The Missing System: Cognitive Outsourcing and the Policy Architecture of AI in the Workplace

The Economy Research Editorial[1,2]

[1]The Economy Research, 71 Lower Baggot Street, Dublin 2, Co. Dublin, D02 P593, Ireland

[2]Swiss Institute of Artificial Intelligence, Chaltenbodenstrasse 26, 8834 Schindellegi, Schwyz, Switzerland

Abstract

Artificial intelligence has increased productivity within firms that can be measured, most notably in software development, where adoption is now nearing saturation. Governance has not kept pace recent industry surveys show a wide gap between reported productivity gains and the extent to which firms formally review AI assisted work. This article argues that this phenomenon, cognitive outsourcing or delegating judgment not only to information itself but also to AI systems, is not primarily a problem of individual discipline, as much current commentary assumes but a structural one, occurring at every operational level from junior workers through middle managers and top executives and increasing because it often remains hidden in short-term output. Using labor-economics analysis of productivity aggregation, a case study of seniority-biased labor-market entry and firm-level evidence of AI governance practices, the article tracks how the phenomenon was created, why organizations have thus far been unable to establish the governing machinery needed for it and what then follows for policy when the nature of the subtask design problem that created it is laid bare. That design problem is not just offloading unnecessary effort to AI but rather involving AI in substitutable judgment tasks at the level of the firm and school rather than individual worker. The article concludes by arguing that the discerning labor market that emerges will be divided not along old lines of seniority but between AI users, who unquestioningly accept even fluent output and AI governors, who have established the authority to direct, verify and be responsible for it.

1 Introduction - Cognitive Outsourcing Beyond the Classroom

[]

Public anxiety about artificial intelligence and the erosion of independent thought has settled, for the most part, in the classroom. The image is a familiar one: a student pastes an assignment prompt into a chatbot, submits the output with minor edits and moves on having learned little beyond how to phrase a request. Universities have responded with new proctoring software, revised honor codes and extensive institutional debate about the future of critical thinking. This is a legitimate worry and it has already produced a serious body of policy analysis. It is also, on its own, an incomplete picture of where the deeper risk lies. The same behavior that concerns educators is occurring inside firms. It offloads the effortful, judgment-forming part of a task while retaining only its polished output among salaried professionals and the executives who direct them, at a scale the classroom framing does not capture and at a pace workplace policy has barely begun to address.[1]

The scale is not speculative. Anthropic's analysis of one hundred thousand real conversations on its Claude platform estimated that the tasks people bring to the model would, on average, take roughly ninety minutes to complete unassisted and that the model cut completion time by something on the order of eighty percent.[2] Extrapolated across the economy, current-generation systems could add close to two percentage points to annual United States labor productivity growth over the coming decade, nearly doubling the recent trend.[3] Software development supplies the clearest illustration of what this looks like at the operational level: Anthropic's index found that software engineers contribute more to AI-linked productivity gains than any other occupation[4] and a mid-2026 survey of applications and engineering leaders by Info-Tech Research Group found that ninety-four percent of respondents now report meaningful productivity gains from AI, with more than four in five reporting a measurable reduction in shipped defects.[5] On the numbers executives actually track, artificial intelligence in software delivery is not a promise. It is already an operating reality.

The same survey, however, recorded something that complicates the celebratory reading. Among developers actually using AI in the build phase of the software lifecycle, only thirty-seven percent described their organization's governance of that use as formal or better. Two in three agreed that AI-written code demands more testing than code a person wrote unaided and security concerns, output-quality worries and basic skills gaps ranked as the leading barriers to adoption, each cited by roughly a third to half of respondents.[6] Productivity has arrived; the surrounding apparatus of verification, accountability and review has not caught up with it. This is not a pattern confined to engineering departments. Commentators on both sides of the Atlantic have begun describing a broader dynamic under the heading of cognitive outsourcing: the habit, described by the technology writer Enrique Dans as a natural extension of a decades-long trend that began with search engines replacing memory and GPS replacing spatial reasoning, of handing over not just information retrieval but the labor of forming a judgment before expressing one.[7] Eric So, a behavioral economist at the MIT Sloan School of Management, has given the underlying pull a name: AI gravity, the mounting and largely unconscious pressure to route ever more thinking through a system that will do it faster and with less friction, than doing it oneself.[8]

The conventional response to this concern, in both public commentary and the first wave of workplace policy proposals, treats cognitive outsourcing largely as a matter of individual discipline: verify AI output more

carefully, train workers to prompt more responsibly, cultivate a habit of skepticism. That response is not wrong so much as aimed at the wrong institutional level. It assumes the central failure is a lapse of personal judgment that better habits can correct, when the evidence increasingly suggests the failure is structural: organizations at every level, from the individual desk to the executive suite, have adopted a capability without building the operating architecture that would make careful verification the default rather than an act of individual willpower running against every incentive pushing the other way.[9] It also assumes, by focusing so heavily on students, that the sharpest version of the problem belongs to people who are still learning their trade. The evidence reviewed in the following pages points the other way: the compounding costs of unverified reliance appear to be accumulating fastest not among novices, who at least know they do not yet know, but among the working professionals and senior leaders whose judgment an organization is least equipped to double-check.

Three developments between 2023 and 2026 make this reframing newly urgent rather than merely plausible. The International Labour Organization's 2026 review of firm-level and macroeconomic productivity data found that the striking task-level gains documented in individual studies have not yet appeared in aggregate output, a pattern it attributes to the same organizational lag that delayed the payoff from electrification and information technology by a generation.[10] Labor-market research from Stanford's Digital Economy Lab, tracking payroll records for millions of American workers, found a roughly sixteen percent relative decline in employment for workers aged twenty-two to twenty-five in the occupations most exposed to generative AI, even as employment for older workers in the same occupations held steady;[11] a separate study spanning résumé and job-posting data for sixty-five million workers across two hundred and eighty thousand firms found the identical pattern in hiring rather than headcount, a phenomenon researchers now call seniority-biased technological change.[12] And a widely discussed Brookings analysis has argued that the productivity gains most visible today are disproportionately produced by a generation of experts who built their judgment before these tools existed. This creates a form of borrowed expertise whose supply is not being renewed at the rate it is being drawn down.[13] Read together, these findings describe something more specific than generalized unease about AI. They describe a pipeline problem with a measurable shape, unfolding on a timeline policymakers can no longer treat as a future concern.

What follows takes that pipeline problem seriously as an object of policy rather than of personal ethics. It begins by documenting the behavioral signature of cognitive outsourcing across the organizational ladder, from the individual contributor through the executive suite and by showing why the same feature that makes AI so useful is its capacity to absorb the offloadable, well-defined portion of cognitive work. That is also what makes over-reliance so easy to fall into and so difficult to detect from the inside.[14] It then turns to the deeper cause: the absence, inside most firms, of an operating system capable of converting raw AI capability into governed, accountable output. The final section draws out what remains, once that diagnosis is accepted, for public policy specifically, which has a narrower and more defensible role than either banning the technology outright or leaving its consequences to individual discretion.

