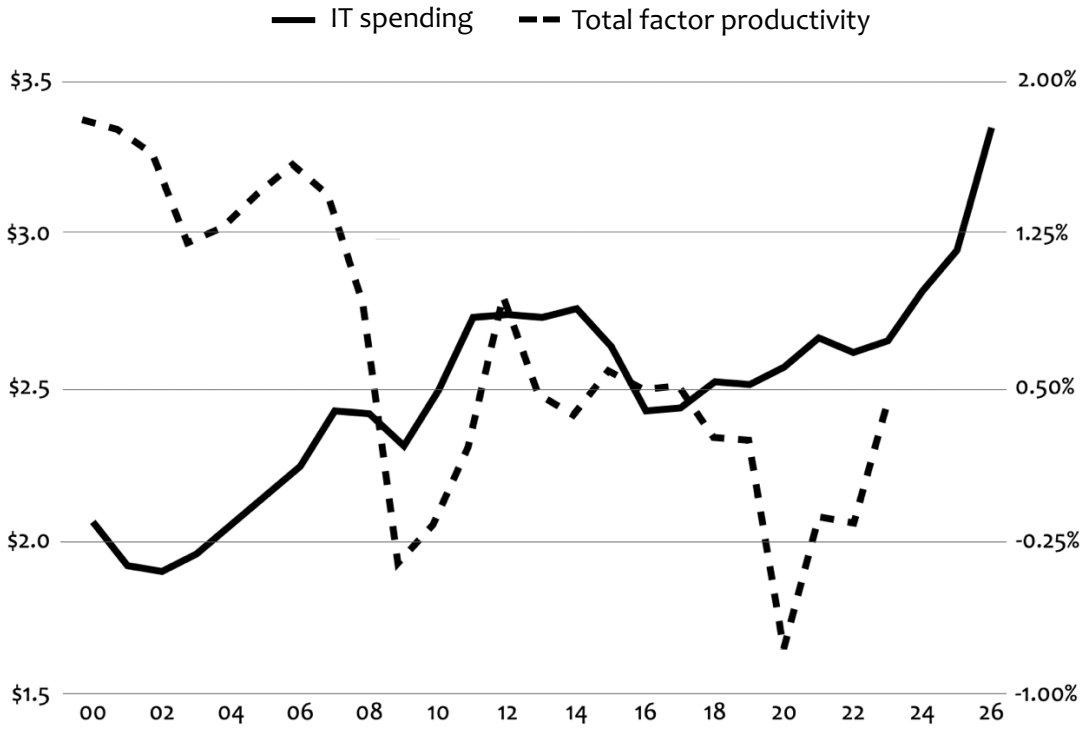
Even before the AI boom, companies and governments were spending roughly $4 trillion dollars a year on information technology. Drawing on data from Gartner and other sources, we estimate that between 2003 and 2023, global IT spending totaled more than $73 trillion in constant 2023 dollars. Yet over this time period, productivity growth in the world's leading industrial economies decelerated. (See Figure 3).

A variety of explanations have been offered for the negative correlation between IT spending and productivity growth, but whatever the underlying causes, the disconnect should temper AI boosterism. To produce a once-in-a-century productivity jackpot, AI would have to be more consequential than the sum of all the digital technologies that have preceded it: semi-conductors, mainframes, CPUs, personal computers, productivity software, ecommerce, GPUs, broadband data, GPS, ERP and CRM systems, cloud computing, the Internet of Things, and smartphones. This seems unlikely.

Conclusion: What Intellectual Honesty Requires

Looking across all three areas—jobs, AGI, and productivity—a consistent pattern emerges. Bold claims made by parties whose financial and reputational interests are served by magnifying AI's potential impact tend to dramatically overstate AI's disruptive impact. On the other hand, when the data is assembled by researchers with no stake in the AI boom, a more cautious and accurate picture emerges.

Figure 3
Growth in global IT spending (billions of 2000 dollars) and the 3-year rolling average growth rate of total factor productivity for the U.S., France, Germany, Italy, Japan and the U.K.



