Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Beyond AI Productivity: Enterprise AI Systems, Human–AI Responsibility, and the Future of Work

Audrey Rah

arahimi@uh.edu

University of Houston https://orcid.org/0000-0001-7501-7809

David W. Johnson

University of Houston

Systematic Review

Keywords: Employee outcomes, enterprise artificial intelligence, human–AI workflow, job insecurity, organizational performance, responsibility transfer

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11670337/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Beyond AI Productivity: Enterprise AI Systems, Human–AI Responsibility, and the Future of Work

Audrey Rah[1,*] · David W. Johnson[2]

[1]Department of Electrical and Computer Engineering, Cullen College of Engineering, University of Houston, Houston, TX 77204, USA
[2]Assistant Vice President, University of Houston, UIT Technology Services and Support, Houston, TX 77204, USA

* Corresponding author: arahimi@uh.edu
David W. Johnson: dwjohnson@uh.edu
ORCID: Audrey Rah — 0000-0001-7501-7809

***Abstract*—A prior U.S. review, *Enterprise AI at the Decision Point*, established that enterprise artificial intelligence (AI) assistants have task-dependent productivity effects and that public evidence does not identify an independent financial return or a complete chain from AI cost to verified workforce replacement. The human side of that productivity record, and any organizational consequences of workforce change, were left open. This study examines responsibilities, training, job insecurity and well-being, human oversight, and performance after U.S. enterprise AI adoption from November 2022 through September 2026. The design is a critical integrative review. Firms that report a task effect most often report exclusive augmentation. Frequent AI use is associated with greater fear of job loss. In one worker panel, displacement concern, rather than use frequency itself, is associated with lower engagement and higher burnout. The contribution is a responsibility-transfer framework with five constructs and four testable propositions. A proposed computational specification of that framework, Human–AI Responsibility-Aware Workflow Optimization (HARWO), places verification and accountability on the workflow graph. Gross task-time savings are separated from net employee time. HARWO is not solved or estimated in this paper. The evidence gap reflects a measurement disconnect: existing studies separately examine task productivity, workforce changes, and employee well-being, but the evidence examined here does not adequately link these outcomes within the same organizations over time. Consequently, whether AI-driven efficiency gains reduce human workload or redistribute responsibilities remains unresolved. The U.S. record is consistent with the separation of task change from employment change, and with anticipatory strain. The performance of organizations that combine AI with workforce reduction remains unmeasured, and the propositions are not empirically validated.**

***Index Terms*—Employee outcomes, enterprise artificial intelligence, human–AI workflow, job insecurity, organizational performance, responsibility transfer.**

I. Introduction

ENTERPRISE AI at the Decision Point, a prior U.S. review of enterprise artificial intelligence (AI), asked whether enterprise AI assistants and agents had produced productivity, cost, workforce, and return-on-investment evidence sufficient for high-stakes deployment decisions [1]. Through 16 September 2026, that review found task-dependent productivity effects, no independently measured financial return for named enterprise assistants, and no complete public case connecting actual AI cost to verified workforce replacement and realized return. It treated minutes saved as distinct from full-time-equivalent reductions, and a staffing cut as distinct from demonstrated replacement of human work. Its forward agenda called for time-use evidence on how remaining work is reallocated, matched employer–employee tests of whether AI expenditure precedes employment change, and job-count studies of hiring.

This paper leaves that investment question where the predecessor left it and asks the next one: after adoption, what happens to human responsibility, and what organizational consequences follow. Automation and augmentation theories allocate tasks between labor and capital [2]–[4]. Technostress and job-insecurity research explain strain from demands and anticipated loss [5], [6]. Human–AI collaboration research explains interaction with a system that remains dependent on human inputs [7]–[9]. None of those accounts defines the post-restructuring ledger: who absorbs the residual work, including the work of prompting, checking, and correcting the system when nobody has been laid off, and whether the firm that combines AI with headcount reduction performs differently from the firm that adopts AI without that reduction.

The contribution is the responsibility-transfer framework in Section II-C. It defines five constructs and states four propositions that can be tested. Section V states Human–AI Responsibility-Aware Workflow Optimization (HARWO), a workflow specification that makes residual responsibility and interaction overhead arguments of an allocation policy. Fig. 3 is the organizing diagram. The U.S. evidence through September 2026 is then used to classify each proposition. Classification is not confirmation, and HARWO is not a fitted model. Coding-assistant experiments, Danish task records, and U.S. training comparisons inform mechanisms [10]–[12]. They do not supply a new productivity estimate.

A. Research questions

1) **RQ1.** After U.S. firms adopt generative AI, what share of reported task change is augmentation, substitution, or role transformation, and does the mode of integration relate to reported employment change?
2) **RQ2.** What evidence distinguishes AI-caused employment change from employer attribution?
3) **RQ3.** What happened to remaining employees' responsibilities, training requirements, workload, and oversight of AI systems?