2 What Are the Signs of AI Reliance and Over-Reliance in the Workplace?

So's account of AI gravity identifies three forces that push a worker toward outsourcing judgment rather than merely information. The first is straightforwardly cognitive: more capable systems amplify an existing human instinct to conserve mental effort, so the temptation to route a task through the model grows precisely as the model improves, not in spite of it. The second is competitive: in an environment where visible performance is what gets rewarded, using AI to mimic the output of a genuine expert becomes a rational individual strategy even when it produces no underlying expertise. The third is social: because it is increasingly difficult to tell from a colleague's output whether they used AI and how much, the pressure to keep pace becomes very hard to resist, since abstaining carries a private cost with no private benefit. None of these three forces requires laziness or bad faith. They describe ordinary incentives operating on ordinary people inside institutions that have not adjusted to compensate for them.[15]

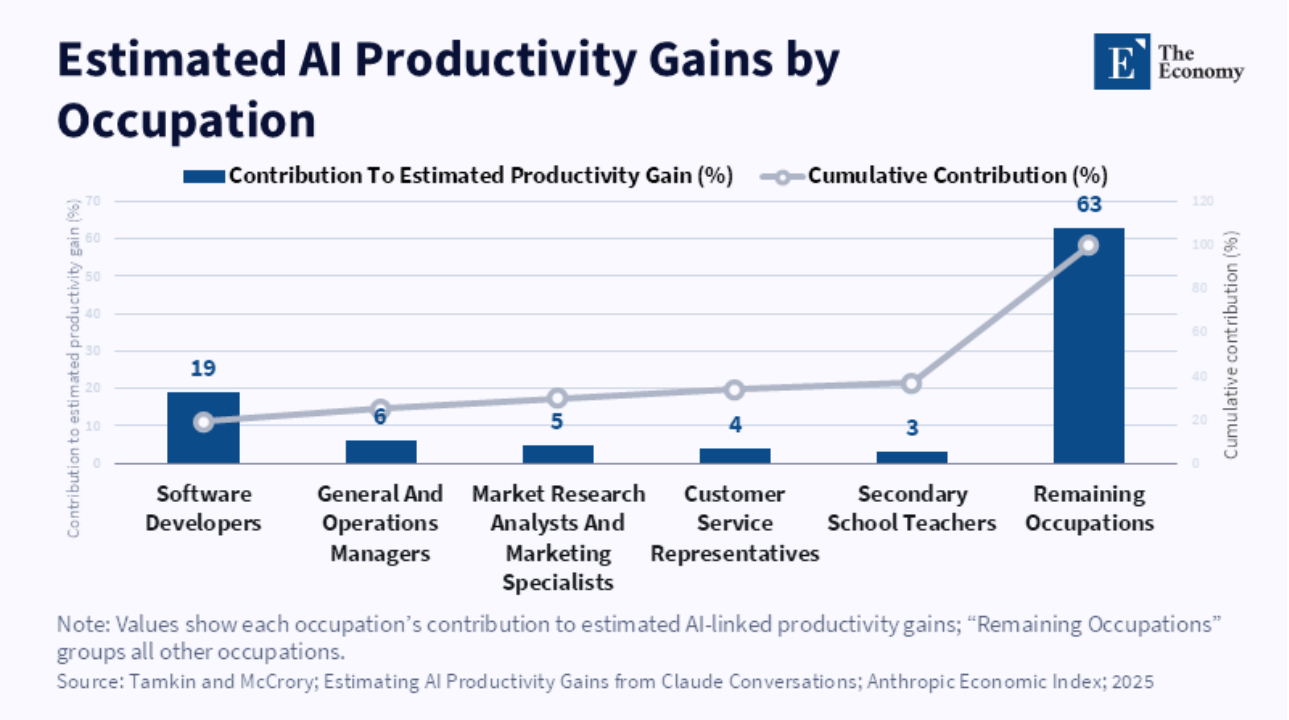

Figure 1: AI-linked productivity gains are led by software work, but the broader effect is distributed across many occupations rather than confined to one sector.

The behavioral evidence is now substantial enough to move beyond anecdote. A preliminary study from the MIT Media Lab found that eighty-three percent of participants who had written an essay with ChatGPT's assistance could not quote a single sentence from the work they had submitted minutes earlier.[16] This summarizes the material as passing from the screen to the assignment without ever entering the writer's mind. A separate survey of knowledge workers conducted by Microsoft Research and Carnegie Mellon University, covering roughly three hundred professionals and drawing on nearly a thousand real work examples, found that a worker's confidence in the AI system was negatively associated with the effort they reported putting into critical evaluation across five of the six cognitive activities the researchers tracked, with the effect strongest

at the evaluative stage where a person judges whether an answer is actually correct.[17] The researchers noted a self-reinforcing pattern: workers tend to withhold critical scrutiny precisely when they lack the underlying expertise to exercise it, which is also the condition under which scrutiny would matter most.[18]

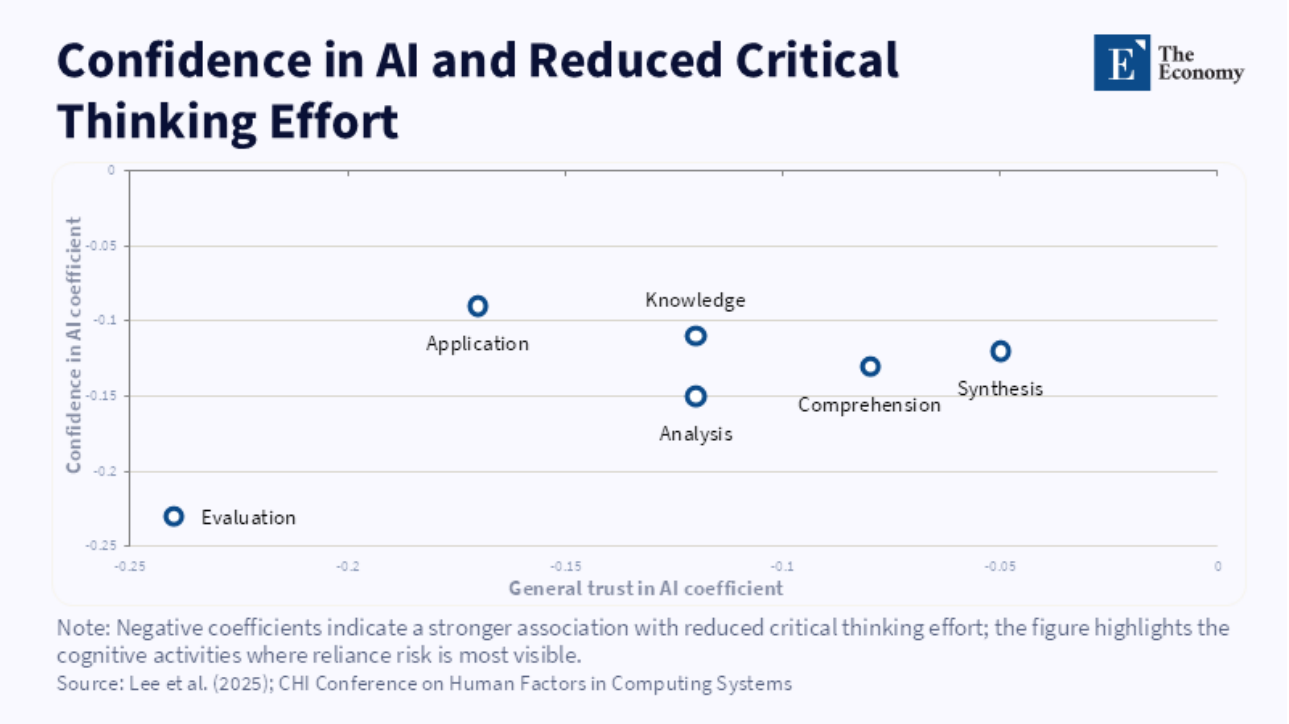

Figure 2: Evaluation is the clearest risk zone, where trust in AI is most closely associated with reduced critical thinking effort.

