

The AI Labor Shift: Replaceable Tasks, Irreplaceable Judgment and the New Politics of Work

The Economy Research Editorial^{1,2}

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

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

Abstract

This paper reframes the debate on Artificial Intelligence and labor by shifting attention from occupation-level forecasts to the task structures through which firms actually reorganize work. AI is conceptualized not as a universal productivity device or a future substitute for all human work, but rather as an institution-forming technology that reshapes scarcity, education, incentives for management, and distributions of opportunity. The key distinction drawn is among replaceable, irreducible, and new tasks that arise as old ones are automated. Replaceable work-tasks that are patterned, uncustomized, and cheap to verify are already a strike one for automation in fields such as generic writing, automated customer service, coding aids, and administrative processing. Irreducible work-tasks that rely on judgment, context, accountability, and trust in institutional decision-making - such as accounting, auditing and regulated professions - continue to operate within established structures of trust and liability. While the automation of some tasks will destroy other jobs, new tasks (metawork) arise from AI applications, such as prompting, task workflow creation, output verification and human-in-the-loop processes. It is unlikely these new tasks can take up the slack of replaced work effectively: these new, "higher-skill" tasks tend to be structured in ways that compensate individuals who possess domain expertise and verification capability. The argument is not that we face a jobs apocalypse; rather, we already see a re-stratification of the labor market. The question is not whether or not AI will destroy work, but rather the governance of how expertise, risk, and career trajectories shift in its wake.

1 Introduction - From Job Forecasts to Task Stratification

The debate over AI and labor is too abstract. Carnegie Endowment for International Peace lists three positions regarding the future of work: the alarmed, the patient and the excited.^[1] This framework may summarize elite sentiment, but it fails to capture firm-level dynamics. Firms don't automate entire occupations. They atomize occupations into tasks,^[2] offload liability, reorganize management and reengineer work processes. The true target is not the occupation as a formal title, but the workflow through which that occupation is performed. Using these three task types, decomposable, irreplaceable and emergent (tasks created by reengineering existing tasks), can highlight the importance of the task as the focus of inquiry and not just a linguistic, regulatory and social group concern.

There is an urgency in this reorientation. Between 2023 and 2026, changes that once appeared likely to unfold gradually compressed into a far shorter institutional adjustment window.^[3] From Stanford HAI, 4.4% of coding problems were successfully completed by AI in 2023, jumping to 71.7% in 2024.^[4] Though global organizations are keen to emphasize that AI doesn't necessarily increase unemployment, the IMF finds 39% of global occupations to be susceptible to AI^[5] and therefore it cannot be said that technology has not had an impact. Technology moves so quickly that the labor markets and their surrounding regulators can't possibly keep up.

This article's central argument is that AI must be conceptualized as an institution-shaping rather than a productivity-increasing tool affecting human capital, management and inequality by altering scarcity dynamics.^[6] The productive impact of AI is less about speeding existing tasks and more about eroding routine cognition, eliminating apprenticeship, shifting verification into hidden forms and creating a new layer of meta-work required to define and repair the AI's output. Therefore, the key questions must be, "How much are the existing tasks increasingly systematizable versus dependent on judgment," and "Will new tasks created by the AI absorb displaced workers." In the 2023-2026 window, the emerging pattern is one of neither Utopia nor dystopia: white collar, routine tasks are vanishing at an accelerating rate, judgmental tasks remain enduring and the new tasks created around AI are narrower and more unequally distributed than tasks which are vanishing.

2 Replaceable tasks: Patterned Work and Cheap Verification

The tasks most amenable to replacement by AI tend to be those with an intuitively familiar economic logic—those for which highly pattern-driven, low-customization, relatively low-noise inputs, with a cheap evaluation of the output are likely to lead to rapid advances in generative AI, such as those employed in draft generation, summarization, templated communications, customer service, simple code generation and other symbolic productions for which the task structure is replicable and for which the "correct" answer is knowable with surface-plausibility or cheap empirical examination.

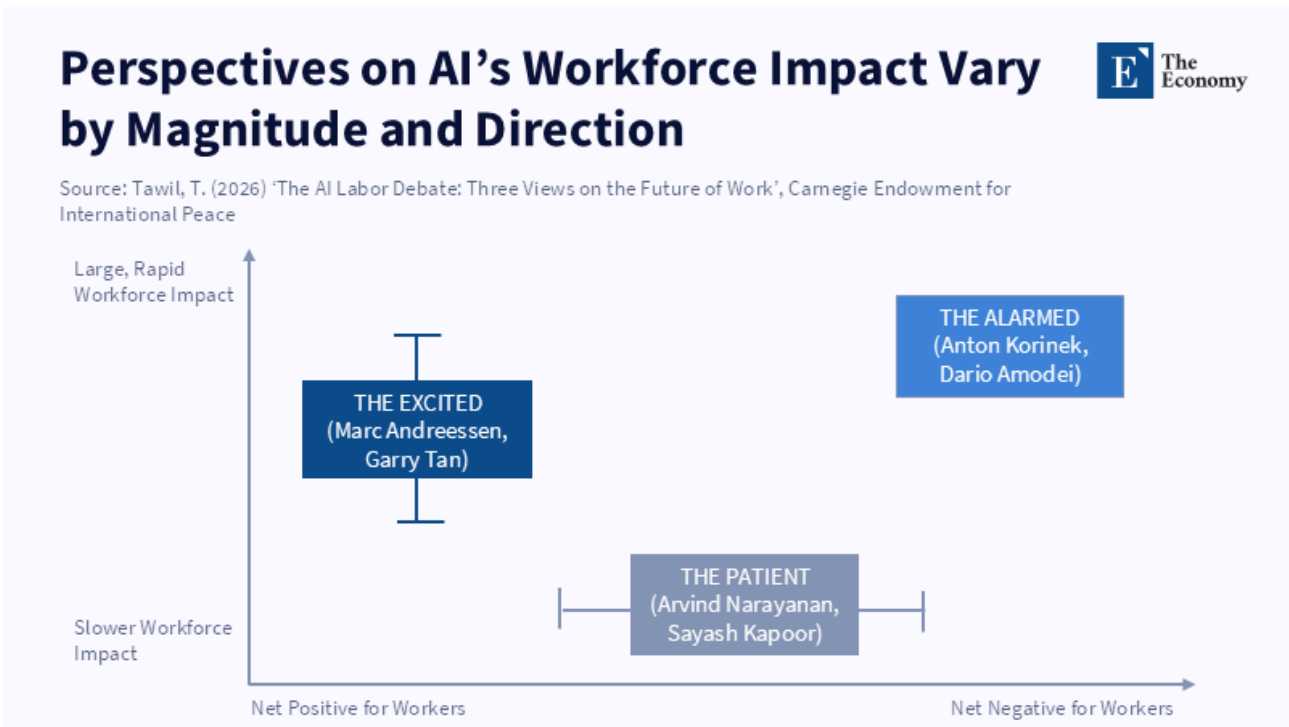


Figure 1: The AI-labor debate is shaped by assumptions about speed, scale and whether the impact is ultimately positive or negative for workers.

