California Management Review

INSIGHT

Artificial Intelligence

Seven Myths about AI and Productivity: What the Evidence Really Says

Dritjon Gruda and Brad Aeon

Image Credit | miss irine

Meta-analytic evidence finds no robust relationship between AI adoption and aggregate productivity gain.

INSIGHT | FRONTIER 16 Oct 2025

Despite widespread enthusiasm for generative AI, empirical evidence reveals inconsistent productivity impacts contradicting popular assumptions. Based on recent meta-analyses and systematic reviews, we debunk seven pervasive myths about AI's workplace benefits. AI's productivity gains are highly context-dependent, varying significantly by user skill level and task complexity. Contrary to expectations, human-AI collaboration often underperforms either agent working independently, except in creative tasks. While AI can accelerate individual work, meta-analytic evidence finds no robust relationship between AI adoption and aggregate productivity gains. We call for research on context-specific organizational deployment strategies to capture genuine value.

RELATED ARTICLES

Jaime Oliver Huidobro, Roberto García-Castro, and J. Mark Munoz. "**AI Automation and Augmentation: A Roadmap for Executives.**" California Management Review Insights, July 17, 2025.

Jack Azagury and Michael Moore. "**Competitive Advantage in the Age of AI**." California Management Review Insights, October 14, 2024.

RELATED TOPICS

Organizational Behavior
Talent Management
Teams & Collaboration
Generative AI

The promise of generative artificial intelligence has captured the imagination of executives worldwide. Despite widespread enthusiasm, however, the actual productivity impact of AI remains murky. This article examines more comprehensive and reliable findings, drawing

exclusively on recent meta-analyses and systematic reviews, to help managers cut through anecdote, hype, and unproven frameworks. This “evidence-about-the-evidence” approach helps neutralize single-study noise and uncover patterns that might otherwise remain hidden due to novelty effects or selective publication biases.

The implicit assumption underlying AI adoption is straightforward: if AI can accelerate individual tasks, organizational productivity will naturally follow. However, as organizations move beyond the initial wave of enthusiasm, empirical evidence indicates that promised productivity gains remain inconsistent across contexts and, in some cases, may impede organizational performance. The findings challenge seven widespread assumptions about AI’s organizational impact and offer concrete guidance for managers navigating this evolving landscape.

The Myth of Universal Productivity Enhancement

Perhaps no belief about generative AI is more pervasive than the assumption that it reliably boosts individual productivity across most contexts and user types. This conviction stems from compelling vendor case studies showcasing dramatic speed improvements and early experiments at customer service desks and developer teams reporting double-digit throughput gains (Brynjolfsson et al., 2023). The narrative appears straightforward: AI automates routine tasks, humans focus on higher-value work, and productivity soars.

However, a July 2025 systematic review of 37 studies examining large-language-model assistants for software development reveals a far more granular reality (Mohamed et al., 2025). While developers did spend less time on boilerplate code generation and API searches, *code-quality regressions and subsequent rework frequently offset the headline gains*, particularly as tasks grew more complex. Senior engineers, in particular, found themselves investing substantial time fact-checking AI output for subtle logic errors that junior developers might have missed entirely.

This pattern extends beyond software development. A 2025 meta-analysis spanning 83 diagnostic-AI studies shows that generative models now match non-expert clinicians yet still trail experts by a statistically significant margin (Takita et al., 2025). Similarly, a randomized controlled trial with 5,000+ agents at a U.S. tech support desk delivered a 35 % throughput lift for bottom-quartile reps but almost no gain for veterans (Brynjolfsson et al., 2024). Together, these findings warn managers not to treat AI as a blanket productivity enhancer (at least not yet) but as a *targeted accelerant* that repays skill-diagnostic deployment strategies.

The managerial response must be equally appropriate. Organizations should instrument their workflows by pairing usage analytics with *quality-of-output* metrics such as bug density or customer-facing error rates. This dual measurement approach reveals not just how much faster work gets done, but whether it meets quality standards. Furthermore, enablement strategies should differentiate between user types: junior and lateral hires may indeed benefit from always-on copilots, while experts derive greater value from fine-tuning capabilities or plugin integrations they can control directly.

The Collaboration Myth: When Human-AI Teams Underperform

The Hollywood imagery of cyborg advantage (combining machine speed with human intuition) has deeply influenced rhetoric around AI deployment. The assumption that human-AI teams inevitably outperform either working alone appears almost self-evident. After all, why wouldn't combining the best of both worlds yield superior results? A comprehensive Nature Human Behaviour meta-analysis covering 106 experiments provides a different answer (Vaccaro et al., 2024). On average, human-AI combinations perform *worse* than the better of the two working solo. Performance improvements emerge only in specific contexts, particularly open-ended content-creation tasks such as brainstorming sessions. Decision-making and judgment tasks, by contrast, suffer from over-reliance on AI suggestions or confusion over authority and responsibility.