This is where the strongest version of the optimistic counterargument deserves a direct hearing, because it is not a foolish one. Skeptics of the cognitive-outsourcing narrative point out that similar warnings accompanied the pocket calculator, then the search engine, then Wikipedia and that in each case workers and students adapted without societies losing the capacity for arithmetic or research. A related claim holds that AI is fundamentally democratizing: by letting a novice produce work that resembles an expert's, it narrows gaps that used to track years of experience. Both claims are correct as far as they go and the second is well supported: the Brookings analysis cited above, drawing on a National Bureau of Economic Research study of more than five thousand customer service agents, found that AI assistance raised the productivity of novice and lower-skilled agents by roughly a third while barely moving output for the most experienced. This produces a compression of the bottom of the skill distribution that is genuinely equalizing.[19] But the comparison to the calculator breaks down in three specific ways that bear directly on a workplace setting. A calculator is reliably correct within its domain; a language model can be confidently, fluently wrong and only a user's own domain knowledge catches the error. A calculator executes an operation the user has already chosen; a language model frequently chooses the framing itself, silently performing the part of the task that used to show expertise: deciding what question is actually being asked. And the convenience differential is far larger: a calculator saves a few seconds of arithmetic, while a model can save an entire memo, brief, or analysis, which means the temptation to skip the developmental work is proportionally much larger too. The democratizing effect on offloadable work is real. It simply does not extend to the frontier and a companion study of consultants working with AI at Boston

Consulting Group found that when a task fell just outside that frontier, performance deteriorated by roughly nineteen percentage points relative to consultants working unaided, precisely because the tool's confident tone gave no signal that it had crossed into territory it could not handle.[20]

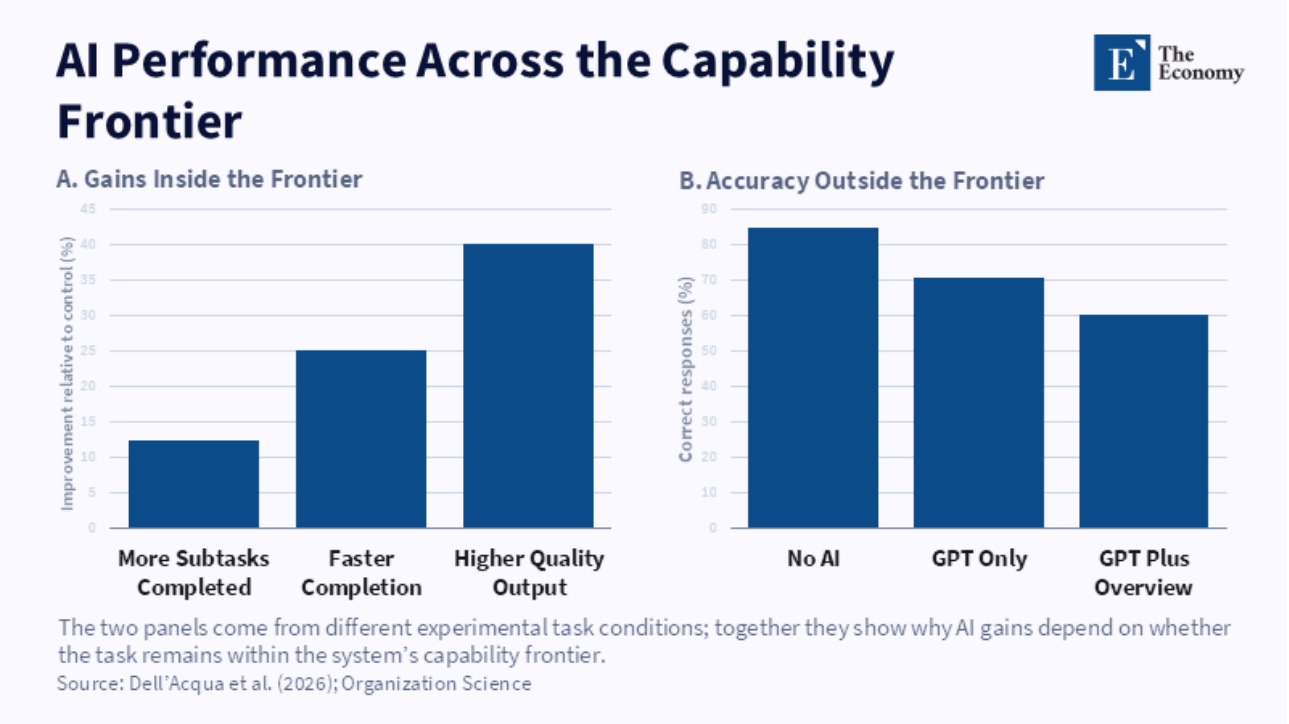

Figure 3: AI improves performance inside its capability frontier, but accuracy falls when users carry the same confidence into tasks beyond it.

That last finding points to the organizing distinction this paper relies on throughout, one that researchers studying classroom learning have arrived at independently of the labor economists studying customer service and consulting. Writing on cognitive offloading in education, Leon Furze, drawing on work by Jason Lodge and Leslie Loble, distinguishes between beneficial offloading, in which a learner hands off extraneous mental load to free up capacity for the actual work of understanding and outsourcing, in which the learner hands off the intrinsic cognitive work itself, the part that was the point of the exercise.[21] The same distinction, translated from a classroom to a payroll, is the difference between asking AI to format a report and asking it to decide what the report should conclude. The trouble is not that workers cannot draw this line in principle. It is that under time pressure, with no structural cue to slow down, the line is invisible from the inside and the tasks that most reward offloading, because they are the most tedious, are frequently adjacent to the tasks where offloading quietly becomes outsourcing.

