Commentary

The LLM Productivity Cliff: Threshold Productivity and AI-Native Inequality

Francesco Bisardi[1]

1. Independent researcher

We define the LLM productivity cliff as a threshold phenomenon: users who adopt an engineering mindset for working with LLMs attain step-change productivity, while others experience modest gains or outright slowdowns. Synthesizing 2025 evidence across software development, customer support, and labor markets, we show that outcomes are diverse and discontinuous, with performance grouped above and below a capabilities threshold. We operationalize this threshold as architectural literacy, not as marginal prompt skill but as a qualitative shift toward decomposition, orchestration, and systematic validation. We identify boundary conditions that make cliffs more likely (task complexity, scaffolding, mental models) and develop a three-level account of inequality at the individual, organizational, and market levels. We conclude with interventions to reduce capability disparities: embedding scaffolding in tool design, institutionalizing architectural literacy training, and promoting equitable diffusion of architectural literacy and AI-native organizational practices.

Corresponding author: Francesco Bisardi, francescobisardi@gmail.com

1. Introduction

In a randomized study with experienced open-source developers, state-of-the-art AI assistants on average made most participants slower; only a developer with extensive prior assistant use registered clear gains[1]. Large surveys of U.S. programmers report a split: self-identified high-proficiency users report far greater improvements than their peers using identical tools[2]. In customer support, the pattern flips: production AI assistants substantially help juniors, while senior professionals see small or even negative effects[3]. The verified solved rate of SWE-bench rose from 4.4% to 71.7% between 2023 and 2025, as reported by hai.stanford.edu[4]. By October 2025, Claude 4.5 reached a rate of 77.2%. Dario Amodei stated in September 2025 that "70, 80, 90% of the code written in Anthropic is authored by

 doi.org/10.32388/5VAOLK

Claude"[5]. Industry-wide adoption metrics have substantiated this transition: the 2025 DORA Report reveals that 90% of software development professionals presently integrate AI into core operations, with 65% expressing significant reliance and 80% noting enhanced productivity[6].

These contradictions resolve when LLM productivity is modeled as a threshold phenomenon rather than a continuous gradient. Performance divergence stems not from incremental skill differences but from a qualitative shift in work practice: from treating LLMs as conversational tools to redesigning workflows around their computational affordances. Below the threshold, users engage models as extended autocomplete or search. Above the threshold, users adopt an engineering mindset: they decompose tasks, orchestrate tools and agents, structure context engineering, gate risks, and evaluate the outputs as a systematic process. Crossing this cliff generates discontinuous productivity gains: not marginal improvements but order-of-magnitude changes in task completion speed and output quality[1][2][7].

1.1. Problem statement

Why do identical LLMs generate radically different outcomes across users and firms? We argue that productivity variance is explained by architectural literacy: the capacity to decompose complex goals into model-tractable subtasks, orchestrate multi-step workflows, agentic flows, and validate outputs systematically. Productivity gains in complex tasks correlate strongly with architectural literacy; its absence predicts stagnation or decline. Where it is present, productivity becomes discontinuous.

1.2. Scope and positioning

We examine forms of knowledge work in which LLMs plausibly complement human labor at scale, focusing on software development, customer support, and labor-market outcomes. The contribution is not a new metric but a coherent account of how and where cliffs arise and what to do about them.

1.3. Contributions

1. **AI architectural literacy.** We introduce AI architectural literacy as a threshold construct distinct from domain expertise or prompt engineering and demonstrate that crossing this threshold generates discontinuous productivity gains across teams and labor market outcomes.
2. **Evidence synthesis.** We compile 2025 evidence showing bimodal or inverted patterns across domains and experience levels (experts slowed in low-scaffold coding tasks vs. novices uplifted in high-scaffold support systems).

3. **Three-level framework.** We connect individual thresholds, organizational capability building, and market concentration into a single structural account of AI-native inequality.
4. **Agenda to flatten the cliff.** We derive design and policy implications: scaffolding by design, literacy as infrastructure, and equitable diffusion.