Experimental evidence beginning in late 2023 demonstrates this. Two commonly cited studies find that average task completion times on medium-skilled professional writing work decreased by 40% and quality improved by 18% after workers were given access to ChatGPT;^[7] a large-scale field experiment of call centers found a 15% increase in total productivity^[8] (roughly 30% for the lowest-skilled jobs) as the AI pushed workers to within 30% of the output frontier of elite workers. But this data doesn't just show higher output; it can mean that the firm learns how to codify heretofore implicit know-how and thus distribute that know-how to all workers via software.

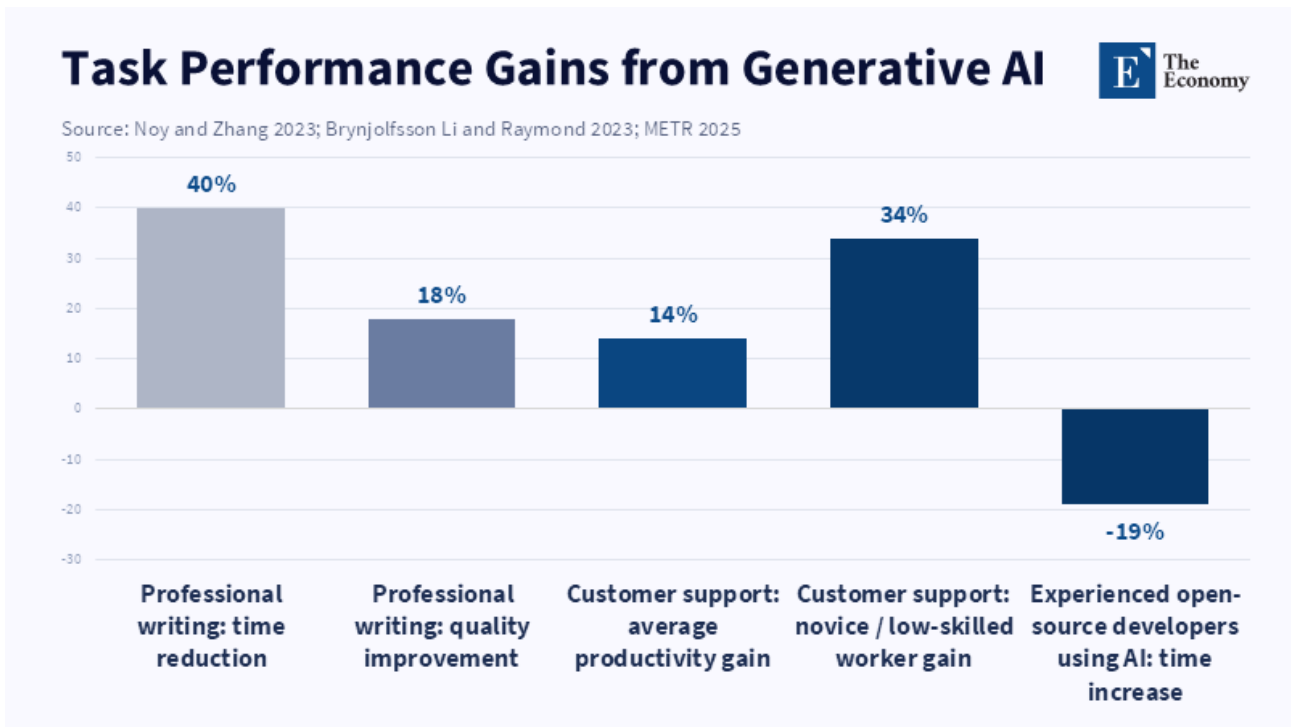


Figure 2: AI produces strong gains in patterned and verifiable tasks, but complex expert work can create hidden review and correction costs.

The codification of implicit know-how via software can explain the disruption in entry-level white-collar work. A late 2025 study at Stanford indicates that employment in AI-exposed occupations for 22-25-year-olds fell 6% from late 2022 to September 2025,^[9] whereas employment among the same cohort in low-AI occupations grew by nearly 9%. Likewise, Johnston and Makridis indicate that jobs needing human-AI collaboration grew by 4% since 2021, while jobs in which AI works alone showed little change in employment.^[10] The Stanford study also differentiates between automation and augmentation: losses were attributed to the former, stable jobs to the latter, indicating that the pertinent dichotomy is not the "human vs. Machine" distinction but that of "automation-prone vs. Non-automation-prone tasks."

This first category, most evident in the domain of software engineering, but applicable to others, requires careful explanation. While the claim that AI will replace junior engineers sounds right, it's not quite accurate: AI systems don't replace the institution of software engineering itself; they automate the easiest, most routine, most readily testable parts of coding: the scaffolding, the templates, the simple bugs, the simple interfaces. Even as of 2025, major platforms showed that the vast majority (79%) of human interaction with advanced coding agents was task automation, not task augmentation^[11] and few users noted errors requiring a back-and-forth; although this last point might just indicate new types of problems, the initial tasks which were the starting ground for new software engineers are now the targets for efficient "outsourcing" by firms via AIs. This isn't quite like spreadsheets replacing human calculators: Spreadsheets can accelerate rule-based, already-defined calculations to perfection, but generativity expands into areas such as language tasks, which weren't easily codified in the first place. However, second-order effects are likely to be much greater. Unlike spreadsheets, which sped up calculation but did not automate higher-level analysis, AI can produce plausible first drafts of memos, simple code suggestions, summaries of texts and other automations of routine tasks; critically, these actions do more

than increase the speed of task performance—they are capable of replicating (to a degree) human competence, potentially altering firms’ incentive structures, causing them to underinvest in training their workers and instead “buy” plausible outputs from an AI at a much lower price than the cost of a full human labor force augmented with new training. Critically, no task is fully “replaceable,” not even in occupations thought to be on the cusp. An experiment with experienced, professional open-source software engineers in 2025 found that access to highly performant AI tools caused them to spend 19% more time writing code,^[12] due to review, prompting and iteration costs, because the task context was complex, had high performance requirements and domain expertise and quality standards were well understood. This does not defeat the point that AI is powerful, but suggests that it is not perfectly or equally potent; jobs are replaceable in the extent that the task environment is not very complex, performance standards are routinely met with readily-checkable work and the cost of fixing any error is low. Thus, while a task might be ripe for replacement, the task’s context might be too path-dependent or deep for AI to do the entire job. Hence, occupation-level predictions are likely too coarse for policymaking.