Sources: Gartner, OECD

In the specific domains where LLMs excel, the gains are real and economically valuable (both authors are heavy users of these tools). Software coding, data analysis, visualization, research synthesis, customer service triage, language translation, medical imaging analysis—these are areas where AI is a genuine boon.

But AI has not produced mass unemployment, and the most credible academic forecasts suggest the labor impact over the next decade will be transitional rather than catastrophic. There will be upheavals for workers in specific roles, but not the wholesale disruption of the knowledge economy predicted by AI doomsters. Likewise, AI has not produced AGI. The technical obstacles to achieving general intelligence remain immensely daunting and there's little evidence they will soon be overcome. Finally, AI has not yet produced the macroeconomic productivity revolution upon which trillions of dollars of investment have been premised, and again, seem unlikely to do so.

It's unarguable that AI capabilities are advancing, adoption is spreading, but intellectual honesty requires us to calibrate our expectations to what the data actually show, and resist the temptation to substitute vivid projections for empirical evidence. This is critical for anyone interested in making wise investment decisions, deploying AI at scale, or crafting relevant public policy.

[1] McKinsey Global Institute. "The Economic Potential of Generative AI: The Next Productivity Frontier." (2023). https://www.mckinsey.com/capabilities/tech-and-ai/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier

[2] International Monetary Fund. *World Economic Outlook* (2024). https://www.imf.org/en/Publications/WEO

[3] VandeHei, Jim and Mike Allen. "Behind the Curtain: A White-Collar Bloodbath." Axios. (May 28, 2025). https://www.axios.com/2025/05/28/ai-jobs-white-collar-unemployment-anthropic

[4] Cutter, Chip and Hayley Zimmerman. "CEOs Start Saying the Quiet Part Out Loud: AI Will Wipe Out Jobs." *The Wall Street Journal*. (July 2, 2025). https://www.wsj.com/tech/ai/ai-white-collar-job-loss-b9856259?utm_source=chatgpt.com

[5] Chartrand, Gabriel, et al. "Deep Learning: A Primer for Radiologists." *RadioGraphics*, 37:7. (November-December, 2017). https://pubs.rsna.org/doi/10.1148/rg.2017170077

[6] Rula, Elizabeth Y. "The Radiologist Shortage: A Workforce Update from HPI." *American College of Radiology Bulletin.* (February 5, 2026). https://www.acr.org/Clinical-Resources/Publications-and-Research/ACR-Bulletin/2026/radiologist-shortage-work-force-update?utm_source=chatgpt.com

[7] Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock. "GPTs Are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models." *Science*, 384(6702), pp. 1306–1308 (2024).

[8] Economic Innovation Group. "AI and Jobs: The Final Word (Until the Next One)." (August 2025). https://eig.org/ai-and-jobs-the-final-word/

[9] Goldman Sachs Research. "How Will AI Affect the Global Workforce?" (Updated 2025). https://www.goldmansachs.com/insights/articles/generative-ai-could-raise-global-gdp-by-7-percent

[10] World Economic Forum. "The Future of Jobs Report." (2025).

[11] Brynjolfsson, Erik, Bharat Chandar, and Ruyu Chen, "Canaries in the Coalmine? Six Facts About the Recent Employment Effects of Artificial Intelligence." Stanford Digital Economy Lab. (November 13, 2025). https://digitaleconomy.stanford.edu/publication/canaries-in-the-coal-mine-six-facts-about-the-recent-employment-effects-of-artificial-intelligence/

[12] Morgan R. Frank, et al. "AI Exposed Jobs Deteriorated Before Chat GPT." arXiv.org. (January 4, 2026). https://arxiv.org/abs/2601.02554v1

[13] Iscenko, Zanna and Fabien Curto Millet, "Looking for the Ladder: Is AI Impacting Entry-Level Jobs?" Economic Innovation Group, (January 2026). https://eig.org/wp-content/uploads/2026/01/TAWP-Iscenko-Millet.pdf