2. Conceptual Framework

2.1. The Cliff

The LLM productivity cliff is a discontinuity: small additional effort below the threshold yields little to negative return in complex work; a qualitative shift in how one works with models yields large, stable gains. Unlike continuous learning curves, which predict monotonic returns to effort, cliffs predict threshold effects: minimal gains below a capability threshold and step-change gains above it. This mirrors evidence of emergent abilities in LLMs themselves, where capabilities appear abruptly once models cross scale thresholds[8]. At the user level, persistence without qualitative practice shifts yields diminishing returns. Three patterns emerge across early field studies:

- **Non-monotonicity.** Early adoption can be net negative for experienced professionals when work is complex and scaffolding is weak[1].
- **Bimodality.** Outcomes cluster: a long tail of modest gains and a concentrated tail of large gains among high-proficiency users[2].
- **Inversions.** When systems incorporate robust structure (scaffolding), beginners can surpass seniors who continue to adhere to outdated practices[3].

2.2. Architectural literacy as an engineering mindset

Architectural literacy is not about clever prompts. It is a work discipline that treats LLMs as programmable reasoning and coordination substrates. Five capabilities characterize the threshold:

1. **Decomposition.** Breaking ambiguous goals into model-agent-tractable subtasks and isolating judgment calls that must stay human.
2. **Workflow design.** Iterative agentic chains, checkpoints, and critique loops instead of one-shot prompting.

doi.org/10.32388/5VAOLK

3. **Output evaluation.** Systematic validation tuned to model failure modes rather than ad hoc spot checks.
4. **System integration.** Binding models to data, tools, and agents so the unit of work is a workflow, not a chat turn.
5. **Adaptive mental models.** Updating assumptions about what models can and cannot do and routing work accordingly.

Architectural literacy reorients practice from prompt optimization (what to ask) to workflow and agentic design (how to structure iterative human-model-agent interaction).

2.3. Boundary conditions: When cliffs are most likely to appear

Cliffs are most visible when three conditions align:

- **High task complexity.** Open-ended design, architecture, research, planning, or multi-module coding are cliff-prone; e.g., summarization and translation are not[9].
- **Low scaffolding.** Unstructured prompts or minimally instrumented tools tend to produce high variance in outcomes, whereas structured and constrained systems reduce this variance.
- **Misaligned mental models.** Experts anchored in legacy workflows, UI-based, under- or over-trust models; novices with scaffolded tools can outperform by adhering to systematized guidance.

This framework reconciles divergent findings across domains: open-ended coding tasks (high complexity, low scaffolding) exhibit cliff effects, while structured high scaffolding (customer support) shows novice gains[1][3].