4) **RQ4.** What is the evidence on job insecurity, satisfaction, burnout, and engagement among workers exposed to workplace AI?
5) **RQ5.** Where managerial expectations and measured outcomes diverge, which dimensions are supported?
6) **RQ6.** After adoption and any workforce restructuring, what evidence shows improvement or deterioration in organizational performance?
7) **RQ7.** Do AI responsibility transfer and a human–agent organizing map survive contact with this evidence, or do they remain unmeasured constructs?

The remainder of this paper is organized as follows. Section II reviews the task, exposure, and collaboration literature that stops short of a fixed-roster ledger, and it states the responsibility-transfer framework and propositions P1–P4. Section III states the review design, eligibility, the searches that were actually exported, and the appraisal limits. Section IV classifies the evidence on adoption, restructuring, remaining employees, insecurity, and performance. Section V specifies HARWO: the workflow graph, the normalized allocation problem, the constraints, and a decision procedure that is not a tested program. Section VI states what the classes imply, why the human consequences remain difficult to measure, and what remains unmeasured. Section VII closes. Appendices A–H, after the references, record the searches, the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) 2020 inventory, screening, study characteristics, the appraisal boundary, the synthesis, the HARWO notation, and the citation audit.

II. Background

A. Tasks, automation, and reinstatement

The labor-economics foundation is the task model. Production requires tasks that can be allocated to labor or capital. Automation reallocates tasks from labor to capital and can reduce labor's share even when productivity rises. New tasks in which labor has comparative advantage produce a reinstatement effect [3], [13], [14]. Occupations typically contain a bundle of tasks, so automation of some tasks has historically been more common than elimination of entire jobs [2].

Pre-generative-AI evidence already separated establishment-level labor-demand changes from aggregate employment. Industrial robots had adverse employment and wage effects in U.S. commuting zones [15]. In online vacancies from 2010–2018, AI-exposed establishments increased AI hiring, reduced non-AI hiring, and changed skill requirements, while occupation- and industry-level employment and wage effects remained too small to detect [16]. That pattern is a prior for 2023–2026 claims, not a generative-AI result.

B. Augmentation, hidden labor, and employee strain

Automation and augmentation are interdependent, and organizations can cycle between them [4]. Exposure is not realized substitution. Standalone large language models were estimated to affect more than half the tasks in only about 1.8 percent of jobs, while complementary software raised the share of jobs with majority-task exposure to just over 46 percent [17]. Those two shares are exposure, not realized substitution. A 2025 International Labour Organization (ILO) index likewise scores occupational exposure, not hours or separations: about one in four workers worldwide is in an occupation with some generative-AI exposure, and 3.3 percent of global employment falls in the highest exposure gradient, with higher shares in high-income countries [18]. A separate pre-generative-AI exposure index ranks occupations, industries, and places; it is not the source of the 1.8 and 46 percent figures [19].

Technology is enacted through organizational routines [20]–[22]. Apparent machine autonomy can rest on uncounted human review, labeling, exception handling, and emotional labor [23], [24]. Algorithms can restrict, recommend, record, rate, replace, and reward, reshaping control without removing work [25]. These are theoretical expectations for residual labor. They are not measurements of generative-AI oversight hours.

Three further peer-reviewed accounts specify the object without supplying that measurement. Managing AI is a decision about autonomy, learning, and inscrutability [26]. Working with AI is a counterpart inside a system of design, implementation, and use, rather than a tool that simply receives a task [7]. Where many algorithms delegate to many humans, delegation is distributed: rights and responsibilities are not absorbed completely by the algorithm, and human inputs remain necessary for the delegation to function [8]. A multilevel review already separates human–AI collaboration, perceptions of capability, worker attitudes, algorithmic control, and labor-market effects [9]. Those distinctions motivate the framework below. The Uber study is qualitative mechanism evidence, and the multilevel review mixes designs and countries. Neither measures hours moved onto U.S. employees who remain after a separation.

Process representations, preference-weighted optimization, and human–AI interaction guidelines each stop earlier than that ledger. Process mining recovers a directed graph of activities from an event log [27]. A weighted sum is one way to state preferences before a search when objectives have different units [28]. Human–AI guidelines require that people be able to correct system output and that accountability for the output remain intelligible [29]. None of those sources defines the human time that remains when the set of employees is held fixed. Two closer accounts still stop short of that ledger. One reallocates call-center supervision from operator cognitive load, stress, and activity [30]. Another treats levels of automation as a shared language for team design and names accountability as a design requirement [31]. In the abstracts retrieved here, neither separates the person who executes a step, the person who authorizes release, and the person who remains accountable, and neither prices prompting and verification as human time distinct from case completion time. HARWO, in Section V, is the proposed joint specification. It is not a result of those literatures and it is not a solved allocation.

C. Responsibility-transfer framework

Responsibility transfer is the reassignment, after AI adoption, of tasks, decision rights, and residual work among the

AI system, workers who leave, and workers who stay. The framework has five constructs.

Task-reallocation mode is the firm's reported mix of augmentation, substitution, and task creation. *Employment object* is one of four non-equivalent records: a firm-reported headcount change, a flow of job postings, an employer's stated reason for an announced cut, or demand in a contract market. *Residual responsibility* is the human work of approval, exception handling, oversight, and integration that remains after a task is shared or removed [8], [24]–[26]. *Attitudinal exposure* is perceived insecurity, fear of elimination, burnout, engagement, and search. It is distinct from residual hours. *Organizational consequence* is the performance of the package that combines adoption with restructuring, as distinct from task speed and from aggregate productivity.