This finding challenges key assumptions about AI deployment strategies. Rather than defaulting to hybrid approaches, organizations should map tasks along a “creation versus evaluation” matrix. AI excels at co-creating first-draft marketing copy or generating divergent design ideas, but high-stakes approvals and risk triage should remain with whichever agent (human or algorithmic) demonstrably outperforms in that specific domain. The key insight is that collaboration is not inherently superior; it depends entirely on task characteristics and the relative strengths of each agent.

The Creativity Paradox: Novelty Without Diversity

Generative AI’s creative capabilities have generated significant excitement, fueled by viral poems, artwork, and world-record scores on creativity benchmarks. The assumption that AI has surpassed human creativity in both quantity and quality has become increasingly common in business discussions about content creation and innovation.

A May 2025 meta-analysis of 28 experiments involving 8,214 participants (Holzner et al., 2025) detected no significant creativity gap between generative AI and humans working independently. Humans augmented by generative AI did achieve modestly higher novelty scores, but this came at a substantial cost: dramatic declines in idea diversity. Both humans and models converge on statistically “likely” answers, creating a homogenization effect that undermines the variety essential for robust innovation.

This pattern has important implications for organizational innovation strategies. While AI can effectively seed ideation sessions and accelerate initial concept generation, organizations need to deliberately engineer “divergence rounds” to recover lost variety. This might involve asking teams to generate counterfactuals or analogies without AI assistance, or implementing reward systems that value unique angles rather than speed alone. The goal is to harness AI’s ability to generate novel combinations while preserving the cognitive diversity that drives breakthrough innovations.

The Macro-Economic Reality Check

Consultancy forecasts routinely trumpet multi-trillion-dollar productivity gains from AI adoption (Bradley et al., 2024), and investors already seem to have priced these optimistic projections into stock valuations across multiple sectors (Williams, 2024). The assumption that AI's micro-level benefits automatically translate to macroeconomic gains appears logical and has influenced significant investment decisions.

Yet, a 2025 meta-analysis pooling 371 estimates published between 2019 and 2024 finds no robust, publication-bias-free relationship between AI adoption and aggregate labor-market outcomes once methodological heterogeneity is controlled (Santarelli et al., 2025). Results vary dramatically depending on how studies define "AI," which sectors they sample, and whether they adjust for capital deepening effects.

This disconnect between micro-level gains and macro-level outcomes should inform organizational scenario planning. Rather than building business cases around single-number ROI promises, managers should present ranges and test adoption pilots using both leading indicators (time saved) and lagging indicators (total factor productivity, customer defection rates). The lesson is not that AI lacks value, but that its economic impact unfolds more gradually and unevenly than early enthusiasts predicted.

The Automation Bias Trap

Automation success stories in aviation and manufacturing have created strong associations between automated systems and reduced human error. The assumption that higher levels of automation automatically reduce both cognitive load and mistake rates seems intuitive and has influenced AI deployment strategies across numerous sectors. Yet, a systematic review of 74 studies on automation bias and complacency documents a concerning pattern (Goddard et al., 2012). While not specific to Generative AI, the findings of this review indicate that when decision-support systems are highly but not perfectly

reliable, users become over-trusting, leading to a 12 percent increase in commission errors (e.g., accepting incorrect AI suggestions) and slower detection of rare anomalies. The very reliability that makes AI useful also creates blind spots in human oversight.

Rather than pursuing maximum automation, organizations should deploy adaptive or mixed-initiative systems that can hand control back to humans when model confidence drops below acceptable thresholds. Equally important is investing in “automation literacy” training that helps employees understand when to trust AI outputs and how to verify them effectively. The goal is not to eliminate human judgment but to calibrate it appropriately for different types of AI assistance.

The Stress Paradox

Automating routine tasks will reduce workplace stress and burnout by freeing up mental energy for more meaningful work. The logic behind this common assumption appears sound: if AI handles the mundane tasks, humans can focus on creative and strategic activities that are more engaging and less stressful. Conversely, Lițan (2025) found substantial correlations between technology overload and job insecurity fears with both psychological strain and performance declines. Large language model (LLM) chatbots and auto-reply engines create new sources of stress: constant notifications, unclear responsibility for AI-generated content, and the mental burden of managing AI interactions, rather than eliminating existing pressures.

Organizations must therefore examine not just workload but mental load when deploying AI systems. This requires rotating responsibilities so no employee spends entire days wrestling with AI prompts, enforcing boundaries such as AI-scheduled replies and after-hours cutoffs, and pairing AI rollouts with explicit recovery periods that provide uninterrupted time for focused work. The recognition that AI can create new forms of workplace stress even as it eliminates other stressors should inform both deployment strategies and employee support systems.

A Field Guide for Implementation

The evidence we reviewed in this article suggests that successful AI implementation requires a more sophisticated approach than early adopters anticipated. Organizations should begin with evidence audits that benchmark performance using metrics matching specific tasks: code-review defects for development teams, customer satisfaction scores for service functions, and so forth. This baseline measurement enables an accurate assessment of AI's actual impact rather than relying on subjective impressions or vendor claims.