Left unaddressed, this pattern compounds in a way that has a close analog in traditional outsourcing economics. Harvard Business Review's 2026 assessment of how generative AI is reshaping outsourcing decisions makes the point that firms increasingly have to evaluate automation and delegation task by task rather than function by function, because the cost of getting the boundary wrong is no longer symmetric.[22] The same asymmetry applies inside a firm that lets employees quietly outsource judgment rather than labor. A junior analyst who produces a memo with heavy, unverified AI assistance and one who produces the same memo through unaided effort turn in indistinguishable work product, as the Brookings analysis observes; the difference

is invisible in any measure that treats the memo as the unit of output. What is not produced, in the first case, is the slow accumulation of pattern recognition, the capacity to sense that a number looks wrong before checking it, that a contract clause is unusually worded, that a client's story does not quite add up, that constitutes the expertise the same organization will need from that analyst a decade later. That capacity, once skipped, is expensive to reconstruct precisely because it was never a line item to begin with and the shortfall shows up only once a case finally lands that the fluent, average-case competence of an AI-assisted worker cannot handle.[23]

At the executive level, the same mechanism operates with less friction and higher stakes. Analysis from the technology newsletter The Data Ecosystem describes a pattern among newly AI-enthusiastic leaders who mistake a rise in apparent throughput for a rise in value: instead of reading a report, a leader skims a five-line AI summary and treats the model's inferred perspective as their own judgment. For a large share of routine decisions, perhaps four in five, this substitution costs little.[24] The trouble concentrates in the remaining share; the complex, high-stakes calls where a leader's lack of independently held context becomes the ceiling on the quality of the decision, precisely when the decision matters most. Because senior leaders are structurally positioned to stay close enough to the work to catch what is wrong while remaining strategically detached, AI used this way collapses that balance rather than preserving it: it lets a leader bypass the team that would ordinarily supply context, producing what the same analysis calls a regression toward micromanagement, in which more topics are nominally covered and less genuine strategic judgment gets exercised. The result is something close to a quality-compounding loop, in which productivity gains remain visible on a quarterly dashboard while the erosion of independently exercised judgment proceeds more quietly, discovered, if at all, only once a consequential decision goes wrong.[25]

None of this is confined to any single rung of the organization. The mechanism is the same whether the actor is a first-year associate or a chief executive: cheap, fluent, confident output on tasks that resemble the tasks experts have always performed, substituting increasingly for the developmental or verificatory labor that built the judgment now being borrowed against. What differs by level is not the mechanism but the organizational consequences and the fact that fewer people sit above a senior leader to catch an error before it hardens into policy. The question that follows is why so many organizations have let this exposure accumulate rather than designing against it and the answer lies less in individual failure than in an absence the next section takes up directly: most firms have adopted a capability without building the system meant to govern it.

3 Firms Have Not Built a Proper Operating System for AI Adoption: The Missing Key Is the System

The previous section cataloged a set of behavioral symptoms. The more consequential diagnosis lies one level up, in the organizational architecture surrounding those behaviors or more precisely, its absence. The International Labour Organization's 2026 research brief on what it calls the aggregation paradox of AI supplies the clearest statement of the puzzle: task-level productivity gains in the range of ten to seventy percent, concentrated among less experienced workers performing well-defined, text-heavy tasks, simply have not shown up yet in firm-level,

sectoral, or national productivity statistics.[26] The brief's explanation draws on economic history rather than technological pessimism. Electrification and the earlier waves of information technology both took a generation to translate into measured productivity growth, because the gains from a general-purpose technology depend on complementary investment in reorganizing work, not merely on installing the technology itself. Economists have long called this pattern the productivity J-curve.[27] The OECD's parallel research on AI and skills reaches a structurally identical conclusion from a different data source: firm-level adoption of AI more than doubled between 2023 and 2025, yet close to forty percent of manufacturing and finance employers who have not adopted AI cite a shortage of relevant skills as the reason and adoption remains heavily concentrated among large, digitally mature firms rather than diffusing evenly across the economy.[28] The pattern in both institutions' data is the same. The technology arrived. The organizational reorganization required to convert it into durable value did not arrive with it.

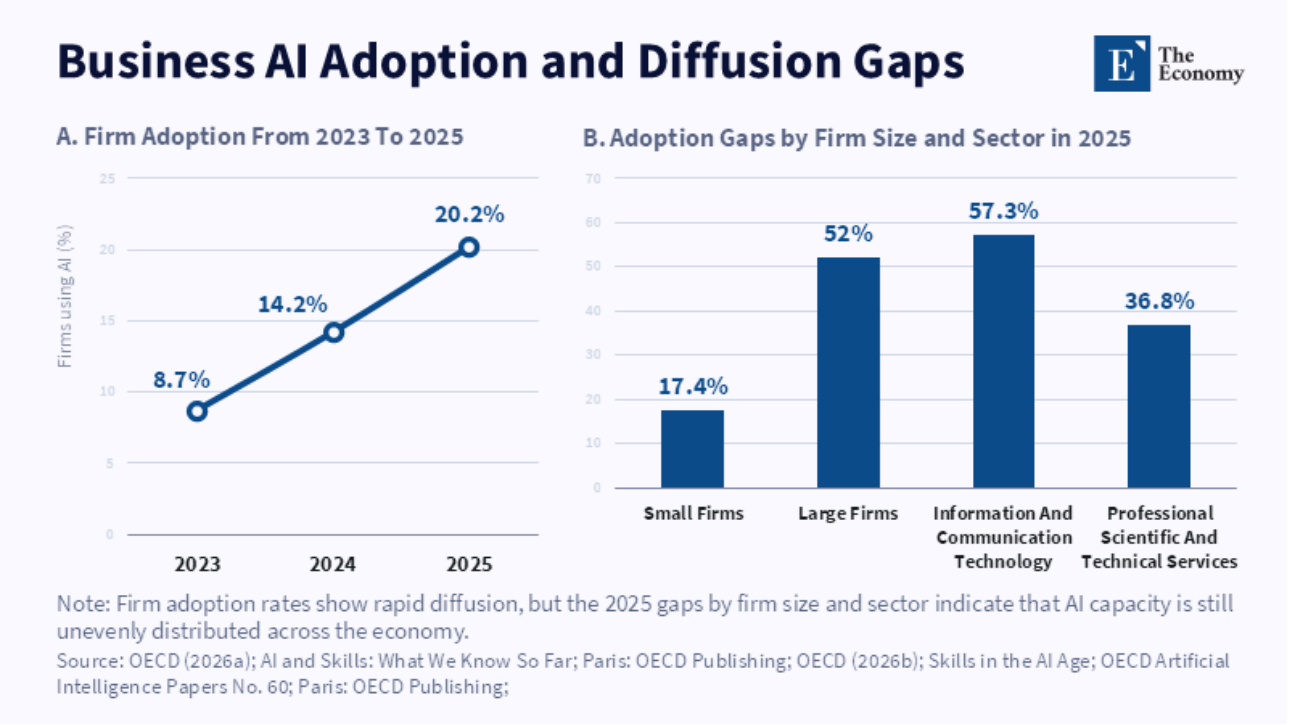

Figure 4: AI adoption is rising quickly, but the 2025 gaps show that adoption remains uneven by firm size and sector.