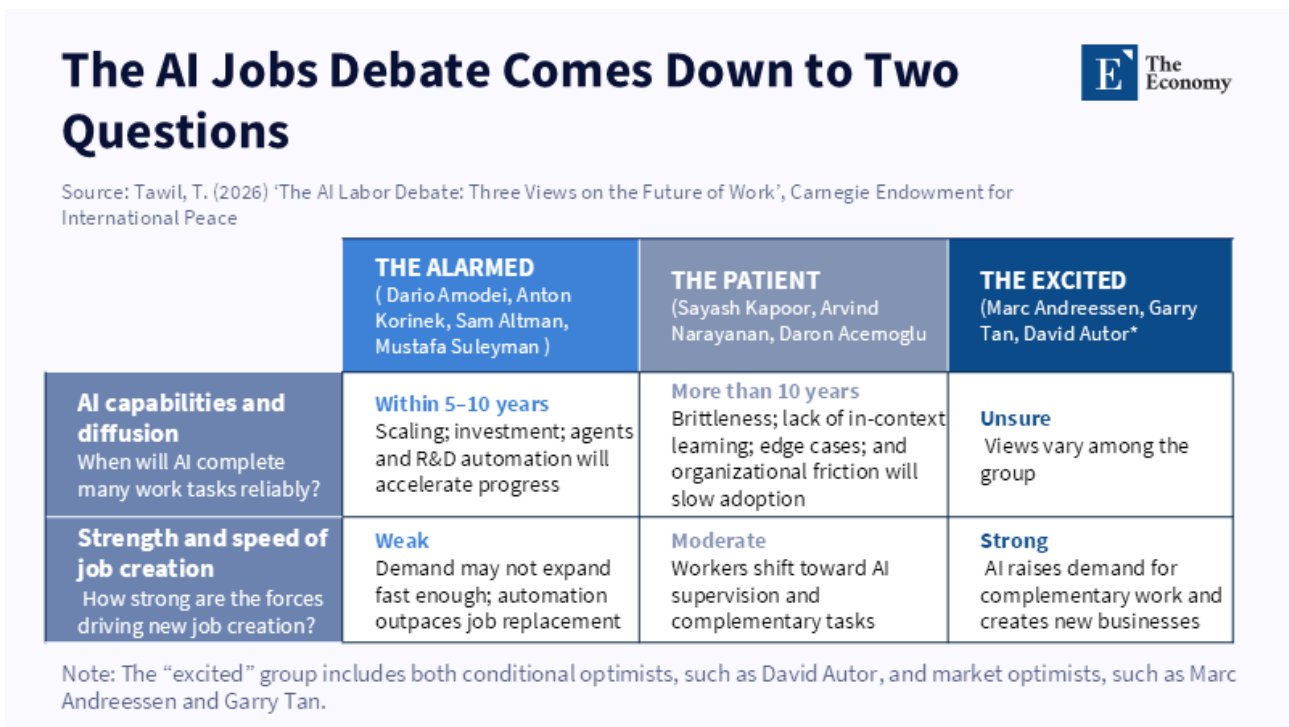


Figure 3: Competing views on AI labor disruption rest on two questions: how reliable AI becomes and whether new job creation is strong enough to offset automation.

The distributional impact is already obvious: Clerk work has the highest risk exposure globally, according to the ILO (3.3% of global employment in the highest risk category and more disproportionately women, among the highest-risk group);^[13] about 27% of jobs in OECD countries fit the highest risk category, ^[14]according to the OECD.

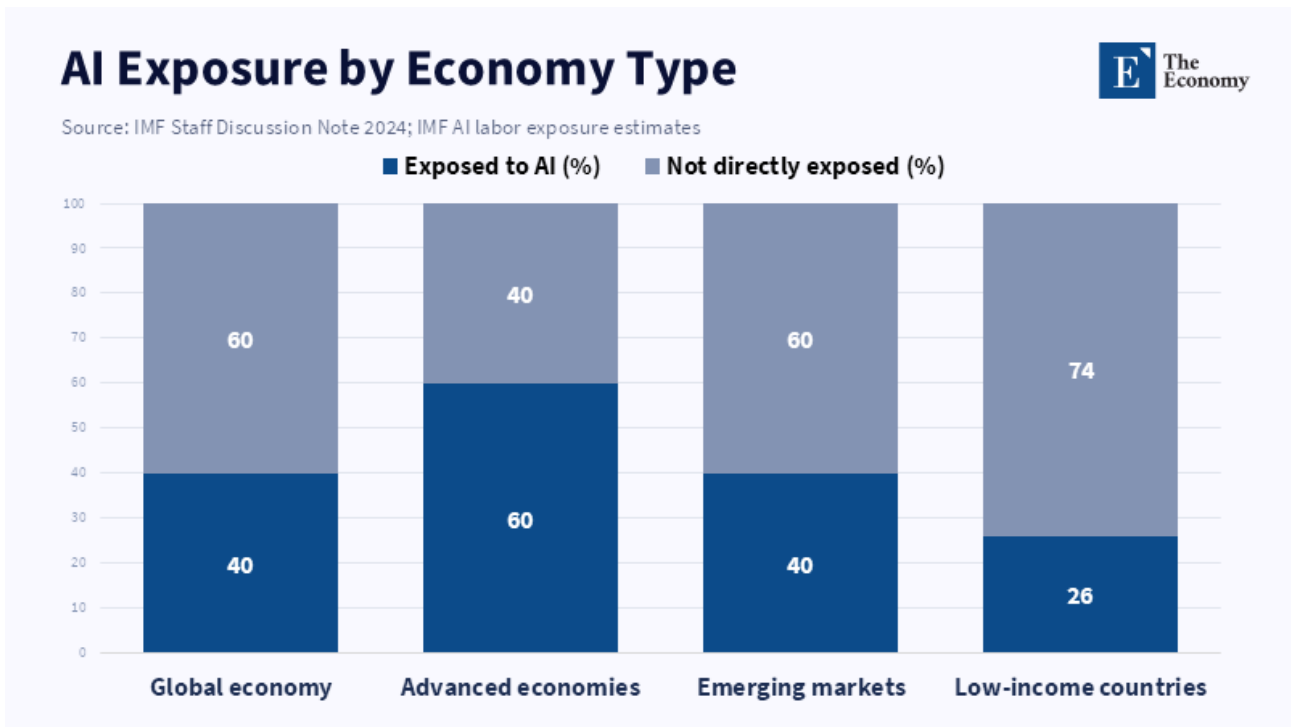


Figure 4: AI exposure is highest where cognitive, office-based and digitally mediated work already dominates employment.

Replaceable tasks are not a description of some part of economic activity; they are a social sorting of labor into work environments defined by routinized processing, clerical work and simple, codified procedures that have historically defined and separated lower parts of the labor market. Unless policymakers can shift focus from total levels of employment toward opportunity, AI’s distributional effects are likely to greatly increase already extant inequality while aggregate employment numbers stagnate.

3 Irreplaceable tasks: Judgment, Customization and Liability

It would be mistaken to assume that any substantial category of work is metaphysically immune to automation. The relevant question is not whether a machine can arrive at an answer, but whether organizations can afford to use its answer without unreasonable risk. What is irreplaceable about certain jobs is precisely this combination of customization, noisiness of evidence, ambiguity of context and real, costly to mitigate, liability. This suggests a core scarcity in many professional fields is not a want of output but a want of reliable judgment. Highly regulated environments have resisted full automation despite a growing level of superficial language competence on the part of LLMs for exactly this reason; the question is not solely one of quantitative limitations but of the integration of numbers, rules, intent, exceptions and accountability, where even a single mistake has serious legal, ethical and/or financial consequences.