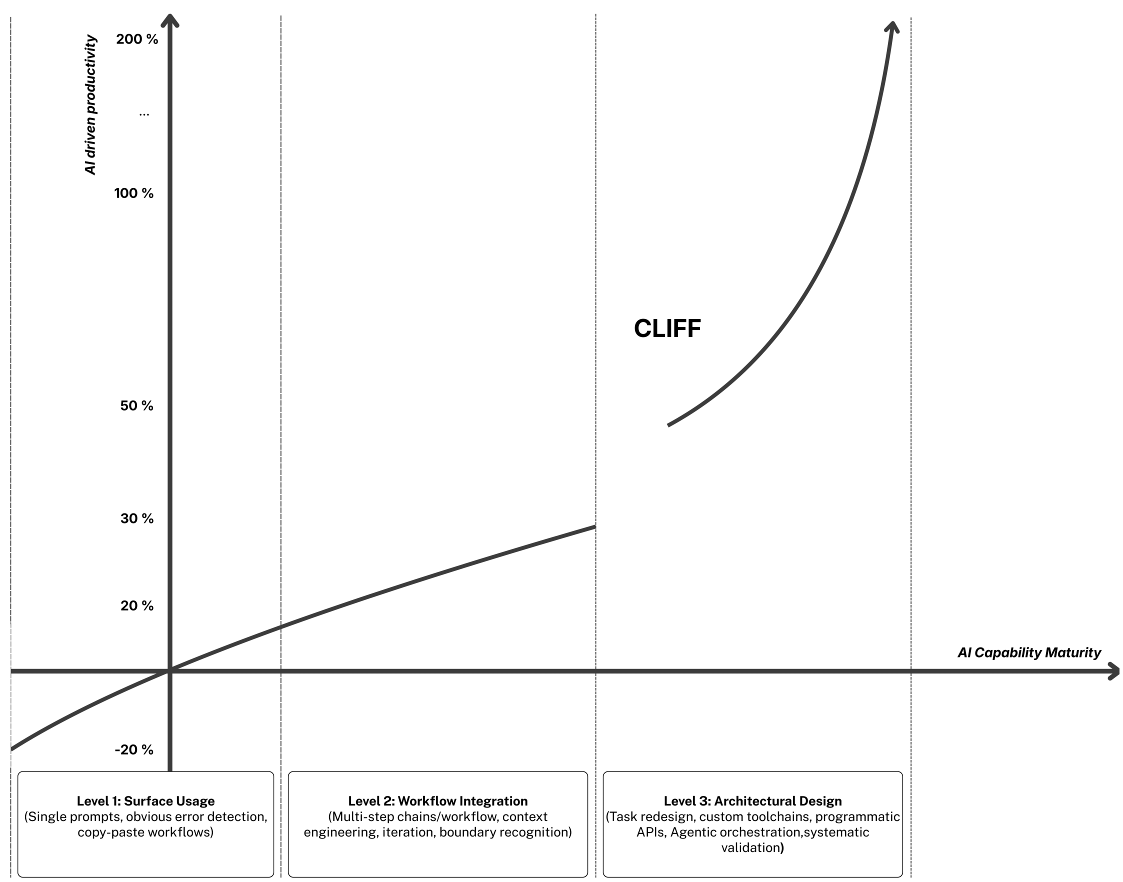
2.4. Levels of practice

To make the threshold/cliff concrete, we use a three-level description of practice. This three-level model operationalizes the cliff[10]: the discontinuity happens in the transition from Level 2 (workflow integration) to Level 3 (architectural redesign)[1][3][11][2][7].

Three-Level Model:

Level / Skills	Capabilities	Productivity Impact	Evidence
Level 1: Surface Usage	Single prompts, obvious error detection, copy-paste workflows	% to +15%	% of devs in [1]
Level 2: Workflow Integration	Multi-step chains/workflow, context engineering, iteration, boundary recognition	+15% to +35%	% “proficient” devs (34% gains)[2]
Level 3: Architectural Design	Task redesign, custom toolchains, programmatic APIs, orchestration, systematic validation	+38% to +200%	＞50h dev [1][38% AI-speedup on my assigned issues. (2025, July 12)], Anthropic (70–90% AI code), [7]

The LLM Productivity Cliff



doi.org/10.32388/5VAOLK

- **Level 1: Surface usage.** One-shot prompts, copy-paste, obvious error catching. Effects range from negative to small positive in complex tasks.
- **Level 2: Workflow integration.** Multi-step prompting, richer context, iterative refinement, basic limits awareness. Gains become consistent but remain bounded.
- **Level 3: Architectural design.** Task redesign, tool/agent orchestration, automated checks, data and API integration, repeatable pipelines creating agentic organization[11]. Gains become step-change.

Movement from Level 2 to Level 3 is the cliff, where the engineering mindset takes place and the distribution "splits."

2.5. Relation to existing theory

Three strands clarify why thresholds, not slopes, dominate:

- **Task-Technology Fit[12].** The same model can be used either as a search-like tool or for structured, orchestrated reasoning, but only the latter suits complex tasks. Fit improves substantially when users redesign tasks to match the tool.
- **Threshold learning & ZPD (Zone of Proximal Development)[13].** Scaffolding moves the boundary of what users can actually do; without it, the self-management and planning demands of LLM use exceed many users' zones of proximal development.
- **Post-routine automation.** When easy tasks are automated, remaining work becomes more expert, raising the bar for meaningful contribution[14][15]. Thresholds, not gradients, follow.

2.6. From individual thresholds to AI-native inequality

Micro-level thresholds may aggregate to macro-level inequality through three mechanisms:

- **Individuals.** Those who cross into Level 3 capture large, persistent gains; others plateau or decline on complex work.
- **Organizations.** Teams that rebuild processes around agents, data, and validation loops pull away from legacy operations; fixed costs and tacit know-how slow catch-up.
- **Markets.** Capability and value concentrate in AI-native startups that can compound workflow advantages; diffusion lags elsewhere.

Productivity inequality at the macro level is structural, not distributional: it reflects cascading effects of architectural literacy gaps across individuals, organizations, and markets[14][15][16]. Our framework

doi.org/10.32388/5VAOLK

suggests that treating LLMs as better autocomplete fails to capture the qualitative shift toward workflow redesign that predicts large productivity gains.

3. Mechanism: A Multi-Level Account of Threshold Effects

We propose a multi-level model in which individual capability thresholds, organizational design choices, and market structure interact to produce distributional bifurcation. Productivity gains are not incremental but discontinuous, contingent on crossing a capability threshold that transforms work organization. We theorize three nested levels that capture this mechanism.