Four mechanisms connect the constructs. Each yields a proposition. None is treated as confirmed.

The task-employment mechanism follows the task model and the automation–augmentation cycle [2]–[4]. Task change can occur without occupation elimination, and broad operational integration is a different exposure from worker-level use.

P1. Among AI-using firms, a reported employment decrease is more likely when AI is integrated across functions and operations than when use is confined to worker tasks, conditional on size and sector.

The incomplete-delegation mechanism follows the management of autonomy and inscrutability, and platform evidence that algorithmic delegation still requires human inputs [7], [8], [26]. Sharing a task with a system does not, on that account, transfer accountability for the task.

P2. Adoption that changes tasks increases human oversight, exception handling, or integration work even when establishment headcount is unchanged.

The anticipatory-strain mechanism follows insecurity and technostress research [5], [6]. Strain can move with the expectation of displacement before any separation occurs, and it need not move one-for-one with use.

P3. Within person, higher perceived displacement risk is associated with lower engagement and higher burnout after AI-use frequency is taken into account.

Interaction overhead is the time and effort spent prompting, verifying output, correcting errors, switching context, and supervising the system inside a task that the employee still holds. It is distinct from residual hours after a separation and from total working hours. Gross task-time savings are the reduction in time to finish a specified task. Net employee time subtracts interaction overhead from that gross saving. The two are not interchangeable. This fourth mechanism asks whether a shorter task still leaves the employee with less work.

P4. Within a task, verification and correction effort can remain, and can track fatigue across repeated use, even when completion time is shorter than without AI. P4 does not imply that total working hours rise, and it does not require a layoff.

The advance is the joint claim. Automation and augmentation stop at task allocation. Technostress and insecurity stop at strain. Collaboration theories stop at the human–AI interaction. The framework requires task time, interaction overhead, attitudinal exposure, and organizational consequence to be reported separately. Consequence stays unidentified until hours absorbed by remaining employees, net of interaction overhead, and performance after AI plus headcount reduction are observed. Fig. 3 orders the constructs. Managerial expectations and organizational readiness remain candidate moderators [32], [33]. The readiness preprint is context for that moderator. It is not an outcome estimate.

Algorithm aversion supplies a further boundary: users may reject an algorithm after seeing it err [34]. A separate preprint states a mid-range theory of delegated commercial agency: organizations may transfer bounded decision rights to human or computational representatives while principals retain residual accountability, and fully autonomous cross-enterprise commerce is not claimed as a present fact [35]. That account is about interorganizational commerce. It does not measure hours absorbed by employees who remain after adoption. Institutional statements cited later describe skills, workforce questions, and audit practice. They record expectations and proposed oversight methods. They do not measure residual hours.

D. Why 2022–2026 rates are not one series

Worker use, firm adoption, employment-weighted adoption, functional integration, worker-task use, utilization intensity, and enterprise-agent use are different constructs. Worker-level generative-AI use exceeded 40 percent of employed adults by late 2025, while firm-weighted Census use was about 18 percent after the late-2025 question revision and employment-weighted firm use was about 32 percent in the 2026 supplement [36]–[38]. Access is not adoption, adoption is not utilization, and none of these is enterprise performance. No comparable national series for enterprise-agent prevalence was found. Fig. 1 keeps the series separate.

III. Methods

A. Design

The design is a critical integrative review [39]–[41]. That class develops a conceptual account across fields. It is the appropriate class because the questions cross labor economics, information systems, and organizational behavior, and because the estimands are not poolable. Reporting follows the PRISMA 2020 statement where the record supports an item [42], together with software-review and protocol guidance on eligibility, search concepts, and appraisal [43], [44]. The review was not registered. Screening was done by one reviewer. Database hit counts were not retained, so they are not imputed, and the predecessor's counts are not copied [1]. Fig. 2 is the historical corpus only. It has no identification count. The counted 8 November 2026 export is Appendix B and is not that figure. Table I is the reporting inventory. Paré et al. and Snyder are cited only for the review class [39], [40].

The verified reference list contains 104 works. One is the predecessor review. The other 103 are follow-up sources: journal articles and working papers, Census, Bureau of Labor Statistics (BLS), Bureau of Economic Analysis (BEA), and Federal Reserve series, workforce surveys, corporate filings and product documents, and institutional guidance. Six of the

TABLE I
PRISMA 2020 REPORTING ITEMS AND THEIR STATUS IN THIS REVIEW.