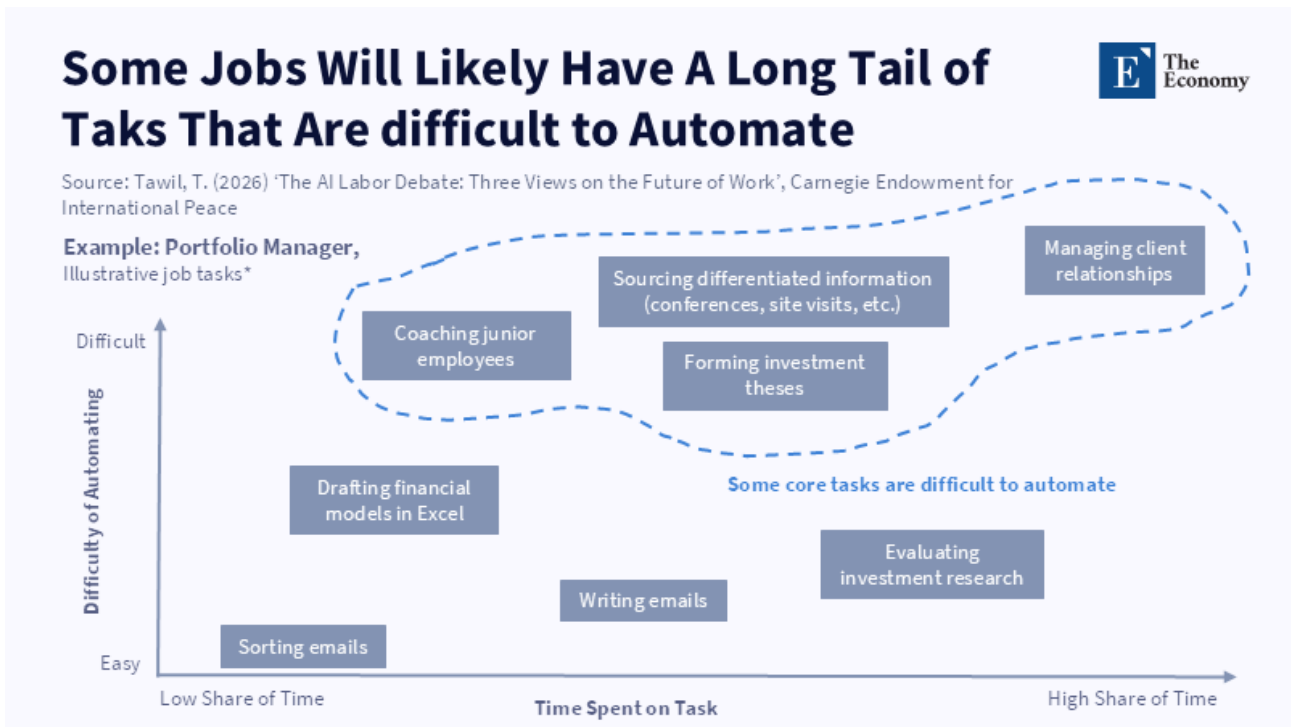


Figure 5: Even highly exposed occupations contain task clusters where judgment, relationships, tacit knowledge and liability remain difficult to automate.

This is visible in the field of accounting, mischaracterized at two extremes. On the one hand, many envision accounting as mechanical bookkeeping with full automation in mind; on the other hand, it has been presented as entirely unsuited to numbers and records for the sheer centrality of quantitative components. But what has changed recently is the surprising performance gains in testing conditions: it was found that a recent (November 2022) instance of GPT-3.5 averaged 53.1% on the professional certification exams and failed all tests.^[15] This could be improved through prompt engineering, multiple model calls and integration of calculators, such that with this latter modification, the performance on average jumped to 85.1% with passing on all tests.^[16] Latest generation models can, in other words, pass a limited version of the accounting examination under controlled conditions. In the 2025 (operational-level) analysis for external auditing, however, while the use of LLMs was found helpful for generating routine workpapers, for example, their use was noted as insufficiently robust or dependable where specific auditing judgments were involved due to potential hallucinations, outdated data, confidentiality, objectivity and liability.^[17] The fact that these tasks remain in human hands is not mysticism, but a reflection of the cost of error.

This distinction is critical; test performance is simply not equivalent to operational capacity.^[18] It is one thing to be able to give an answer regarding revenue recognition, for example, when test conditions and well-formed questions hold; it is another entirely to interpret unstructured client records, flag discrepancies, cross-reference between contradictory pieces of evidence, or determine if escalation is warranted by virtue of a lack of certainty in the underlying evidentiary landscape. This has been empirically shown in research on accounting reasoning, which indicates that quantitative accuracy, rule-bound logic, cross-statement correlation and multi-step reasoning are confounded with the interpretation of language; they are not merely questions for interpretation. As such, the continuing hesitance towards full automation hinges on our inability to guarantee that anything a

model produces is verifiable in terms of audit trail, theoretical grounding and line of responsibility.

It is on this count that the matter of replaceability begins to clarify: the issue is not that LLMs are not capable of achieving desired outputs, but that they lack the governance and control framework required of institutional responsibility. NIST's work on AI risk management emphasizes a lifecycle perspective and includes an emphasis on verification and ongoing assessment:^[19] such checks are not merely additions to system design and deployment, but crucial components designed to mitigate against the tendency for probabilistically designed systems to simply push risk outward unless control is actively introduced. For fields like accounting, auditing and compliance, control through humans is not a vestigial practice but a core production input: the more a model is able to generate output, the greater will be the incentive for humans to trust it and if delegated output authority outstrips institutional capacity for verification, then perceived efficiency will transmute into unknown governance risk.

There are also subtler policy ramifications of this. AI will not only fail to replace certain fields in their entirety; it will transform the structure of these professions. This is likely to produce an acceleration of routine data entry, drafting and preliminary analysis, while the focus will be driven increasingly toward areas requiring greater-level judgment: oversight, complex problem-solving, client relations and interpretation, monitoring compliance with standards and making factually grounded decisions on the basis of relevant criteria. In many fields, this will likely produce higher productivity for senior professionals at the expense of the roles of mid-level and junior positions that are founded on a combination of repetitive tasks and stepwise skill development. Such patterns are visible in other areas now suffering from AI-based automation and in the regulated sectors themselves. The result will be what may be termed a split in job roles: both fewer entry and intermediate positions and an increased need for the role of a hybrid specialist that bridges a domain specialty with the skills required for AI oversight, documentation and modeling. If not actively developed, the tendency may be for institutions to fully automate away even the apprenticeship function through which reliable experts are produced.

4 New tasks by replaced tasks - Prompt Engineering: Interface Work After Automation

The emergence of novel roles, like prompt engineering, is generally framed around AI's impact on the labor force, but contrary to what one might assume about new skills emerging in regions where AI adoption is high, employment-to-population ratios have dropped more in regions with higher AI adoption,^[20] showing that the creation of these new roles does not balance out the loss of jobs from automation. A better framing is that these new roles exist, but mass stand-alone roles are overwhelming, not their future; rather, it's a temporary intermediary layer of interface work that only emerges when an organizational shift occurs with a general-purpose system that hasn't yet solidified its operating norms, its task decomposition with its introduction into workflow, or internal controls over its use. It is workers who learn to translate ill-defined organizational tasks into a set of structured prompts, know how to give their context, guide their search space and identify that what

on the surface seems like the correct answer is, in fact, the wrong one who get rewarded during this transition. That is why it first seems esoteric and quickly becomes mundane.