3.1. Individual level: Skill acquisition thresholds

At the individual level, threshold effects may stem from cognitive and procedural factors. Productivity variance is better predicted by architectural literacy than by domain expertise or tool access, conditional on task complexity[1]. Field studies show that below this threshold, complex task performance often worsens: Becker et al.[1] found experienced developers became 19% slower on average when using AI assistants, while the single developer with >50 hours of prior experience achieved significant acceleration. These results are consistent with a threshold model: performance clusters bimodally, with gains concentrated among users exhibiting architectural practices[1][2]. Users who reach this stage move from treating the model as a conversational partner to treating it as a component in a system (designing context, iteration, and validation loops that the model executes within). Ganuthula[17] frames this shift as Capital Structure Transformation: AI compresses the efficient scale of production, allowing individuals to substitute for small teams when they possess sufficient architectural literacy. This pattern diverges from continuous skill acquisition models and aligns with threshold learning theories[10], where conceptual advancements yield nonlinear enhancements in performance.

3.2. Organizational level: Capability development

At the meso level, organizations exhibit the same bifurcation. Firms that are AI-native[18] (designed around integrated AI pipelines, contextual data, and agentic workflows) pull ahead of incumbents still layering AI onto legacy systems. Wang X. et al.[7] provide direct evidence that organizational design, not mere tool access, governs AI spillovers: in lean, AI-native firms, AI adoption generates additional firm-level productivity gains beyond directly affected tasks, with spillovers two to three times larger than those associated with traditional IT firms, even after accounting for firm size. Brodzicki[19] formally

doi.org/10.32388/5VAOLK

models this mechanism, showing that substantial fixed costs of effective AI implementation combined with relatively modest average impacts bias market dynamics toward concentration, allowing AI-native enterprises to accumulate competitive advantages and pushing market structures toward oligopoly. Empirical enterprise studies[20] indicate that although roughly 90% of professionals report using AI, only a small minority achieve substantial productivity gains, largely because organizational infrastructure and team structures have not been redesigned for AI systems. Productivity gaps may reflect differences in system design capability rather than tool access[20][21]. Organizations that redesign processes and integrate AI into core infrastructure (connecting model outputs to databases, APIs, and decision/orchestration agentic systems) may experience superlinear returns, though empirical validation of scaling dynamics remains limited[22].

3.3. Market level: Structural concentration

At the macro level, the cliff manifests as AI-native inequality: value concentrates in actors that can cross capability thresholds at scale. Market bifurcation is already emerging. The WEF *Future of Jobs* [23] report indicates that roughly 60% of enterprises now demand some form of AI literacy, a pattern consistent with increasing concentration of opportunities and returns among AI-capable firms and workers (the cliff). Autor and Thompson[14] demonstrate that the automation of ordinary operations increases the required skill for the remaining tasks. As LLMs automate routine cognitive tasks, residual work requires higher-order skills[14]. Mid-skilled workers may be displaced if they cannot shift to supervisory or orchestration jobs[14][24]; this will concentrate gains among those who complement rather than compete with AI. Fan[15] formalizes a similar mechanism: AI-related productivity gains cluster around certain skills, making it harder for workers to switch roles and increasing wage gaps. Calibrated macro models show a similar pattern: AI may narrow some wage differences but still raise wealth inequality, because high-skilled AI workers and capital owners capture most of the gains from AI adoption[25]. Johnston and Makridis[16] experimentally demonstrate that salary increases disproportionately benefit younger, highly educated, AI-proficient workers in areas where AI serves as a complement rather than a substitute. The cliff hypothesis predicts that individual capability thresholds aggregate to organizational and market-level inequality through three reinforcing mechanisms: skill concentration, capability lock-in, and returns to scale in AI-native workflows.

4. Evidence: Empirical Patterns of Divergence

Preliminary empirical evidence from individuals, organizations, and markets aligns with the cliff hypothesis; however, causal interpretation is limited. Research indicates mixed productivity results, with performance aggregating at both high and low extremes instead of following a normal distribution.