Item	Status in this manuscript
Objectives	Seven questions are stated.
Eligibility	Include and exclude rules are stated.
Information sources	Indexes that were queried, and those that were not, are named. The pre-8 November 2026 hit list was not archived.
Search strategy	Concept sets are stated. The 8 November 2026 strings, dates, and export counts are in Appendix A. Earlier strings were not archived.
Selection	One reviewer. No second screener and no adjudication log.
Data items	Venue, place, sample, period, AI type, outcomes, stated effects, and a causality class were recorded when the source stated them.
Risk of bias	Five-question checklist only. The Risk of Bias 2 tool (RoB 2), Risk Of Bias In Non-randomized Studies of Interventions (ROBINS-I), and Grading of Recommendations Assessment, Development and Evaluation (GRADE) were not applied.
Synthesis	Concept-centric. No pooling. Conflicts follow Table II.
Registration	None.

103 are author preprints used only as context. Seven further journal articles, added in this revision, are theory, review methodology, qualitative delegation evidence, a multilevel review, or one non-U.S. wellbeing association. Two Conference on Human Factors in Computing Systems (CHI) studies inform interaction overhead only. Two Control Objectives for Information and Related Technologies (COBIT) publications are governance design, not employee-outcome evidence. They are not added to the causal count. The 104 works are not 104 causal evaluations, and they are not 104 systematically screened studies. Two further sources, an Organisation for Economic Co-operation and Development (OECD) working paper and a 2026 ILO research brief, are cited only for measurement limits stated by their authors. They are not added to the causal count or to the 643-record screen. Eight carried-forward studies were classed as causal or strong quasi-causal workplace or labor-market designs. Three further designs are used for mechanism or task-time evidence: randomized developer experiments [10], a Danish difference-in-differences study used only as a mechanism [11], and a matched U.S. training comparison [12]. Further preprints by the first author, on agentic-system governance, enterprise generative-AI security, a governance-priority architecture, semantic tracing with a human-review gate, and delegated commercial agency, are cited later as capability, control, or boundary context. They are not outcome studies and are not added to the causal count. Related digital-transformation notes that restate the readiness preprint are not cited again. Surveys by AICPA and CIMA, NACD, Wharton Human-AI Research, and MIT Sloan Management Review with Boston Consulting Group, an ISACA audit proposal, and an MIT Sloan Executive Education program description are perceptions, guidance, or a curriculum statement. They are not evidence that AI improves financial performance. Five 2026 official sources close measurement gaps on postings, tasks, expectations, utilization, and announced cuts [38], [45]–[48].

B. Eligibility

English-language sources were eligible when all of the following held.

1) The source provided empirical, theoretical, or official-statistical evidence on AI and workforce change, employee outcomes, or enterprise performance.
2) The evidence was relevant to the U.S. window of November 2022 through September 2026, or it was foundational theory from an earlier period. International empirical studies were eligible only as mechanism evidence and were not used in U.S. totals.
3) The source was retrievable as a journal article, working paper, government publication, transparent survey, product-capability document, or official filing.

Sources were excluded when any of the following held: a news story was the sole support for a major conclusion; the source was vendor marketing of return on investment or job replacement; the document could not be retrieved; or the citation served only to lengthen the reference list. Two review-methodology articles are the stated exception to the workforce-evidence rule. They name the review type and the reporting table, and they are excluded from every outcome claim [39], [40]. Microsoft product documentation was eligible for capability and control description and was ineligible as proof of productivity, return on investment, displacement, satisfaction, or layoff effects. Incompatible time series were not spliced. Null and contradictory findings were in scope. The 2026 update was limited to Census, BLS, BEA, Federal Reserve, and gap-filling peer-reviewed or NBER evidence, plus the author preprints identified above, which are labeled as preprints.

C. Search and reproducibility

Search concepts combined an AI set (generative AI, enterprise AI, large language model, AI agent, Copilot), a workforce set (employment, layoff, job posting, training, workload, oversight), and an outcome set (job insecurity, burnout, engagement, technostress, training, organizational performance). Discovery used Google Scholar and publisher platforms. Verification required the original journal, working-paper, government, or corporate document. Sources consulted included publisher pages for Web of Science and Scopus listings, IEEE Xplore and the ACM Digital Library, ScienceDirect, SpringerLink, Wiley, Sage, Taylor & Francis, PubMed, the National Bureau of Economic Research (NBER), Census, BEA, BLS, the Federal Reserve, the Government Accountability Office (GAO), the National Institute of Standards and Technology (NIST), the Cybersecurity and Infrastructure Security Agency (CISA), the Office of Management and Budget (OMB), the U.S. Securities and Exchange Commission Electronic Data Gathering, Analysis, and Retrieval system (SEC EDGAR), and Pew, Gallup, and American Psychological Association (APA) pages. SSRN, arXiv, Research Square, and Preprints.org were used for emerging studies and are labeled as such.