Info-Tech's software-development survey, discussed above for its behavioral findings, doubles as the sharpest available illustration of this gap. Ninety-four percent of respondents report meaningful productivity gains; only thirty-seven percent describe their governance of that gain as formally structured. Separate industry research, though methodologically distinct enough to be treated as directional rather than precisely comparable, tells a consistent story: developer-productivity researchers tracking well over one hundred thousand engineers across hundreds of companies have found adoption rates approaching universal even as measured productivity gains plateau in the low double digits, a gap the researchers attribute to the untracked time now spent reviewing, validating and fixing AI-assisted work rather than producing it. The self-reported sense of acceleration and the measured, validated outcome are drifting apart and the drift is largest precisely where governance is weakest.[29]

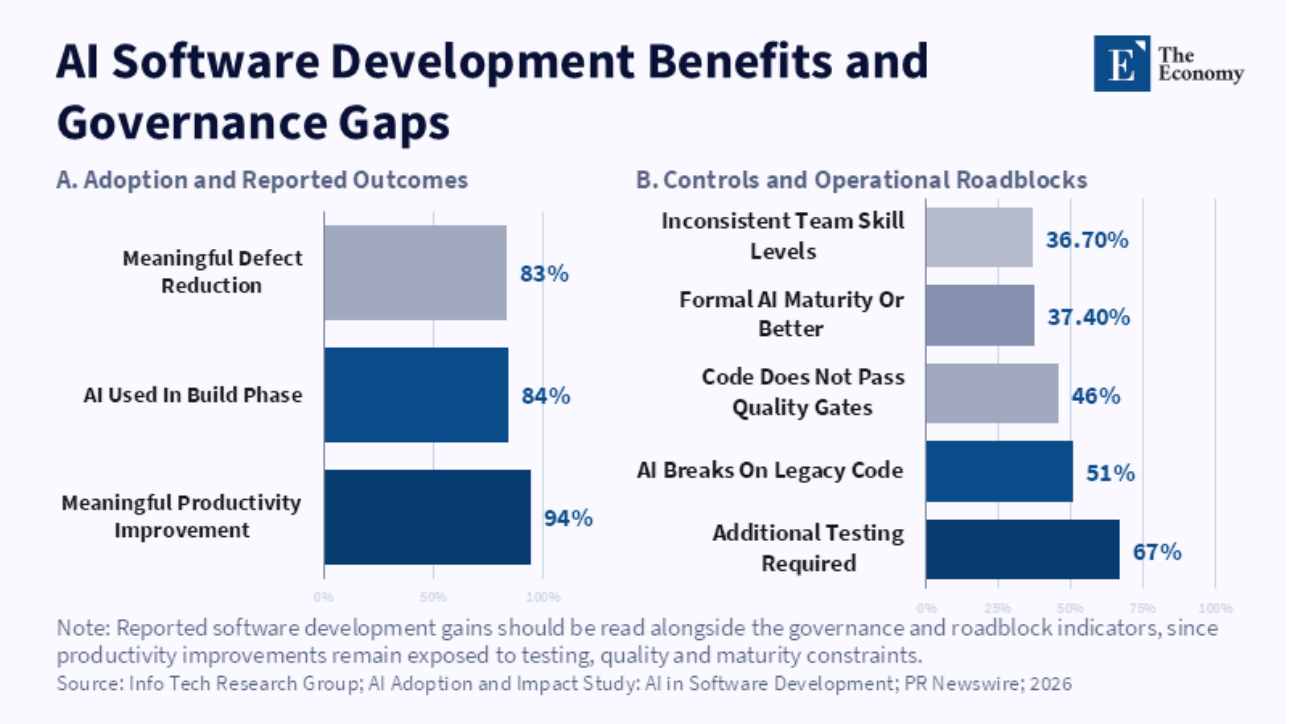

Figure 5: Reported productivity gains are strong, but they remain exposed to testing demands, quality limits and weak governance maturity.

What is missing, in other words, is not a more sophisticated way of extracting output from the models themselves. It is not, to borrow language increasingly common in engineering circles, a matter of optimizing token usage or prompt design. The missing element is a system in the more basic organizational sense: clarity about who is responsible for an AI-assisted output before it is treated as finished, a defined process by which that output is checked against domain knowledge before it is implemented and a governance layer that decides, deliberately rather than by accident, which categories of AI-based work require that scrutiny and which do not. A firm that has not settled these three questions has not really adopted AI in any sense that will show up in a decade-long productivity series; it has merely distributed a powerful, ungoverned tool to a workforce operating under exactly the incentives described in the previous section.

Building that system requires attention at three organizational tiers, each of which fails in a different way when neglected. At the level of individual work, the missing piece is a verification protocol built into the workflow itself rather than left to each employee's discretion under deadline pressure. Info-Tech's data on who actually reviews AI output is instructive here: developers with more than fifteen years of experience are roughly twice as likely as those with three to seven years to use dedicated AI security and review tooling and more experienced staff disproportionately report shifting their own time toward validating AI-generated work rather than producing it, while less experienced staff disproportionately report only the speed gains.[30] Verification behavior, in other words, tracks expertise closely, which means it cannot be assumed of a workforce that is, by construction, still building that expertise. A system at the working level has to specify, structurally, which categories of task require senior sign-off before AI output is trusted, rather than relying on every individual worker to correctly self-assess whether a given task sits inside or outside their own competence under time pressure.