From 2021 to now, "job postings seeking AI-related skills (prompt engineering, fine-tuning, model validation) have 'surged', unlike postings for tasks that the AI handles for people (data entry, routine coding)," according to a 2026 study by Popa, Oprea and Bra. A study of 20,662 job postings in 60 countries only flagged 72 (just shy of 0.5% of the sample) as prompt-engineer jobs.^[21] The lesson is indisputable: prompting is not an emerging, separate profession; instead, it's diffusing as an underlying skill.^[22]

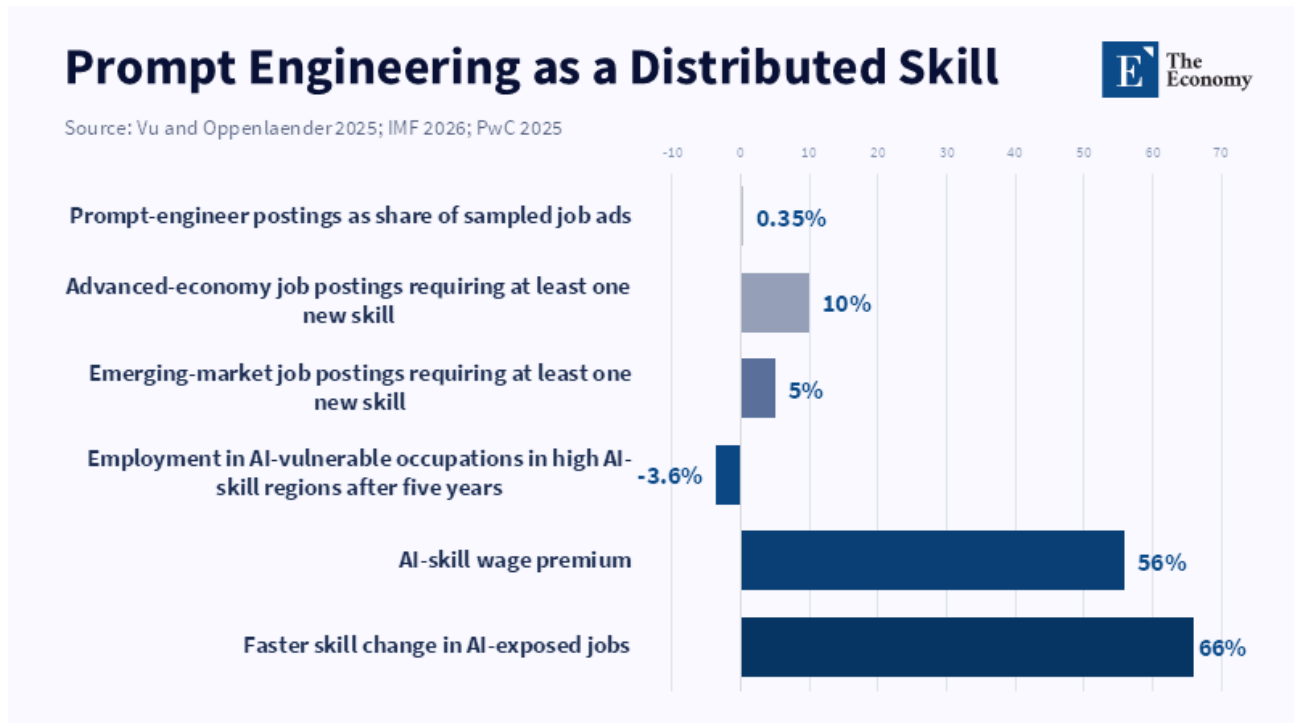


Figure 6: Prompting is not becoming a mass occupation; it is becoming a wage-relevant skill layer inside existing professional work.

The signals from the labor market are unmistakable: A 2025 industry survey found that workers with AI skills earn 56% more than those otherwise identical;^[23] IMF research on millions of jobs revealed that roughly 1-in-10 jobs in advanced economies require at least one new AI skill^[24] and those jobs pay a wage premium that reflects the specific application of the domain with the AI-mediated work, not a specific job title.

Here is why this new role is not a permanent priesthood of prompters, as some suggest; it's an office-software certification. The reality is that any new general-purpose interface will be rewarded as long as there is an organizational learning gap and so there will be workers rewarded in that transitional moment. For AI, this transition is already shifting beyond prompt writing to task specification, context-sensitivity, search strategy, task ordering and judgment of output accuracy. Data on US jobs postings from Harvard show that roles highly vulnerable to automation were declining, but roles capable of AI-mediated augmentation increasingly sought prompt writing ability and AI tool proficiency.^[25] And, though the introduction of AI doesn't mean mass, stand-alone jobs will emerge, the core significance to work comes in that prompting enables individual workers to do much more and smaller work than previously. This is the "super-human labor" many companies imagine, but it's also enabling a pace where being unproductive may not be tolerated.

The distributional consequences of prompting are likely the hardest to ignore. As prompting's only true utility is to complement subject matter expertise, workers already in possession of it will be further rewarded. It is unlikely to function as a great equalizing force in this manner but will rather bifurcate workers by the addition of prompt expertise, further distancing subject-matter experts in their respective domains who have mastered use of the models from those without. The semblance of democracy in it is institutional stratification; the pragmatic course is not to fund a new generation of inflated certifications in prompt writing, but to refocus vocational and professional training on AI interaction as a transversal discipline and incorporate it into sector-level standards and management guidance.

5 New tasks by replaced tasks - Eval: Verification as the New Scarcity

The prompt phase of market adjustment to generative AI is likely to be the first step and evaluation may prove to be the long-standing second step. As AI gets better at writing text, code, summaries and recommendations, the scarce commodity will not be text, code, or recommendations, but judgment. This is not trivial. An element of the highly valued human work of draft writing may become predominantly deciding whether or not a draft is "correct, robust, safe, compliant and fit for purpose." Evaluation is indeed already being discussed as a problem in exactly this sense; a 2026 paper entitled 'What can we know from evaluating LLM prompts?' argues that existing metrics of task performance accuracy will not work for text narratives, multi-step reasoning or quality dimensions that are intrinsically subjective.^[26] This is, in essence, what the National Institute of Standards and Technology has already been doing in the realm of official policy; generative AI requires vetting, validation, benchmarking and ongoing monitoring as a condition, not a possibility.^[27] As an example, in 2025, a NIST pilot study requested specific standardization of benchmarks for continuous performance.

The sets of activities appearing as evals, HITL or model QA in business, may not yet appear as steady jobs, but as supervisory work such as specifying the quality parameters, setting up the test sets, checking the output against established 'golden' answers, looking for false positives and false negatives, maintaining the level of quality, verifying claims or calculations and referring back complex or subjective issues to specialist humans. According to Stanford HAI, the interactions that are emerging in major coding platforms, such as the 'directive use' of code generators, feedback loops, iteration, validation and learning processes through feedback from human users are indicative of how these tools change workflows,^[28] though any resulting productivity gains may not be obvious for some time to come and the initial impact is likely to be on entry-level staff.