[14] International Center for Law and Economics. "AI, Productivity, and Labor Markets: A Review of the Empirical Evidence." (February 2026). https://laweconcenter.org/resources/ai-productivity-and-labor-markets-a-review-of-the-empirical-evidence/

[15] Edwards, Benj. "Elon Musk: AI Will be Smarter Than Any Human Around the End of Next Years." Ars Technica. (April 9, 2024). https://arstechnica.com/information-technology/2024/04/elon-musk-ai-will-be-smarter-than-any-human-around-the-end-of-next-year/?utm_source=chatgpt.com

[16] Altman, Sam. "Reflections." Blog post (January 2025). https://blog.samaltman.com/reflections Altman wrote that OpenAI was "now confident that we know how to build AGI as we have traditionally understood it."

[17] SWE-Bench progress statistics from 4.4 percent (2023) to 71.7 percent (2024), reported across multiple technical analyses. See also: AI Multiple (2026) citing benchmark data from major model providers. Bar exam and MMLU performance figures are drawn from public model evaluation reports by OpenAI, Anthropic, and Google.

[18] Karpathy, Andrej. "2025 LLM Year in Review." (December 2025). https://karpathy.bearblog.dev/2025-llm-year-in-review/
[19] Center for AI Safety, Scale AI and HLE Contributors Consortium. "A Benchmark of Expert-Level Academic Questions to Assess AI Capabilities." *Nature* 649:1139-1146. (2026).
[20] "Claude Mythos Benchmarks Explained." NxCode. (April 8, 2026). https://www.nxcode.io/resources/news/claude-mythos-benchmarks-93-swe-bench-every-record-broken-2026
[21] AAAI (Association for the Advancement of Artificial Intelligence). Survey of AI researchers on scaling and AGI prospects, 2025. Reported in Niall McCarthy, "Why Might the LLM Market Not Achieve AGI." (July 2025).
[22] Marcus, Gary, Walter Quattrociocchi, and Valerio Capraro. "Rumors of AGI's Arrival Have Been Greatly Exaggerated." (February 2026). https://garymarcus.substack.com/p/rumors-of-agis-arrival-have-been
[23] Shojaee, Parshin, et al. "The Illusion of Thinking: The Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity." (May 2025). arXiv:2506.06941. https://arxiv.org/abs/2506.06941
[24] Metz, Cade. "An A.I. Pioneer Warns the Tech 'Herd' Is Marching into a Dead End." *New York Times*, January 26, 2026. https://www.nytimes.com/2026/01/26/technology/ai-limitations-yann-lecun.html
[25]Jonvet, Jon. "A Brief History of LLM Scaling Laws and What to Expect in 2025." Blog post (December 2024). https://jonvet.com/a-brief-history-of-llm-scaling-laws/
[26]Ge, Haosen, Hamsa Bastani, and Osbert Bastani. "Are AI Capabilities Increasing Exponentially? A Competing Hypothesis." (February 4, 2026). arXiv.org.
[27] Witkin, Nathan. "Against the METR Graph." *Transformer*, January 2026, transformernews.ai/p/against-the-metr-graph-coding-capabilities-software-jobs-task-ai.
[28] Kwa, Thomas, et al. "Measuring AI Ability to Complete Long Software Tasks." (2026). arXiv:2503.14499. https://arxiv.org/abs/2503.14499
[29] Karpathy, Andrej. Talk transcript. AI Native Summit (2026).
[30] Murgia, Madhumita, et al. "AWS Hit by 13-Hour Outage Caused by Amazon's Kiro AI Coding Agent." *Financial Times* (February 20, 2026).
[31]LeCun, Yann. Post on X. March 27, 2023. https://x.com/ylecun/status/1640122342570336267.
[32] Kwa, Thomas and Vincent Cheng, "How Does Time Horizon Vary Across Domains," METR Blog https://metr.org/blog/2025-07-14-how-does-time-horizon-vary-across-domains/#summary
[33] Edwards, supra note 15.
[34] Goldman Sachs Research, "How Will AI Affect the Global Workforce?" *op. cit.* (fn. 5).
[35] McKinsey Global Institute, "The Economic Potential of Generative AI," *op. cit.* (fn. 1).
[36] Solow, Robert. "We'd Better Watch Out." *New York Times Book Review* (July 12, 1987).
[37] Slok, Torsten. "AI and Productivity." Apollo Global Management Chief Economist Blog. Reported in: "Thousands of CEOs Just Admitted AI Had No Impact on Employment or Productivity," *Fortune* (February 17, 2026). https://fortune.com/2026/02/17/ceos-ai-no-impact-productivity-employment/
[38] [38] Ivan Yotzov et al., "Firm Data on AI." National Bureau of Economic Research Working Paper 34836 (March 2026), http://www.nber.org/papers/w34836.
[39] Penn Wharton Budget Model. "The Projected Impact of Generative AI on Future Productivity Growth." (September 2025). https://budgetmodel.wharton.upenn.edu/issues/2025/9/generative-ai-productivity
[40] Acemoglu, Daron. "The simple macroeconomics of AI," *Economic Policy*, Vol. 40, No. 121 (January 2025), pp. 13–58. DOI: 10.1093/epolic/eiae042.