Task triage emerges as a critical capability. Managers must systematically evaluate which work a) should remain human, b) would benefit from a hybrid human-AI approach, and c) can become fully autonomous. The creation versus evaluation matrix provides a useful framework, but each organization must develop its own mapping based on organizationally specific contexts and capabilities. In addition, adaptive governance structures become essential as AI systems mature. Installing confidence disclosures and explanatory mechanisms helps users understand when to override AI suggestions, while well-being guardrails (e.g., mandatory pauses, no-notification zones, and rotation systems for intensive prompt work) help prevent the new forms of AI-related technostress.

Finally, organizations must maintain an equity lens throughout AI deployment, tracking distributional impacts to ensure that expert disengagement or uneven acceleration among novices doesn't undermine overall organizational capability. Training and incentive structures should evolve to support both AI-augmented novices and experts who choose to work with or without AI assistance.

Implications for Management Practice

The Solow paradox ("You can see the computer age everywhere but in the productivity statistics", referring to the puzzling disconnect between rapid advances in information technology and the sluggish growth of measured productivity in the economy) remains (at least to some extent) relevant in the age of AI. The myths examined here persist partly

because success stories are salient, failure stories are quietly patched, and rigorous evidence aggregations rarely make headlines. Meta-analyses tell a subtler tale: AI's productivity dividend is real in specific contexts, for specific users, and under specific workflow designs. But it is far from automatic or universal.

For management practitioners, the challenge is not to slow innovation but to design for heterogeneity. This means accepting that AI's value curve is U-shaped, that synergies emerge only under appropriate task structures, and that the human in the loop remains both the greatest asset and the weakest link. Approaching AI deployment with the same analytical rigor applied to capital budgeting or safety engineering will help ensure that these myths remain myths rather than becoming costly organizational realities.

AI's transformative potential remains substantial, but realizing it requires more sophisticated management approaches than early enthusiasm suggested. Organizations that move beyond simplistic assumptions about AI's universal benefits and instead develop nuanced, context-specific implementation strategies will be better positioned to capture genuine value while avoiding the pitfalls that have trapped less thoughtful adopters.

References

Bradley, C., Chui, M., Russell, K., Ellingrud, K., Birshan, M., & Chettih, S. "**The Next Big Arenas of Competition**." McKinsey Global Institute. October 24, 2024.

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. "**Generative AI at Work**." No. W31161. National Bureau of Economic Research, 2023.

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. "**Generative AI at Work**." Preprint, arXiv, 2023.

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. "**Generative AI at Work**." arXiv:2304.11771. Preprint, arXiv, November 6, 2024.

Goddard, Kate, Abdul Roudsari, and Jeremy C Wyatt. “A**utomation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators**.” Journal of the American Medical Informatics Association 19, no. 1 (2012): 121–27.

Holzner, Niklas, Sebastian Maier, and Stefan Feuerriegel. “**Generative AI and Creativity: A Systematic Literature Review and Meta-Analysis**.” Preprint, arXiv, 2025.

Liţan, Daniela-Elena. “**Mental Health in the ‘Era’ of Artificial Intelligence: Technostress and the Perceived Impact on Anxiety and Depressive Disorders—an SEM Analysis**.” Frontiers in Psychology 16 (June 2025): 1600013.

Mohamed, Amr, Maram Assi, and Mariam Guizani. “**The Impact of LLM-Assistants on Software Developer Productivity: A Systematic Literature Review**.” Preprint, arXiv, 2025.

Santarelli, Enrico, Emanuela Carbonara, and Ishita Tripathi. “**Assessing the Impact of Ai on Labor Market Outcomes: A Meta-Analysis**.” Preprint, 2025.

Takita, Hirotaka, Daijiro Kabata, Shannon L. Walston, et al. “**A Systematic Review and Meta-Analysis of Diagnostic Performance Comparison between Generative AI and Physicians**.” Npj Digital Medicine 8, no. 1 (2025): 175.

Vaccaro, Michelle, Abdullah Almaatouq, and Thomas Malone. “**When Combinations of Humans and AI Are Useful: A Systematic Review and Meta-Analysis**.” Nature Human Behaviour 8, no. 12 (2024): 2293–303.

Williams, Sean. “**Prediction: The Artificial Intelligence (AI) Bubble Will Burst in 2025. Here’s Why**.” The Motley Fool, December 19, 2024.

Dritjon Gruda Follow

Dritjon (Jon) Gruda holds a joint PhD in Management and Psychology and serves as an Associate Editor and Editorial Board member of multiple journals. His work has appeared in several outlets, including Harvard Business Review and Nature.

Brad Aeon Follow

Brad Aeon holds a Ph.D. in management from Concordia University and serves as a productivity consultant at Highline Productivity. His research applies behavioral science to help high-performing teams optimize time management and enhance focus.