Appendices A–H of this PDF record the retrospective protocol, the search that was actually exported, the PRISMA 2020 inventory, eligibility and screening, study characteristics, the

appraisal boundary, the synthesis by question, the HARWO notation, and the citation audit. The Boolean strings, export dates, and hit counts for the corpus compiled before 8 November 2026 were not archived. Scopus, Web of Science, IEEE Xplore, and the ACM Digital Library were not queried, because no institutional session for those indexes was available. On 8 November 2026 the public PubMed E-utilities interface was queried with the title/abstract string "generative artificial intelligence" or ChatGPT, and employment, workload, burnout, or layoff, and worker, employee, or firm, limited to publication dates 1 November 2022 through 30 September 2026. The query returned one record, a healthcare scoping review, PubMed identifier (PMID) 41037981. It was excluded on setting and was not added to the synthesis. An OpenAlex relevance count of 43,131 works was not exported and was not screened. A Crossref bibliographic query the same day returned 1,878,646 works and was not exported. Neither count is a screening denominator. Three OpenAlex exports were screened by one reviewer and are recorded in Appendix A. The first returned 68 works and added one cross-sectional study to the synthesis [49]. The second returned 90 works on human-in-the-loop task allocation and added two comparison papers, not outcome estimates [30], [31]. The third returned 3 works and added none. Those exports do not replace the earlier corpus. A separate protocol, written the same day and not registered, repeated the seven questions and the same eligibility rules. The reference-list classification below is not a claim that those works were retrieved by the new strings. Scopus, Web of Science, IEEE Xplore, and the ACM Digital Library remained unavailable. Four queries were exported in full and screened by one reviewer. OpenAlex title-and-abstract search for enterprise, generative, artificial intelligence, and employment, from 1 November 2022 through 30 September 2026, returned and exported 39 records. The same dates for generative artificial intelligence, employee, burnout, and workload returned and exported 2. A search for human, artificial intelligence, responsibility, accountability, and workflow, from 1 January 2018 through 30 September 2026, returned and exported 330. PubMed, with "generative artificial intelligence" or ChatGPT in the title or abstract and employment, workload, or burnout, for publication dates 1 November 2022 through 30 September 2026, returned and exported 272. The four files contain 643 records. One duplicate title was removed, and 642 titles were screened. Fifty-three titles remained after that pass. Abstracts were then requested from OpenAlex for all 53. Forty-two abstracts were returned and read. Eleven records had no abstract on OpenAlex. All eleven are SSRN working papers. Crossref returned an abstract for each on 8 November 2026, and those abstracts were read. None was added. One multi-country firm study reports employment interactions, but it is a working paper, the sample is not a U.S. payroll contrast, and the abstract coefficients were not taken from a re-read table (doi:10.2139/ssrn.7537464). One security-operations simulation reports a workload change; the setting is outside the enterprise question and the full text was not retrieved, so the figure was not used (doi:10.2139/ssrn.7512166). The other nine abstracts are conceptual, governance, or secondary syntheses without a primary enterprise sample. Full texts of the 53 were not retrieved, because no abstract met the inclusion rule for a new outcome study. None of the original 42 abstracts stated a checkable enterprise employee, workload, or organizational-performance design inside RQ1–RQ7. Education and clinical samples, conceptual essays, and syntheses without a primary sample were excluded. Two examples are a European public-administration monitoring case (doi:10.1007/s00146-026-02978-y) and a patent study of earnings calls from 2011 to 2021 (doi:10.1111/jpim.70015). This search added no study. All 102 works then in the reference list were classified from the stored bibliography and from the role already assigned in the manuscript. The classes separate studies used for an employee or performance estimate from theory, review methods, official series, surveys and filings, governance documents, and author preprints. Full texts of those 102 works were not downloaded again. The OECD working paper and the ILO brief were read from the issuing PDFs and were not part of that earlier sort. The classification is not a second full-text appraisal, the works were not given new RoB 2 or ROBINS-I domain scores, and they are not entered in the identification total of the new flow. They remain cited. No second reviewer repeated either screen. On 8 November 2026, eight public PDFs already cited in the bibliography were downloaded and are listed in Appendix E. RoB 2 was applied only to the randomized Copilot experiment, and only the randomization domain was given a concern rating. The other domains, and the other downloaded papers, were not scored. Journal PDFs without an open copy were not retrieved. The manuscript remains a critical integrative review. Two further queries, the same day, separated intensification without a layoff from mental health after a reduction. OpenAlex title-and-abstract search for generative artificial intelligence, employee, workload, and verification returned 1 record, a consumer-commerce preprint, which was excluded. The search for artificial intelligence, layoff, remaining employees, and burnout returned 2 records, a literature essay and a podcast summary, which were excluded. The matching PubMed strings returned 0. APA PsycInfo was not queried. Tables V and VI therefore use only evidence already in the corpus.

D. Appraisal and synthesis

Appraisal used five questions: venue class, design, geography, explicit numerators and time periods, and independence from the vendor whose product is evaluated. The questions are a transparency checklist. They do not generate a risk-of-bias score, and they are not a GRADE rating. An SCImago Journal Rank (SJR) quartile classifies the outlet. It does not score bias in an individual estimate. Working papers, outplacement reports, vendor studies, and author preprints were not labeled Q1. Table II is the hierarchy used when sources conflict, unless the higher-tier study is off-scope.

For each included source the reviewer recorded the bibliographic record, venue class, country, sample, period, AI type, workforce outcome, enterprise outcome, author-stated effect sizes where the source stated them, and a causality class. Synthesis is concept-centric and configurative [41]. Effect

TABLE II
EVIDENCE HIERARCHY USED IN THIS REVIEW.