4.1. Individual productivity heterogeneity

A survey of ~2,000 U.S. developers[2] shows a 2.4× productivity differential between high- and low-proficiency users: 34% vs. 14% self-reported improvement. Becker et al.[1] provide experimental evidence: even among seasoned open-source programmers, only those with deep prior assistant experience achieved speedups; the rest slowed down. Brynjolfsson et al.[3] observe an inverted pattern in customer support. Junior agents using an AI generative assistant improved issue resolution by 30%, while senior agents showed negligible or negative gains. The tool embedded strong scaffolding, effectively codifying expert behavior into the system, and in doing so, it upskilled novices by imitation. Large-scale field experiments with nearly 5,000 industry developers similarly find that AI code assistants increase task completion rates by about 26% on average, with substantially larger gains for less-experienced programmers than for senior engineers (Cui et al., 2025). Together, these results suggest a U-shaped pattern: large gains for low-skill workers in scaffolded environments (Cui et al., 2025) and for high-skill workers who achieve architectural fluency (a 21% increase in demand for AI roles, with a 23% wage premium[26]), but stagnation in the middle. Mid-career professionals report lower productivity gains than either novices or experts[2], consistent with the hypothesis that intermediate skill levels face higher adaptation costs. Wang et al.[27] add mechanistic clarity: when tasks are approached through programmatic, multi-agent workflows rather than prompt-response interactions, agents achieved 88% faster completion and 90% lower cost. The performance jump is architectural, not behavioral.

4.2. Organizational capability divergence

Firm-level dynamics mirror the individual cliff. AI-native startups, structured around agentic integration and feedback loops, are scaling faster than legacy enterprises still layering AI on top of older infrastructure. At the startup level, Shi[18] reports that a subset of AI-native firms are achieving, in some cases, $1M revenue/employee, attributing this to intensive AI automation, senior technical talent, high-value contracts with outcome-based pricing, rapid iteration, and highly AI-scalable, automated go-to-

doi.org/10.32388/5VAOLK

market processes run by small AI generalist teams rather than narrow specialists. While the underlying data are not peer-reviewed, they illustrate how aggressive workflow redesign and automation can shift the productivity frontier for small teams. Conversely, a [21] report found that 95% of generative AI pilots at corporations are failing, with the core issue not being model performance but a "learning gap": organizations lack experience integrating AI into workflows, and generic conversational tools such as ChatGPT tend to stall in enterprise settings because they are not integrated with firm-specific workflows and data and therefore cannot adapt to local processes. Field experiments in large enterprises show similar heterogeneity in realized gains: workers given access to Copilot reallocate substantial time away from routine solo tasks, but usage intensity and benefits vary sharply across firms[28]. Brodzicki[19] predicts oligopolistic drift due to high setup costs and limited knowledge diffusion. Using an inter-firm mobility network of AI workers among more than 16,000 U.S. companies, Wang X. et al. [7] showed that productivity spillovers from AI talent are conditional on organizational context: hiring from lean startups yields transferable AI generalists and outsized productivity gains, whereas hires from traditional, multi-layered incumbents transmit little advantage. Doddapaneni et al.[20] empirically confirm that enterprises with legacy architectures capture little of the theoretical productivity gain, even with high AI adoption rates, pushing firms toward strategic partnerships or acquisitions to obtain generative-AI capabilities[21]. In short, access parity masks systemic inequality: the productivity frontier shifts from who has the tools to who can rebuild processes around agentic and AI-native workflows.

4.3. Labor market restructuring

At the labor market level, multiple 2025 studies detect polarization consistent with the cliff hypothesis[3][29][16][24]. Using high-frequency ADP payroll data, Brynjolfsson et al.[3] find that employment in highly AI-exposed jobs fell sharply for entry-level workers (a 6–13% drop among 22–25-year-olds) after ChatGPT's release, while mid- and senior-level workers remained stable or gained. Johnston and Makridis[16] similarly find that wage gains in high-AI-exposure sectors concentrate among educated, AI-fluent employees. Dominski and Lee[24] show that as model capabilities expand, tasks within affected occupations are reclassified from "human" to "AI-capable," raising unemployment and reducing hours for those unable to complement models. In contrast, Humlum and Vestergaard[30] detect minimal average wage change across the Danish labor market, implying that national aggregates hide strong within-sector variance: the gains are captured by a narrow subset of workers and firms. The labor evidence parallels the micro results: only those who can operate at the architectural level, supervising,

doi.org/10.32388/5VAOLK

integrating, and verifying AI outputs, retain or increase value up to a 23% wage[26]. The rest face gradual or sudden displacement.