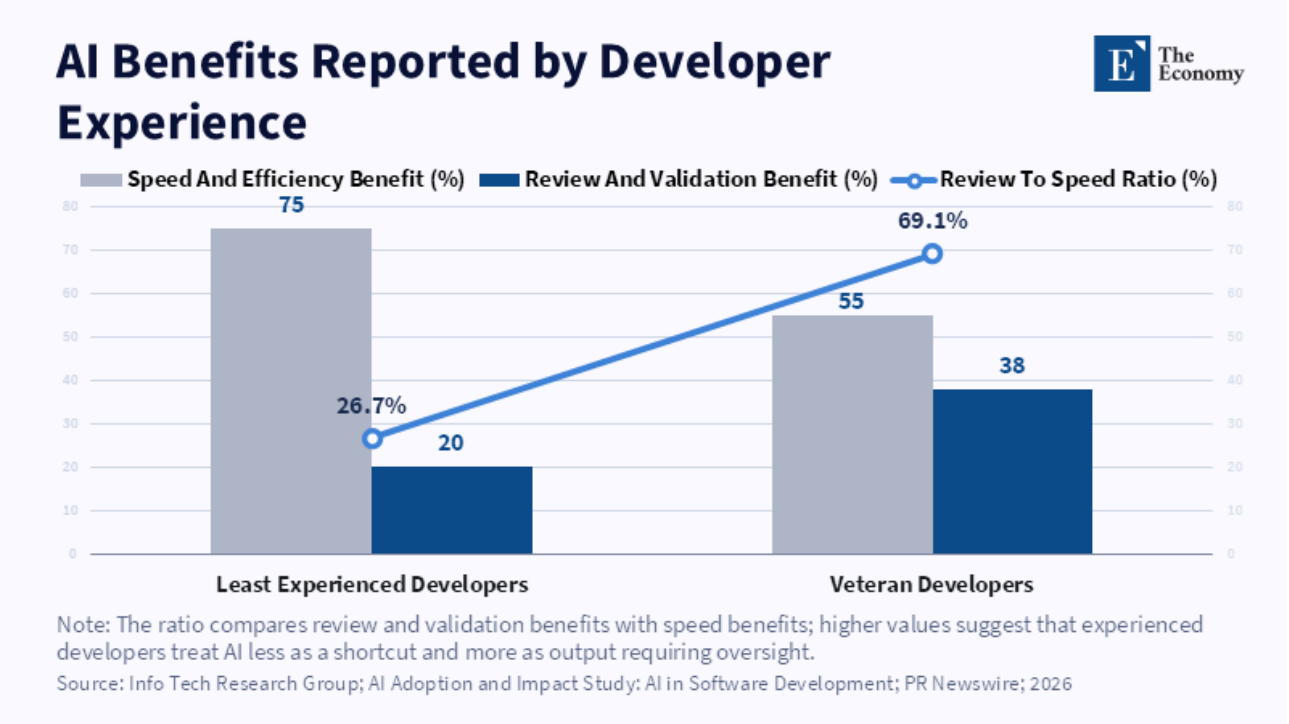

Figure 6: Experienced developers appear more likely to treat AI output as something to review, not simply as a shortcut.

At the management tier, the failure mode is the one described in the previous section: a collapse of the balance between closeness to the work and strategic distance from it. A useful way to see what a functioning system would protect is to look at what a recent labor-economics briefing from the Swiss Institute of Artificial Intelligence identifies as the durable human layer that persists even as routine tasks are automated. The briefing groups this layer into five capacities that AI does not substitute for even as it becomes more capable: domain context, the accumulated sense of what actually matters in a given industry or client relationship, which models can retrieve fragments of but do not own; judgment under genuine uncertainty, where interpretation matters more than pattern matching; accountability, in the literal sense that an organization needs a specific person or legal entity who owns the outcome; coordination across teams, vendors and stakeholders, which depends on a trust that a system cannot manufacture; and ownership of exceptions, the edge cases and novel failures where automation is weakest and human value concentrates.[31] A management-level system worthy of the name deliberately protects a leader's direct exposure to these five things rather than allowing AI-mediated summaries to substitute for them by default. That does not mean prohibiting the use of AI-generated briefings; it means designing explicitly for which decisions still require unmediated contact with the underlying material and treating a manager's judgment failures, when they occur, as at least partly a symptom of an unmanaged substitution rather than a purely individual lapse.

At the company level, the failure mode is governance itself: the absence of a clear answer to who is accountable when AI-assisted output turns out to be wrong and the absence of a deliberate framework for deciding which functions to automate, which to keep fully in-house with human ownership and which sit in between. Harvard Business Review's 2026 analysis of outsourcing economics captures the shift this requires: the operative question companies face is no longer whether an entire function like finance or human resources should be

handled internally or externally, but which specific tasks within that function can be safely automated, which still require outside expertise and which have become more strategically valuable to keep under direct human control precisely because AI has commoditized everything around them. Firms that have not made this task-level determination explicit are, by default, making it implicitly and inconsistently, task by task, decision by decision, in exactly the way the Info-Tech data suggests is currently happening across the software industry.[32]

The cost of neglecting this three-tier architecture is not confined to the present quarter. The same seniority-biased hiring pattern discussed earlier, firms quietly reducing entry-level recruitment because AI now performs the routine work that used to train juniors. Is itself a company-level governance failure with a long fuse: the juniors a firm declines to hire today were the internal candidates for the senior judgment that firm will need in fifteen years and that judgment, once lost, cannot simply be purchased back from the outside, because it is inseparable from years of accumulated context specific to the firm.[33] Any individual company that continues hiring juniors while its competitors do not will look, in the short run, less efficient than firms that have cut the cost. This is a coordination failure in the technical sense: rational for each firm acting alone, corrosive for the industry as a whole[34] and exactly the kind of problem no single company can solve unilaterally. It is also, as the next section argues, one of the clearer points at which public policy has a legitimate and fairly narrow role to play.

4 Policy Implications

It is worth being honest about why reasonable, specific policy proposals for cognitive offloading are so hard to find in the current debate. Most existing commentary defaults to one of two unsatisfying positions: restrict the technology, or train people to use it more responsibly. Both miss what the previous two sections establish, that the boundary between beneficial offloading and harmful outsourcing is a property of how a specific task or workflow is designed, not a property of the technology in general or of an individual worker's character. A government cannot legislate that distinction into existence at the level of a single spreadsheet macro or client memo, any more than it can legislate which software features a firm should ship. That redesign is inescapably a firm-level undertaking and, for the parallel question inside schools and universities, an institutional one, closely related to the credentialing and procurement recommendations a companion Brookings analysis has already developed for the education and research sector specifically.[35] The task here is narrower: to identify what a government can usefully do once it accepts that the primary redesign work belongs to firms and educational institutions rather than to itself.

The clearest such role is correcting coordination failures that no individual actor, however well-intentioned, can solve alone. The hiring pattern described in the previous section is the textbook case. No single firm can profitably commit to training juniors for an industry-wide pipeline whose benefits it will only partly capture,[36] since a well-trained junior is free to leave for a competitor; the arithmetic that makes sense for each firm individually produces an outcome, an empty senior pipeline a decade out, that is worse for every firm collectively. This is precisely the kind of externality that professional bodies, accreditation standards and public procurement rules exist to correct. Licensing boards in law, medicine, engineering and accounting can require that candidates

demonstrate a baseline of unaided competence at defined career gates, independent of how much AI assistance they used along the way, which preserves an incentive for firms to provide genuine developmental experience rather than AI-mediated throughput alone. Public procurement, which already sets labor and safety standards for contractors in many jurisdictions, could extend similar logic to professional-services contracts, favoring firms that can demonstrate structured, AI-free developmental phases for their own junior staff. None of this requires government to specify how a firm organizes its internal workflow; it requires government to fix the incentive that currently rewards every firm for free-riding on a pipeline it is simultaneously depleting.

A second, related role concerns disclosure. The gap Info-Tech documented between self-reported productivity and formally governed practice is invisible to the outside world precisely because nothing currently requires a firm to report it. The International Labour Organization's own recommendations point toward collective bargaining and social dialogue as vehicles for surfacing exactly this kind of information: provisions on training rights, transparency about how AI is used in performance evaluation and basic data protection.[37] There is no obvious reason the same logic could not extend to lighter-touch disclosure standards for publicly traded firms, comparable to the reporting regimes that have developed around cybersecurity incidents or safety practices in other industries. The point of such disclosure is not to prescribe a governance model. It is to let capital markets, insurers and customers begin pricing governance maturity the way they already price other forms of operational risk, giving firms a market-based reason to close the gap between what Info-Tech's ninety-four percent are reporting and what its thirty-seven percent are actually practicing.[38]

The most direct role for government and the least controversial is funding the transition itself, because the economics of training are a well-understood case of market underprovision: a firm that pays to train a worker captures only part of the return, since the worker can leave, so firms systematically underinvest relative to the social value of that training. The scale of the problem is not abstract. The World Economic Forum's 2025 survey of more than a thousand global employers found that they expect thirty-nine percent of workers' core skills to change by 2030, that fifty-nine of every hundred workers will need meaningful retraining over that period and that eleven of those fifty-nine are unlikely to receive it under current arrangements;[39] a gap concentrated, unsurprisingly, among workers whose employers are least able or willing to invest in them. A related survey by the Federal Reserve Bank of Boston found that fewer than three in ten workers believed they could adapt to AI on their own, against roughly half who believed they could adapt with proper training,[40] which is as clear a statement of unmet demand for public or publicly subsidized training infrastructure as labor economics tends to produce. Portable training accounts, sectoral training partnerships that spread the free-rider risk across an industry rather than concentrating it in any one employer and targeted subsidies for apprenticeship-style programs in AI-exposed occupations all follow directly from this diagnosis and follow considerably more directly than either a prohibition on workplace AI use or a general appeal to individual responsibility.