Tier	Definition and use
A. Causal experiment	Randomized or naturally randomized assignment to AI access. Used for task and selected workplace time-use claims.
B. Strong quasi-causal	Difference-in-differences, staggered adoption, or a comparable design with a defined counterfactual.
C. Official statistics	Census, BLS, BEA, and Federal Reserve series. Used for adoption, employment, and productivity levels. Not AI-identified unless the agency study says so.
D. Longitudinal association	Repeated observation without identification.
E. Cross-sectional association	Single-wave correlation.
F. Self-reported perception	Survey attitudes and feared effects.
G. Corporate attribution	10-K statements, earnings claims, and named layoff reasons.
H. Product documentation	Microsoft Learn and similar. Capability and governance only.

sizes were not pooled because tasks, samples, and estimands are not commensurable. Each major relationship is classed as causal evidence, strong quasi-causal evidence, longitudinal association, cross-sectional association, self-reported perception, corporate attribution, descriptive co-occurrence, or insufficient evidence. A higher tier does not convert an off-scope estimate into a U.S. total. A corporate attribution does not become a causal employment effect because the count is large. The same record informs HARWO only as a list of which terms can be observed and which cannot. An unobserved duration, error rate, or cost is left missing. It is not imputed, and the resulting specification is not counted as a causal evaluation.

IV. RESULTS

A. Exposure context

Table III retains the 2025–2026 contrasts that later employee-outcome claims require as denominators. Earlier points on the same incompatible series, already used as decision context in the predecessor [1], are stated here so the path is not dropped. Business Trends and Outlook Survey (BTOS) firm use in goods or services production was 3.7 percent of firms in September 2023 and 5.4 percent in February 2024, with 6.6 percent then expected by fall 2024 [50]. The legacy production question, carried to September 2025, was about 10 percent use and 14 percent planned [37]. In late 2024, the Real-Time Population Survey recorded any generative-AI use by 45 percent of adults aged 18–64 and work use in the prior week by 27 percent of employed adults, of whom 10 percent used it every workday and 17 percent on some days; assisted hours were 1–7 percent of work hours and reported hours saved were 1.4 percent [36]. Those late-2024 figures were confirmed against the published *Management Science* abstract (doi:10.1287/mnsc.2025.02523). In 2026:Q2, the St. Louis Fed series of employed adults aged 18–64 who report using generative AI for their job was 45.199 percent, not seasonally adjusted [51]. That question is not the prior-week work-use measure reported for late 2024. A later vintage of the same series, updated 20 October 2026, records 40.545 percent for 2026:Q3. The Q2 figure is therefore not the latest published point, and the Q3 reading falls after this review's September 2026 close. Employment-weighted firm AI and large-language-model use in the Survey of Business Uncertainty were about 78 percent and 54 percent of employment at adopting firms [37], [52].

Series	2023–24	Late 2024	Nov. 2025	2026
BTOS firms	3.7–5.4%	∼10%	18%	—
RPS workers	—	27% prior week	∼41%	45.2% job use
Agents	Capability evidence; no prevalence series			

Fig. 1. U.S. AI measurement timeline, 2022/2023–2026. Worker, firm-weighted, employment-weighted, and agent series are not interchangeable. The 45.2 percent 2026 figure is a job-use question, not the late-2024 prior-week measure.

Historical corpus, assembled before 8 November 2026.
Identification and exclusion counts for this corpus were not archived and are not shown.

Records were checked against the eligibility rules. A news story, an unverified number, or vendor marketing of return or job replacement was not used as a major conclusion.

This figure is not the 8 November 2026 export. That export is Appendix B: 643 records, 1 duplicate removed, 642 titles screened, 589 excluded on title, 53 abstracts assessed, 0 studies added.
The reference list is the cited corpus. It is not the set of records identified by that export.

Fig. 2. Historical corpus. Counts for this path were not retained. The counted 8 November 2026 export is Appendix B and is not drawn here.

Classification: descriptive official-survey evidence. The November 2025 BTOS break is a measurement change, not an adoption shock [37], [46]. Worker use is widespread and shallow. Firm function use is lower and still narrow. Worker-task use can occur without formal firm adoption, and the reverse also occurs [38]. Census authors also note that a firm-level respondent may miss decentralized worker use, so task-level adoption can be understated [38].

B. Workforce restructuring

Job postings, hiring, employment, announced cuts, and executed separations are not interchangeable. This subsection distinguishes those objects. It does not estimate a causal payroll effect of generative AI.

1) Labor demand: Before generative AI, AI-exposed establishments reduced non-AI hiring while aggregate occupation- and industry-level employment effects were undetectable [16]. After ChatGPT, highly exposed freelancers had short-run employment and earnings reductions. In CESifo Working Paper 10601 the estimates are a 2 percent decline in the number of monthly jobs (standard error 0.004) and a 5.2 percent decline in monthly earnings (standard error 0.016) [54], [55]. Those two percentages were re-read from the working paper. They were not re-confirmed on the *Organization Science* PDF. The result is a spot-market contrast with less-affected freelancers, not a U.S. payroll census.