4.4. Cross-level synthesis

The empirical picture aligns with the threshold theory:

- **Micro level:** Productivity is discontinuous across individuals. The gains scale only after architectural literacy is achieved[1].
- **Meso:** Firms integrating or redesigning AI into core workflows report higher productivity than incumbents, consistent with capability-based concentration[19][20].
- **Macro:** Labor market studies find wage gains concentrated among AI-complementary roles[16][26], consistent with skill-biased technical change[14]. At the same time, cross-country experiments suggest that generative AI can narrow productivity gaps between high- and low-wage regions by ‘leveling up’ workers in lower-income countries[31].

Early evidence across various levels of analysis supports the threshold hypothesis, indicating that productivity gains are concentrated among individuals, firms, and workers who redesign tasks and workflows to capitalize on AI capabilities, while others face stagnation or decline.

5. Methodological Considerations

Until such data exist, the LLM productivity cliff should be treated as a structural hypothesis: it is supported by convergent early evidence but requires validation through sustained and transparent measurement. Studies such as Becker et al.[1] and Brynjolfsson et al.[3] rely on short-term field or workplace trials, typically lasting weeks. They reveal heterogeneity but cannot yet establish persistence or causality. Survey data[2] depend on self-reported proficiency and perceived gains, which may conflate skill with confidence. The current empirical base is therefore temporally narrow and demographically skewed. Early adopters tend to be younger, highly educated, and technically inclined[2][32][3]. Their performance patterns may overstate both potential gains and the magnitude of inequality. Longitudinal observation is needed to determine whether current dispersion stabilizes or widens as organizational scaffolding matures and training diffuses. Measurement practices also vary. “Productivity” spans different denominators (speed, quality, or output volume) and is rarely benchmarked against controlled

doi.org/10.32388/5VAOLK

baselines. Most datasets are observational, not experimental, limiting inference about thresholds versus continuous effects.

Future research should:

1. Track performance and compensation longitudinally across experience cohorts.
2. Instrument variation in scaffolding and workflow redesign to isolate the causal effects of architectural literacy.
3. Develop cross-domain metrics of AI fluency analogous to digital literacy indices, enabling comparison beyond software work.

6. Flattening the Cliff: Design and Agenda

The observed discontinuities are not technologically inevitable. They arise from the speed of the industry and mismatched design, training, and diffusion regimes. Three complementary interventions (design scaffolding, literacy infrastructure, and equitable diffusion) can mitigate these effects.

6.1. Scaffolding by design

Interfaces and development environments can embed progressive scaffolds that translate architectural practices into guided workflows. Evidence from office productivity suites suggests that when such tools are integrated into everyday workflows, they systematically shift how workers allocate time rather than only improving isolated tasks[28]. Tools that visualize context state, support iteration checkpoints, or automatically test outputs reduce metacognitive load and help users cross the threshold without full engineering expertise. Empirical parallels already exist in high-scaffold domains such as customer support, where structured tools collapsed the performance gap between novices and experts[3]. Similar scaffolding could be generalized to creative and analytical work[27].

6.2. Literacy as individual capability infrastructure

Crossing the cliff requires individual architectural literacy, not just prompt fluency. According to the World Economic Forum's Future of Jobs Report[23], 60% of companies would demand AI literacy for competitive advantage[23]. Crossing the cliff appears to require more than short tutorials; it likely depends on the curricular integration of computational and systems thinking. Educational programs and professional certification should frame LLMs not as conversational aids but as programmable systems.

doi.org/10.32388/5VAOLK

Experiments in enterprise training in 2025 show that limited exposure (<50 hours) yields minimal change[1]; depth and repetition are critical. National or sectoral "AI fluency frameworks" could standardize competencies in decomposition, orchestration, and validation, paralleling the digital literacy initiatives of earlier decades.

6.3. Equitable diffusion as organizational and market capability

Productivity gaps persist even when individuals acquire baseline literacy because the larger bottleneck lies in organizational redesign capability. Firms vary dramatically in their capacity to restructure workflows, integrate AI-native processes, and accumulate complementary organizational capital. Diffusion therefore requires investment in institutional supports: firm-level retraining programs, operational redesign, and strategic acquisitions that transfer AI-native practices into legacy environments. Wang's evidence that labor mobility from lean firms is an actual diffusion mechanism[7]. These mechanisms address the structural threshold that determines whether organizations can absorb and scale architectural literacy. Without intentional diffusion, productivity gains will remain concentrated among AI-native firms, amplifying market power and wage dispersion[19][7][14]. Rockall et al.[25] explicitly characterize this as an efficiency–equity trade-off: firms endogenously over-adopt AI in high-wage tasks, which raises aggregate productivity but widens wealth gaps absent policy intervention. Policy interventions should aim to extend organizational capability, not only individual skill.