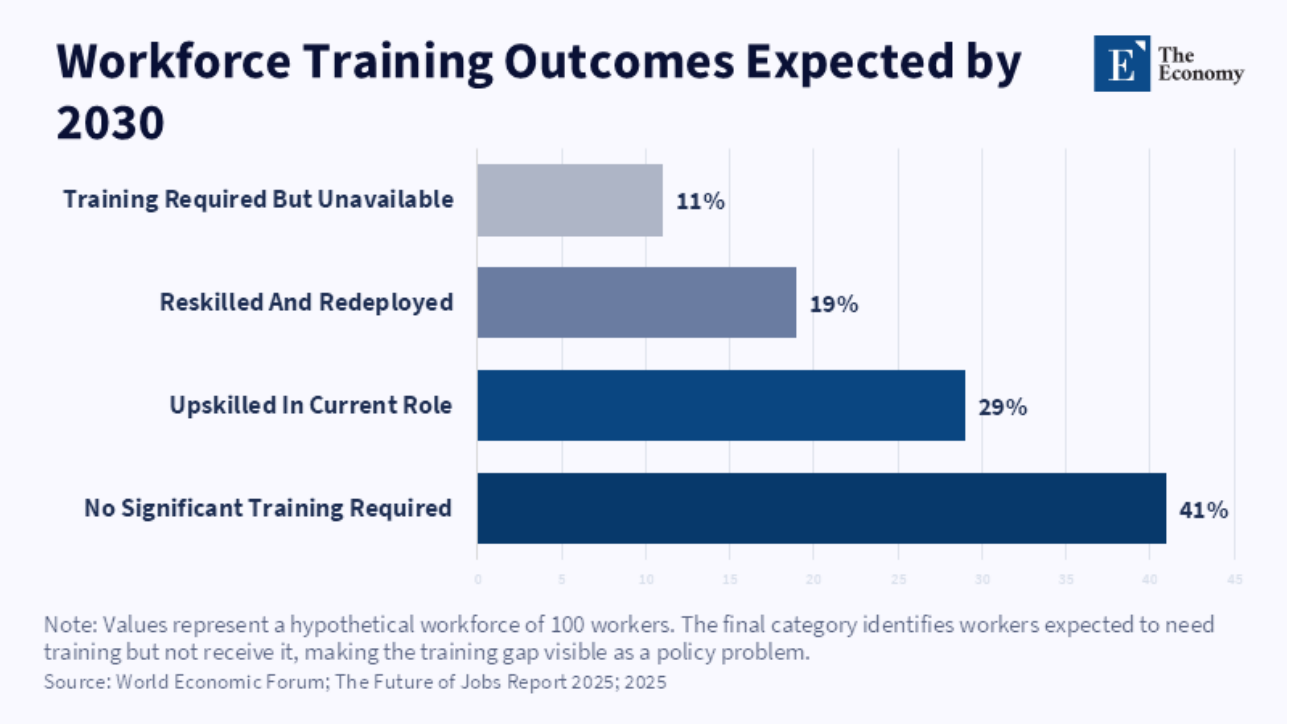

Figure 7: The training problem is not only the scale of reskilling, but the share of workers expected to need support and not receive it.

Regulatory design supplies a further, more direct lever and here the more defensible instrument is accountability rather than prescription. Rather than specifying how a firm's internal review process should work. That decision belongs to the firm and should be calibrated to its own tasks and risk tolerance. Policy can specify who is liable when unverified AI output causes a defined harm and require disclosure when AI materially shaped a regulated decision such as a medical judgment, a credit determination, or a legal filing. This kind of regulation does not try to design the operating system the previous section describes; it makes the absence of one costly enough that firms have a direct incentive to build one, a meaningfully different and more durable form of intervention than a compliance checklist. The International Labour Organization's warning is the relevant caution here: absent this kind of deliberate policy action, the gains from AI adoption are more likely to widen productivity and income gaps across firms, occupations and countries than to close them, since the firms and workers already best positioned to build governance capacity will simply extend their advantage while everyone else absorbs the downside risk described earlier in this paper.[41]