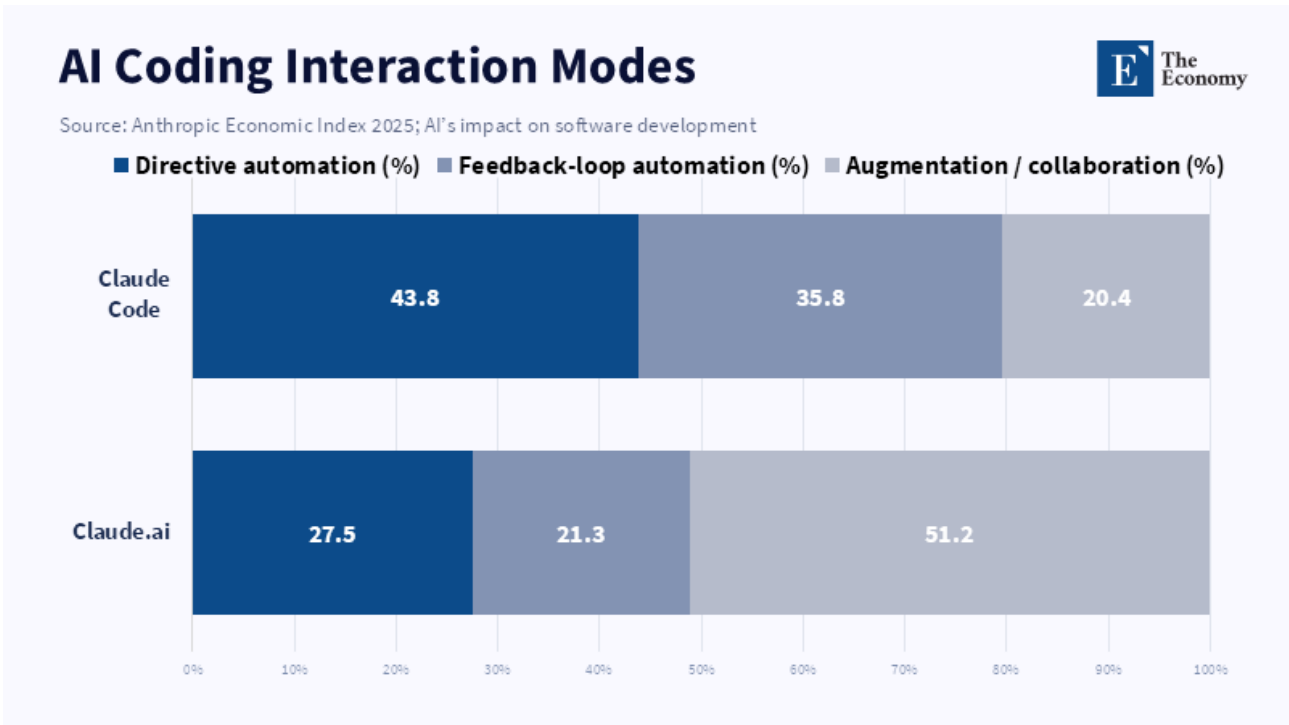


Figure 7: Agentic coding tools shift work toward delegation, but validation and feedback remain embedded in the human workflow.

It offers a critical view of human-in-the-loop work; touted and encouraged as the 'must-do' prerequisite of trustworthy AI, HITL actually describes what may prove to be low-value, mentally taxing, unspecified work.^[29] This work is not necessarily complementary to tasks being automated; it may directly replace them. An individual worker might be able to produce more sophisticated and responsible work, performing a greater amount of detail work, checking ten instances of AI-generated analyses than he would have to produce writing a single analysis, even where this involves the careful and conscious elimination of undetected errors rather than any construction of original thought processes. New forms of work emerging from these activities are not supplementary labor, nor residuals of automated tasks; they are entirely new forms of labor with specific new demands for expertise, skeptically focused attention, precise calibration and explicit accountability—a recipe for high-growth potential if well managed and accounted for, but also for inflated value because organizations, in the name of speed and efficiency, have tended to see evaluation tasks as a cheap add-on to true costs of managed output.

This argument itself therefore, seems to be a most damning critique of the thesis of simple labor displacement by AI: The burden shifted from doing work to overseeing it, from easily quantifiable work, to often difficult-to-value, cognitively demanding, inherently less standardizable cognitive work. Although the work is assigned to more senior people (senior analysts, domain experts, technology policy hybrids), compared to the entry-level positions being replaced (data entry, rote calculations, standard form completion), there is indeed equal, or perhaps more, human work being performed, but of lower quantity and higher skill. Therefore, one key policy consideration for institutions of all types moving forward must be whether, in their efforts to measure productivity gain, they will also consider and value the increased level of human work required for testing, validating, red-teaming and exception handling, or whether they will be tempted by false perceptions of pure

labor-substitution.

6 AI Job Apocalypse Won't Happen - Only Certain Jobs Will Disappear

The best conclusion to reach on the evidence to date, through the end of early 2026, is neither that AI will result in an immediate, mass unemployment shock, nor that it will unleash a new Gilded Age of human work. Instead, a more defensible conclusion-and therefore, a more useful analytical one-is narrower: generalized job apocalypse is improbable for some time to come, yet it is undoubtedly true that job titles, jobs and career ladders, at the minimum their tasks, components and constituent jobs will shrink, or decrease in their task and constituent jobs respectively, in specific economic sectors and levels. Even with continued total growth in employment, such declines will be consequential. Macro evidence: A WEF forecast of 170 million jobs created and 92 million displaced by 2030, producing a projected net gain of 78 million jobs^[30] and continued discussion of job transformation, as opposed to total job elimination, from both the International Labor Organization and the OECD, suggests a more modest effect. There also appears to be little net negative employment impact to date,^[31] although the OECD did caution an already present risk of continuing asymmetric and differential impacts.