7. Conclusion

Despite the rapid diffusion of generative AI, productivity outcomes remain heterogeneous and discontinuous across users, organizations, and sectors. Across empirical studies from 2025, the pattern is suggested: identical models produce divergent trajectories because only a fraction of users and organizations adopt the engineering mindset required to harness them. The resulting productivity cliff is both cognitive and structural. Below the architectural literacy threshold, professionals experience limited gains[2]; above the threshold, systematic workflow redesign yields order-of-magnitude improvements[1][27]. Individual-level thresholds may aggregate to organizational and market concentration through capability lock-in and returns to scale, though longitudinal evidence is needed to test this hypothesis. Current evidence should be interpreted as a signal of transition. Cross-country evidence suggests that AI can narrow productivity gaps between countries[31], so who gains globally depends on where and how AI is adopted. As tools mature and scaffolding, literacy, and diffusion mechanisms evolve, the cliff can flatten. Absent interventions, threshold effects may drive a competency-based divide, where productivity and compensation diverge based on architectural literacy rather than tool access[26][15][16]. Zhang[33] similarly documents that AI reshapes power dynamics and inequality within occupations, even when outright job loss is limited. Addressing this divide will require coordination between developers, educators, and policymakers to embed scaffolding, mandate transparency, and institutionalize learning pathways. Mitigating capability disparities requires democratizing not only model access but also the architectural literacy to design, orchestrate, and validate AI-augmented workflows.

References

1. a, b, c, d, e, f, g, h, i, j, k, l, m, n, o *Becker J, Rush N, Barnes E, Rein D (2025). "Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity." arXiv. https://arxiv.org/abs/2507.09089.*
2. a, b, c, d, e, f, g, h, i, j, k *Rethink Priorities (2025). "Adoption of LLMs Among U.S. Tech Workers — Rethink Priorities." Rethink Priorities. https://rethinkpriorities.org/research-area/adoption-llms-tech-workers/.*
3. a, b, c, d, e, f, g, h, i, j *Brynjolfsson E, Li D, Raymond L (2025). "Generative AI at Work." Q J Econ.* ***140****(2):889–942. doi:10.1093/qje/qjae044.*
4. ^ *Stanford HAI (2025). "2025 AI Index Report." Stanford HAI. https://hai.stanford.edu/ai-index/2025-ai-index-report.*
5. ^ *Axios (2025). "Anthropic's Dario Amodei & Jack Clark & Axios' Jim VandeHei" [Video]. YouTube. https://www.youtube.com/watch?v=nvXj4HTiYqA.*
6. ^ *Salva RJ (2025). "How Are Developers Using AI? Inside Our 2025 DORA Report." Google. https://blog.google/technology/developers/dora-report-2025/.*
7. a, b, c, d, e, f, g *Wang X, Feng C, Sun T (2025). "AI Spillover Is Different: Flat and Lean Firms as Engines of AI Diffusion and Productivity Gain." arXiv. https://arxiv.org/abs/2511.02099.*
8. ^ *Havlík V (2025). "Why Are LLMs' Abilities Emergent?" arXiv. https://arxiv.org/abs/2508.04401.*
9. ^ *Gulati K (2025). "Benchmarking the Future of Work: Mapping AI Progress to Occupational Tasks." SSRN Electron J. doi:10.2139/ssrn.5452354.*
10. a, b *Land R (2003). "Threshold Concepts and Troublesome Knowledge: Linkages to Ways of Thinking and Practising Within the Disciplines." Oxford Brookes University. https://www.research.ed.ac.uk/en/publications/threshold-concepts-and-troublesome-knowledge-linkages-to-ways-of-.*