[41] Svanberg, Maja S., et al. "Beyond AI Exposure: Which Tasks Are Effective to Automate with Computer Vision?" Social Science Research Network (January 19, 2024). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4700751
[42] Mohamed, Amr, Maram Assi, and Mariam Guizani, "The Impact of LLM-Assistants on Software Developer Productivity: A Systematic Literature Review," (March 23, 2026). https://arxiv.org/abs/2507.03156
[43] Monahan, Kelly, and Gabby Burlacu. "From Burnout to Balance: AI-Enhanced Work Models." Upwork (July 23, 2024). https://www.upwork.com/research/ai-enhanced-work-models
[44] Carbonara, Emanuela, Enrico Santarelli, and Ishita Tripathi. "Assessing the Impact of AI on Labor Market Outcomes: A Meta-Analysis." Social Science Research Network (February 6, 2025). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5126345
[45]Big Go Finance, (May 11, 2026). https://finance.biggo.com/news/OV2hGp4BaoGGrU-ITr7D See also: Palumbo, Angela. "AI-Development Spending Is Huge. Why Few Companies Are Using It." *Barron's* (December 17, 2025). https://www.barrons.com/articles/ai-development-spending-79d225f5
[46] Mills, Madison. "AI Sticker Shock Hits Corporate America." *Axios* (May 28, 2026). https://www.axios.com/2026/05/28/ai-spending-roi-enterprise-costs
[47] Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. "Generative AI at Work," *The Quarterly Journal of Economics,* 140:2 (May 2025), pp. 889–942.
[48] Dell'Acqua, Fabrizio, et al. "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality." *Organization Science*. 37:2 (March–April 2026).
[49] Dell'Acqua, Fabrizio, et al. "The Cybernetic Teammate: A Field Experiment on Generative AI Reshaping Teamwork and Expertise." Harvard Business School Strategy Unit Working Paper No. 25-043 (March 28, 2025).
[50] The Enterprise AI Playbook: Lessons from 51 Successful Developments, Stanford Digital Economy Lab, April 4, 2026. https://digitaleconomy.stanford.edu/publication/enterprise-ai-playbook/
[51] Dell'Acqua, Fabrizio, et al. "Navigating the Jagged Technological Frontier," *op. cit.* (fn. 39).
[52] Federal Reserve Bank of San Francisco. "The AI Moment? Possibilities, Productivity, and Policy." Economic Letter (February 2026).
[53] Narayanan, Arvind, and Sayash Kapoor. "A Guide to Understanding AI as Normal Technology." *AI Snake Oil* (Substack), September 9, 2025. https://www.normaltech.ai/p/a-guide-to-understanding-ai-as-normal