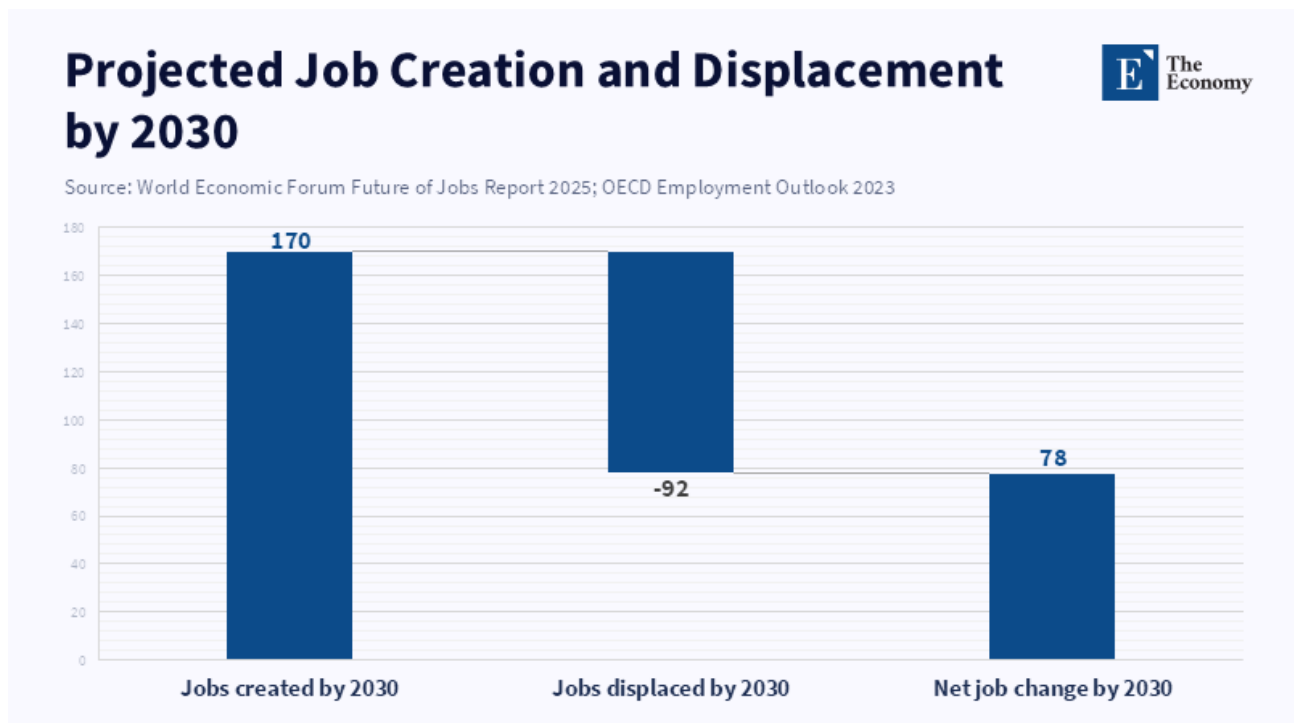


Figure 8: The aggregate forecast is not job apocalypse, but it still implies major churn across tasks, sectors and career ladders.

The total guarantees at the macro level do not erase structural damage at the micro level. It will not be workers who are affected evenly; clerical and administrative jobs remain in the center of task-exposure grids and younger workers appear to already fare more poorly than older ones at these jobs, while women are disproportionately represented in the highest-exposure occupational categories, especially in high-income

countries. Meanwhile, businesses simultaneously express optimism about AI (nearly four in five businesses already using AI report performance increases)^[32] and fears (three in five workers surveyed express concern about job loss in the next decade).^[33] Both responses make sense in their own way, but also imply that the same tool that makes work in one area faster can also undermine job security, lower bargaining power, or diminish career ladder progression in others. A boom in productivity need not necessarily coexist with a bust in employment: in the current institutional context, they are often two sides of the same reorganization.

Arguments to the contrary must also be engaged directly. First, AI must create general prosperity simply by raising individual productivity. This claim is not false, as experimental evidence clearly and repeatedly indicates performance gains from the tool across a range of delimited tasks (writing, customer service and others more advanced). Second, the technology "democratizes" capabilities by allowing less experienced and less proficient individuals to perform as well as or better than high performers after training. Again, it does enable some gains for lower-skilled and less experienced workers and some of the greatest improvements appear at the less experienced level. Third, it is sometimes compared to previous general-purpose technologies like calculators, spreadsheets and the Internet, all of which were initially feared but later successfully integrated as complementary technologies. All three of these counterarguments fail to fully address the matter at hand: Individual productivity increases do not automatically translate to overall worker employment and making task performance easier does not obviate the role of learning, on-the-job training or practice it displaces in some cases. Unlike calculators or the Internet, generative AI synthesizes and produces content that often replicates that of experts and thus offers a broader and deeper potential for replacement across task types than earlier technologies, especially by compressing expert pathways.

A more promising version of the optimist's case is that the technology will generate new jobs and forms of work by creating an entirely new set of tasks and processes, including but not limited to integration, evaluation, design and complementary services. This point by the Carnegie paper should not be entirely dismissed. However, those new tasks and forms of work are unlikely to be socially isomorphic with old ones; the task structure is based more on higher skill, more scarce labor and higher cognitive loads-integration, verification, compliance, service and AI-enabled consultation-than the eliminated tasks were. Thus, a much more pressing issue is the quality of labor market transition, rather than aggregate job totals. In the short term, companies that eliminate entry and mid-level jobs by way of internal reorganization and retooling, rather than new job creation, may produce yet another familiar dilemma: efficiency at the expense of the institutions that enable it over the longer term.

The policy implications are clear. Instead of broad-spectrum forecasting, policymakers should pursue granular task-level measurement in order to identify who is likely to be most vulnerable in entry-level jobs, how job ladders are likely to change and how those impacts are distributed across age, gender and industry sectors. Rather than chasing the disaster when it has arrived, supported transition measures such as wage insurance and skills upgrading ought to be implemented preemptively. Regulators, working closely with industry associations, should develop clear delineations between assistive, supervised and authoritative uses of AI and establish respective liabilities. Companies should not substitute internal organizational changes at the behest of efficiency from AI to the exclusion of developing junior staff and should be required to retain apprenticeship and review

structures, particularly on government contracts. Finally, public administration must design detailed explicit governance structures, including limitations on acceptable and prohibited uses of AI, requirements for validation and robust reporting mechanisms that currently do not exist, instead of allowing workers to bear the immediate costs.

This added force underlines the initial point of departure. The question has never really been about whether AI will 'displace workers,' but whether organizations would use it to restructure work more quickly than the labor market, institutions and society would adapt. Increasingly, the evidence points toward on the precipice of the latter, in which jobs displaced through automation are eliminated because they lend themselves to the task structure of AI when pattern, standardization and verification can be conducted cheaply, while jobs that are difficult to automate because they rely on judgment, context and liability remain irreplaceable; the new jobs and work streams created (e.g., prompting, integration, evaluation) will thus almost certainly involve narrower ranges and fewer openings than the displaced ones, suggesting that the impact is less likely to be a single event such as a job apocalypse and more likely to be a long and uneven process of re-stratification, with fewer low-skilled positions, increased demand for 'meta-work' (or work overseeing automated systems) and greater segmentation between those who can direct machines and those who cannot. Acting ahead of time is not alarmism; it is good governance.

7 Conclusion - Governing the Re-Stratification of Work

The question that AI poses is not whether work will be diminished globally, but in what ways the organization of work beneath the average labor-market figures will be transformed. Work is not replaced by AI in a neat, systematic fashion. The technology bifurcates tasks based on whether they are characterized by pattern-recognition, customization, verification-costs and institutional risk. Those tasks that lend themselves to systems and reproducibility (because they are routine and standardized) are vulnerable to automation. Those tasks which involve judgment, responsibility, context and accountability, are not exempt, but the value of human involvement has increased because of the inherent difficulty and cost of automation as well as the value of trustworthiness. Most important, then, is the transformation in the middle: the hollowing out of entry-level, task-based ladders and the development of novel, "meta" work-places in systems prompting, interfacing, and verifying tasks. We will need new jobs, perhaps not at the same numbers or skill levels, nor at the same rates, to do this work. If we allow AI to advance without governance, then AI-powered work will lead to greater labor-market stratification because it will create a labor market rewarding for those who can prompt, verify, and control systems, and leaving others behind-the ladder that they would previously climb has been destroyed. We are not facing a job apocalypse, nor a period of increased productivity gains for which we should wait passively. We are facing task-based regulation, better job training systems, secured pathways to learning through apprenticeship and direct institutional accountability for tasks mediated by AI.