11. [a, b] *Chi Z, Dong L, Dong Q, Hao Y, Wu X, Huang S, Wei F (2025). "The Era of Agentic Organization: Learning to Organize With Language Models." arXiv. https://arxiv.org/abs/2510.26658.*
12. [^] *Goodhue DL, Thompson RL (1995). "Task-Technology Fit and Individual Performance." MIS Q.* **19***(2):213. doi:10.2307/249689.*
13. [^] *Verenikina I (2003). "Understanding Scaffolding and the ZPD in Educational Research." Research Online (University of Wollongong). https://ro.uow.edu.au/edupapers/381.*
14. [a, b, c, d, e, f, g] *Autor D, Thompson N (2025). "Expertise." doi:10.3386/w33941.*
15. [a, b, c, d] *Fan T (2025). "The Labor Market Incidence of New Technologies." SSRN Electron J. doi:10.2139/ssrn.5160195.*
16. [a, b, c, d, e, f] *Johnston A, Makridis C (2025). "The Labor Market Effects of Generative AI: A Difference-in-Differences Analysis of AI Exposure." SSRN Electron J. doi:10.2139/ssrn.5375017.*
17. [^] *Ganuthula VRR (2025). "The Solo Revolution: A Theory of AI-Enabled Individual Entrepreneurship." arXiv. https://arxiv.org/abs/2502.00009.*
18. [a, b] *Shi H (2025). "Official Lean AI Native Companies Leaderboard: The Future of 1-Person Billion-Dollar Startups." leanaileaderboard.com. https://leanaileaderboard.com/.*
19. [a, b, c, d] *Brodzicki T (2025). "Heterogeneous Firms and AI Adoption: Dynamic Insights Into Market Structure and Global Trade." doi:10.13140/RG.2.2.29182.91202.*
20. [a, b, c, d] *Doddapaneni P, Radzevych B, Breeden S, Bansal B, Rao T (2025). "From Pilots to Payoff: Generative AI in Software Development." Bain. https://www.bain.com/insights/from-pilots-to-payoff-generative-ai-in-software-development-technology-report-2025.*
21. [a, b, c] *Nanda M, Challapally A, Pease C, Raskar R, Chari P, Project NANDA (2025). "The GenAI Divide." mlq.ai. https://mlq.ai/media/quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf.*
22. [^] *Gassmann O, Wincent J (2025). "The Non-Human Enterprise: How AI Agents Reshape Organizations." Calif Manage Rev. https://cmr.berkeley.edu/2025/10/the-non-human-enterprise-how-ai-agents-reshape-organizations/.*
23. [a, b, c] *World Economic Forum (2025). "Future of Jobs Report 2025." World Economic Forum. https://www.weforum.org/reports/the-future-of-jobs-report-2025/.*
24. [a, b, c] *Dominski J, Lee YS (2025). "Advancing AI Capabilities and Evolving Labor Outcomes." arXiv. doi:10.48550/arxiv.2507.08244.*
25. [a, b] *Rockall E, Tavares MM, Pizzinelli C (2025). "AI Adoption and Inequality." IMF Work Pap.* **2025***(068):1. doi:10.5089/9798229006828.001.*

26. [a, b, c, d] *Bone M, Ehlinger EG, Stephany F (2025). "Skills or Degree? The Rise of Skill-Based Hiring for AI and Green Jobs." Technol Forecast Soc Change. **214**:124042. doi:10.1016/j.techfore.2025.124042.*

27. [a, b, c] *Wang ZZ, Shao Y, Shaikh O, Fried D, Neubig G, Yang D (2025). "How Do AI Agents Do Human Work? Comparing AI and Human Workflows Across Diverse Occupations." arXiv. doi:10.48550/arxiv.2510.22780.*

28. [a, b] *Dillon EW, Jaffe S, Immorlica N, Stanton CT (2025). "Shifting Work Patterns With Generative AI." arXiv. doi:10.48550/arxiv.2504.11436.*

29. [^] *Chen D, Kane C, Kozlowski A, Kunievsky N, Evans JA (2025). "The (Short-Term) Effects of Large Language Models on Unemployment and Earnings." arXiv. doi:10.48550/arxiv.2509.15510.*

30. [^] *Humlum A, Vestergaard E (2025). "Large Language Models, Small Labor Market Effects." doi:10.3386/w33777.*

31. [a, b] *Bjerkan-Wade B (2025). "Generative AI as a Global Productivity Equalizer: Evidence From Cross-Country Experiments." SSRN Electron J. doi:10.2139/ssrn.5303111.*

32. [^] *Hartley J, Jolevski F, Melo V, Moore B (2025). "The Labor Market Effects of Generative Artificial Intelligence." SSRN Electron J. doi:10.2139/ssrn.5136877.*

33. [^] *Zhang Y (2025). "From Automation Technology to Generative AI: Skill Heterogeneity in Technology's Impact on Laborers." J Chin Sociol. **12**(1). doi:10.1186/s40711-025-00249-9.*

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.