Liu and Webber link Lightcast postings to BTOS industry AI use and to firm-level AI or machine-learning skill requests, September 2023–November 2025, before the BTOS wording break. Industry- and firm-level coefficients are small once scaled by the low mean AI-posting share (1.6 percent of postings). Firm-level models include 1,053,991 firms. The

TABLE III
2025–2026 U.S. ADOPTION CONTRASTS USED AS DENOMINATORS FOR EMPLOYEE-OUTCOME CLAIMS. THESE ROWS ARE NOT ONE TIME SERIES.

Period	Estimate	Source
Nov. 2025	RPS work generative AI: 40.7 percent any use; 12 percent daily last week	[37]
Nov. 2025	BTOS after the question revision: about 18 percent of firms use AI in any business function; about 21 percent planned	[37]
Nov. 2025–Jan. 2026	BTOS supplement: 18 percent of firms; 32 percent employment-weighted; 22 percent expected within six months	[38]
Nov. 2025–Jan. 2026	Worker-task use: 23 percent of firms; 41 percent employment-weighted	[38]
Nov. 2025–Jan. 2026	Among adopters, 57 percent use AI in three or fewer functions; Sales/Marketing 52 percent, Strategy 45 percent, IT 41 percent	[38]
Very large firms, three sectors	Information, professional services, and finance: firm use 50–60 percent (60–70 percent employment-weighted)	[38]
Aug. 2025–May 2026	More than 80 percent of occupations have adoption above 20 percent; adoption exceeds 50 percent in only about 40 percent of occupations	[53]

authors treat those coefficients as precisely estimated nulls and state that the only statistically significant results are a small positive relationship between AI exposure and postings. They report no evidence in this window that industries or firms with higher AI adoption reduced overall job postings. They do not examine occupation-specific postings, they do not claim causality, and they describe the exercise as backward-looking monitoring [45]. Classification: longitudinal association and official monitoring. The result is not proof that AI does not reduce jobs. The firm count, the 1.6 percent share, and that conclusion were re-read from the 27 March 2026 Finance and Economics Discussion Series (FEDS) Note. The note's regression table was not re-extracted cell by cell, so a more specific lag claim is not restated.

2) Firm-reported headcount and task change: In the 2026 BTOS AI supplement, firm-weighted shares are 2.3 percent reporting an AI-related headcount increase and 2.0 percent reporting a decrease (4.3 percent reporting either change). Employment-weighted shares are 3.7 percent and 2.4 percent. The authors summarize the no-change share as 95 percent of AI-using firms and describe each tail as about 2 percent [38]. Capital or software replacement is more common (16 percent of AI-using firms). Summaries that the authors say have no causal interpretation find that functional breadth and operational investment are positively associated with reported employment decreases, whereas worker-task integration is not once those organizational measures are included [38]. The 2023–24 snapshot likewise found few reported employment reductions [50]. Classification: cross-sectional association and self-report.

The same supplement is the direct evidence for RQ1. Forty-four percent of AI-using firms report augmenting worker tasks. Among firms that report any task effect, 66 percent report exclusive augmentation. Pure substitution and pure task creation each occur in 5 percent of firms, and about 7 percent both augment and substitute. Among substituting firms, the share reporting a "large number" of replaced tasks rose from 2.5 percent to 7 percent since the first supplement [38]. Writing and editing is the leading generative-AI task (85 percent of firms with task-level use), followed by information search (50 percent). These percentages were re-read from CES-26-25. Classification: official descriptive evidence that augmentation is the modal reported mechanism; insufficient evidence for general U.S. job elimination.

3) Announced cuts and projections: Challenger, Gray & Christmas began coding AI as a standalone reason for announced U.S. job-cut plans in May 2023. Table IV records planned announcements and the employer's stated reason. It does not record completed separations. Contemporary news accounts of the May 2023 release reported 3,900 AI-cited cuts in that month. The May 2023 reasons table was not re-extracted here, news is not eligible as the sole support for a major conclusion, and 3,900 is a monthly figure rather than a May–December cumulative. It is therefore omitted from Table IV. The December 2025 Challenger report states that employers announced 1,206,374 U.S. job cuts in 2025, of which 54,836 were AI-cited, and that AI had been cited in 71,825 announcements since 2023 [56]. Those three figures were re-read from the year-end report. That report does not publish standalone full-year 2023 or 2024 AI totals. The arithmetic residual, $71{,}825 - 54{,}836 = 16{,}989$, is a combined 2023–2024 implication. It is not a Challenger-published annual split, and it is not entered in the table.

Through August 2026, total announced cuts were 529,914, down 41 percent from the same period of 2025, while AI remained the leading year-to-date stated reason at 116,175 announcements (about 22 percent). In August itself, AI was the fourth-stated reason, at 3,462, the lowest monthly AI total since December 2025 [48]. The May, June, July, and August 2026 AI totals were re-read from Challenger's own releases. Classification: corporate attribution. Rising attribution can coexist with the posting result and with the rarity of Census employment-decrease reports because the three series measure different objects.