References

- [1] Tawil, T. (2026) ‘The AI Labor Debate: Three Views on the Future of Work’, *Carnegie Endowment for International Peace*, 23 April.
- [2] Acemoglu, D. and Autor, D.H. (2011) ‘Skills, Tasks and Technologies: Implications for Employment and Earnings’, in Ashenfelter, O. and Card, D. (eds.) *Handbook of Labor Economics*, Vol. 4B. Amsterdam: Elsevier, pp. 1043–1171. doi:10.1016/S0169-7218(11)02410-5.
- [3] Bick, A., Blandin, A. and Deming, D.J. (2024) ‘The Rapid Adoption of Generative AI’, *NBER Working Paper* No. 32966. doi:10.3386/w32966.
- [4] Stanford Institute for Human-Centered Artificial Intelligence (2025) *Artificial Intelligence Index Report 2025*. Stanford, CA: Stanford HAI.
- [5] Cazzaniga, M., Jaumotte, F., Li, L., Melina, G., Panton, A.J., Pizzinelli, C., Rockall, E.J. and Tavares, M.M. (2024) *Gen-AI: Artificial Intelligence and the Future of Work*. IMF Staff Discussion Note No. 2024/001. Washington, DC: International Monetary Fund. doi:10.5089/9798400262548.006.
- [6] Cazzaniga, M., Jaumotte, F., Li, L., Melina, G., Panton, A.J., Pizzinelli, C., Rockall, E.J. and Tavares, M.M. (2024) *Gen-AI: Artificial Intelligence and the Future of Work*. IMF Staff Discussion Note No. 2024/001. Washington, DC: International Monetary Fund. doi:10.5089/9798400262548.006.
- [7] Noy, S. and Zhang, W. (2023) ‘Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence’, *Science*, 381(6654), pp. 187–192. doi:10.1126/science.adh2586.
- [8] Brynjolfsson, E., Li, D. and Raymond, L.R. (2023) ‘Generative AI at Work’, *NBER Working Paper* No. 31161. doi:10.3386/w31161.
- [9] 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.
- [10] Johnston, A. and Makridakis, C.A. (2026) ‘AI, Output, and Employment’, *CESifo Working Paper* No. 12579. Munich: CESifo.
- [11] Anthropic (2025) *Anthropic Economic Index: AI’s Impact on Software Development*. San Francisco: Anthropic.
- [12] Becker, J., Rush, N., Barnes, B. and Rein, D. (2025) ‘Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity’, *Model Evaluation and Threat Research*. doi:10.48550/arXiv.2507.09089.
- [13] Gmyrek, P., Berg, J., Kamiński, K., Konopczyński, F., Ladna, A., Nafradi, B., Rosłaniec, K. and Troszyński, M. (2025) *Generative AI and Jobs: A Refined Global Index of Occupational Exposure*. Geneva: International Labour Organization. doi:10.54394/HETP0387.
- [14] OECD (2023) *OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market*. Paris:

OECD Publishing. doi:10.1787/08785bba-en.

- [15] Eulerich, M., Sanatizadeh, A., Vakilzadeh, H. and Wood, D.A. (2024) ‘Is It All Hype? ChatGPT’s Performance and Disruptive Potential in the Accounting and Auditing Industries’, *Review of Accounting Studies*, 29, pp. 2318–2349. doi:10.1007/s11142-024-09833-9.
- [16] Eulerich, M., Sanatizadeh, A., Vakilzadeh, H. and Wood, D.A. (2024) ‘Is It All Hype? ChatGPT’s Performance and Disruptive Potential in the Accounting and Auditing Industries’, *Review of Accounting Studies*, 29, pp. 2318–2349. doi:10.1007/s11142-024-09833-9.
- [17] Fotoh, L.E. and Mugwira, T. (2025) ‘Exploring Large Language Models in External Audits: Implications and Ethical Considerations’, *International Journal of Accounting Information Systems*, 56, 100748. doi:10.1016/j.accinf.2025.100748.
- [18] Eulerich, M., Sanatizadeh, A., Vakilzadeh, H. and Wood, D.A. (2024) ‘Is It All Hype? ChatGPT’s Performance and Disruptive Potential in the Accounting and Auditing Industries’, *Review of Accounting Studies*, 29, pp. 2318–2349. doi:10.1007/s11142-024-09833-9.
- [19] Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P. and Roberts, K. (2024) *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*. NIST AI 600-1. Gaithersburg, MD: National Institute of Standards and Technology. doi:10.6028/NIST.AI.600-1.
- [20] Jaumotte, F., Kim, J., Koll, D., Li, E., Li, L., Melina, G., Song, A. and Tavares, M.M. (2026) *Bridging Skill Gaps for the Future: New Jobs Creation in the AI Age*. IMF Staff Discussion Note No. 2026/001. Washington, DC: International Monetary Fund.
- [21] Vu, A. and Oppenlaender, J. (2025) ‘Prompt Engineer: Analyzing Skill Requirements in the AI Job Market’, *arXiv*. doi:10.48550/arXiv.2506.00058.
- [22] Vu, A. and Oppenlaender, J. (2025) ‘Prompt Engineer: Analyzing Skill Requirements in the AI Job Market’, *arXiv*. doi:10.48550/arXiv.2506.00058.
- [23] PwC (2025) *The Fearless Future: 2025 Global AI Jobs Barometer*. London: PricewaterhouseCoopers.
- [24] Jaumotte, F., Kim, J., Koll, D., Li, E., Li, L., Melina, G., Song, A. and Tavares, M.M. (2026) *Bridging Skill Gaps for the Future: New Jobs Creation in the AI Age*. IMF Staff Discussion Note No. 2026/001. Washington, DC: International Monetary Fund.
- [25] Azpúrua, A.E. (2026) ‘Enhance or Eliminate? How AI Will Likely Change These Jobs’, *Harvard Business School Working Knowledge*, 20 February.
- [26] Hong, M., Lee, E., Park, S. and Kim, J. (2026) ‘PEEM: Prompt Engineering Evaluation Metrics for Interpretable Joint Evaluation of Prompts and Responses’, *arXiv*. doi:10.48550/arXiv.2603.10477.
- [27] Iyer, H., Seo, S., Diduch, L., Peterson, K., Awad, G. and Lee, Y. (2025) *2024 NIST GenAI Pilot Study: Text-to-Text Evaluation Overview and Results*. NIST AI 700-1. Gaithersburg, MD: National Institute of

Standards and Technology. doi:10.6028/NIST.AI.700-1.

- [28] Anthropic (2025) *Anthropic Economic Index: AI's Impact on Software Development*. San Francisco: Anthropic.
- [29] Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P. and Roberts, K. (2024) *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*. NIST AI 600-1. Gaithersburg, MD: National Institute of Standards and Technology. doi:10.6028/NIST.AI.600-1.
- [30] World Economic Forum (2025) *The Future of Jobs Report 2025*. Geneva: World Economic Forum.
- [31] OECD (2023) 'Artificial Intelligence and Jobs: No Signs of Slowing Labour Demand Yet', in *OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market*. Paris: OECD Publishing. doi:10.1787/08785bba-en.
- [32] EY (2025) *Responsible AI Pulse Survey: Companies Advancing Responsible AI Governance Linked to Better Business Outcomes*. London: Ernst & Young Global.
- [33] McClain, C., Kennedy, B., Gottfried, J., Anderson, M. and Pasquini, G. (2025) *How the U.S. Public and AI Experts View Artificial Intelligence*. Washington, DC: Pew Research Center.