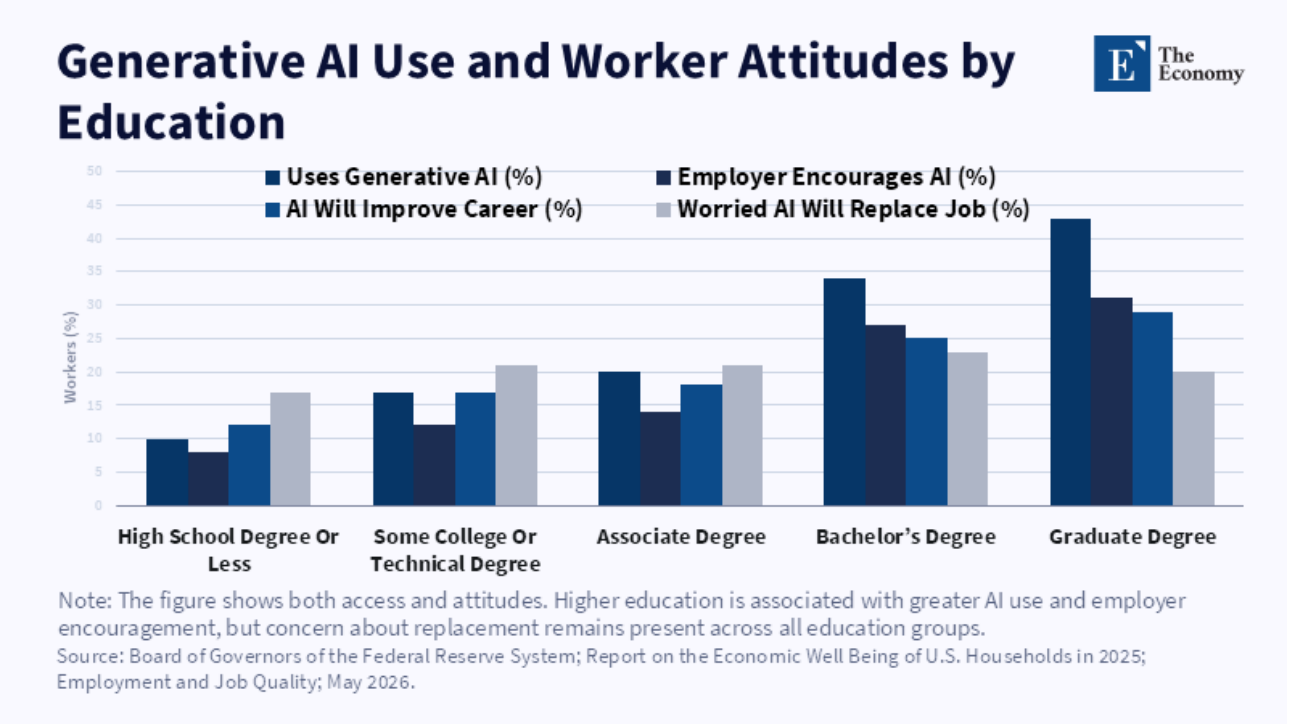

Figure 8: AI use and employer encouragement rise with education, suggesting that early workplace gains may reinforce existing human-capital advantages.

Put together, these roles describe a government that supports the two sectors doing the primary construction work. Firms must redesign their internal operating architecture while educational institutions must redesign how expertise is built. Government can support both without attempting to substitute its own judgment for theirs on questions neither legislators nor regulators are well positioned to answer. What government can do and currently is mostly not doing is remove the coordination failures, close the information gaps, fund the transition and set the accountability rules that make it rational for firms and workers to build toward becoming the kind of organization and the kind of professional capable of directing AI rather than being directed by it.

5 Conclusion - The AI User and the AI Governor

The productivity gains documented throughout this paper are real and nothing here argues otherwise. The risk was never that artificial intelligence would fail to accelerate work; it is that the acceleration is being financed by drawing down a form of capital faster than any current arrangement replenishes it. That capital includes verified judgment, institutional memory and the slow accumulation of expertise that only unaided effort builds rather than any current arrangement replenishes it. That imbalance surfaces first as a set of individual habits, then as a gap between reported and governed productivity inside firms and finally as a policy vacuum that neither prohibition nor exhortation can fill.

Filling it responsibly means firms building the three-tier system the second section describes, schools building the parallel architecture for how expertise forms and government confining itself to the coordination failures, disclosure gaps and training externalities that only it can fix. The stakes differ by seat but not in kind. Employees who take AI output on faith make their own roles the easiest to automate; managers who extend that same trust upward meet the same fate at a higher salary; firms that never build the system are simply

out-competed by rivals that do; governments that neglect either the corporate or the educational side of this transition inherit the resulting instability regardless. The labor market now sorting itself is not dividing along old lines of seniority or credential. It is dividing between AI users and AI governors and the sorting has already begun.

References

[1, 7] Dans, E. (2026) 'AI is creating the first generation of cognitively outsourced humans', *Medium*, 1 April.

[1, 13, 19, 23, 33, 42] Yaraghi, N. (2026a) 'Borrowed expertise: Why AI's productivity boom may not survive the generation that built it', *Brookings Institution*, 10 July.

[2, 3, 4] Tamkin, A. and McCrory, P. (2025) 'Estimating AI productivity gains from Claude conversations', *Anthropic Economic Index*, 25 November.

[5, 6, 29, 30, 32, 38] Info-Tech Research Group (2026a) *AI Adoption and Impact Study: AI in Software Development, June 2026 Top 10 Insights*. Arlington, VA: Info-Tech Research Group.

[5, 6, 29, 30, 38] Info-Tech Research Group (2026b) '94% of developers report AI productivity gains, but governance maturity lags behind adoption', *PR Newswire*, 20 July.

[8, 9, 14, 15, 16] Stackpole, B. (2026) "'AI gravity" is pulling you toward dependency. Here's how to push back', *MIT Sloan School of Management*, 2 June.

[9, 17, 18] Lee, H.P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R. and Wilson, N. (2025) 'The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers', *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, Article 1121, pp. 1–22.

[10, 26, 27, 37, 41] Chan, C.Y.C. and Shedania, K. (2026) *The Aggregation Paradox of AI: Why Do Micro-Economic Productivity Gains from AI Disappear at Scale?* Research Brief. Geneva: International Labour Organization.

[11] Brynjolfsson, E., Chandar, B. and Chen, R. (2025) *Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence*. Stanford, CA: Stanford Digital Economy Lab.

[12, 33, 34] Hosseini Maasoum, S.M. and Lichtinger, G. (2025) *Generative AI as Seniority-Biased Technological Change: Evidence from U.S. Résumé and Job Posting Data*. SSRN Working Paper 5425555, 31 August, revised 6 June 2026.

[14, 21] Furze, L. (2026) 'Resistance as a framework for combating cognitive offload', 22 March.

[16] Kosmyna, N., Hauptmann, E., Yuan, Y.T., Situ, J., Liao, X.H., Beresnitzky, A.V., Braunstein, I. and Maes, P. (2025) *Your Brain on ChatGPT: Accumulation of Cognitive Debt When Using an AI Assistant for Essay Writing Task*. arXiv preprint, 10 June.

[19] Brynjolfsson, E., Li, D. and Raymond, L.R. (2025) 'Generative AI at work', *The Quarterly Journal of Economics*, 140(2), pp. 889–942.

[20] Dell'Acqua, F., McFowland III, E., Mollick, E.R., Lifshitz-Assaf, H., Kellogg, K.C., Rajendran, S., Krayer, L., Candelon, F. and Lakhani, K.R. (2026) 'Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality', *Organization Science*, 37(2), pp. 403–423.

[21] Lodge, J.M. and Loble, L. (2026) *Artificial Intelligence, Cognitive Offloading and Implications for Education*. Sydney: Australian Network for Quality Digital Education and University of Technology Sydney.

[22, 32] Agrawal, A. (2026) 'AI is rewriting the economics of outsourcing', *Harvard Business Review*, 5 June.

[24, 25] Anderson, D. (2026) 'Leaders are starting to outsource their thinking to AI', *The Data Ecosystem*, 28 May.

[28] OECD (2026a) *AI and Skills: What We Know So Far*. Paris: OECD Publishing.

[28, 41] OECD (2026b) *Skills in the AI Age*. OECD Artificial Intelligence Papers, No. 60. Paris: OECD Publishing.

[31, 42] Lee, K. (2026) 'AI labor and human labor: Why replacement is slower than the hype, but more serious than workers think', *Swiss Institute of Artificial Intelligence*, 16 July.

[35, 36] Yaraghi, N. (2026b) 'Repaying the inheritance: How education and research policy can address AI's borrowed expertise', *Brookings Institution*, 20 July.

[39] World Economic Forum (2025) *The Future of Jobs Report 2025*. Geneva: World Economic Forum.

[40] Bracha, A. and Tang, J. (2025) *Shaping the Future of Work: Workers' Optimism and Pessimism about AI*. Current Policy Perspectives 25–16. Boston, MA: Federal Reserve Bank of Boston.