BLS projections for 2024–34 anticipate a 33.5 percent increase for data scientists and a 15.8 percent increase for software developers [57]. Both percentages were re-read from the 16 July 2026 Economics Daily page. Projections are not realized outcomes.

TABLE IV
AI-CITED ANNOUNCED U.S. JOB-CUT PLANS (CHALLENGER).

Period	AI-cited announced cuts
Full year 2025	54,836; cumulative since 2023: 71,825. Total announced cuts: 1,206,374 [56]
YTD May 2026	87,714 (22 percent of 2026 announcements through May) [58]
YTD June 2026	101,743 (about 23 percent); cumulative since 2023: 173,568 [59]
YTD July 2026	112,713 (about 24 percent); cumulative since 2023: 184,538 [60]
August 2026	3,462 (fourth reason; lowest monthly AI total since Dec. 2025) [48]
YTD August 2026	116,175 (about 22 percent of 529,914); still the leading year-to-date reason [48]

These rows are corporate attributions, not completed separations and not causal identification. The May 2023 monthly figure of 3,900 and the reconstructed 2023–2024 residual are omitted because they are not primary annual totals.

C. Remaining employees

Direct evidence is separated here from theory and from hypotheses that the corpus cannot test.

1) Empirical findings: The revised Microsoft 365 Copilot working paper, which is vendor-affiliated, reports an intention-to-treat reduction of 1.4 Outlook hours per week (12 percent) and a local average treatment effect of about two hours (17 percent) [61]. The November 2025 revision of the working paper, which was re-read for this version, also states that more than 90 percent of treated workers used the tool at least once, while weekly use in the later window stabilized at just under 40 percent of treated workers. The NBER abstract compresses the user result as two fewer email hours among the 80 percent of treated workers who used the tool in the second half of the experiment, with less time outside regular hours. The same paper does not detect a change in the number of Teams meetings or Word documents, or in the quantity and composition of tasks. The authors note that major new responsibilities would be expected to leave a trace in those applications. That is an interpretation of telemetry, not a personnel-file measure of job scope [61], [62]. An earlier statement of a 3.6-hour (31 percent) email effect is not in the November 2025 revision and is withdrawn. Classification: causal evidence for email-time and off-hours changes among users; insufficient evidence for responsibility expansion.

A separate customer-service rollout, already cited by the predecessor for issues resolved per hour, also reports lower attrition, driven by retention of newer workers, and facilitation of worker learning [1], [63]. That is an employee-outcome result inside a productivity experiment. It is not a finding about workers who remain after a layoff.

In Denmark, employer and worker records show widespread chatbot initiatives in exposed occupations, worker-reported productivity benefits, and new tasks in content generation, oversight, and integration. Difference-in-differences estimates rule out earnings and recorded-hours effects larger than 2 percent two years after ChatGPT's launch. Too few workers move into higher-paying occupations for average earnings to change [11]. Classification: strong quasi-causal evidence for a null on Danish earnings and hours, with descriptive evidence of task reorganization, including oversight tasks. Not a U.S. total. The 2 percent bound was re-read from the NBER abstract and the March 2026 revision.

2) AI Interaction Overhead and Work Intensification Without Layoffs: The question in this subsection is whether enterprise AI reduces the employee's work or moves prompting, verification, correction, and supervision onto the person who still holds the job. No layoff is required for that move. Gross task-time savings and net employee time are separated. Neither CHI study measures a U.S. payroll workday, and neither is treated as proof that total hours rose.

Lee et al. surveyed 319 knowledge workers who used generative AI at least weekly and collected 936 task examples [64]. In most examples, participants rated cognitive activities as requiring less effort with generative AI than without it: 72 percent for knowledge and 55 percent for evaluation, the least reduced activity. Higher confidence in the AI was associated with less critical thinking and with larger perceived effort reductions. Higher confidence in one's own ability was associated with more critical thinking and with more effort when applying and evaluating responses. Qualitatively, the work of thinking shifted toward information verification, response integration, and task stewardship. The study is self-report. Several authors are affiliated with the vendor. It supports a change in the kind of effort. It does not measure hours, and it does not show that prompting increased total working time. A separate cross-sectional study of 276 corporate research-and-development employees reports that workload and organizational support were positively associated with generative-AI use, and use with innovative performance [49]. That association is not a U.S. payroll result, not a measure of net hours, and not evidence that a heavier workload was caused by the tool.

Fan et al. held one model fixed and varied the interface while 60 participants completed three Python tasks, with a no-AI control [65]. Pooled AI assistance reduced National Aeronautics and Space Administration Task Load Index (NASA-TLX) workload by 18.2 points and completion time by 22 percent (25.0 versus 32.1 minutes) and raised the odds of correctness (odds ratio 1.71). Inline help was fastest on simple work. Chat was more accurate as complexity rose, without a time penalty in that contrast. A verification-load index, built from failures, time to first compile, code churn, pauses, and switches, partially mediated increases in stress and fatigue across repeated tasks. The workload-reducing result and the verification-load result are both in the study. The first is a gross task-time and task-load saving inside a lab session. The second is a within-session cost of checking and repairing output. The design does not observe whether those minutes were added back elsewhere in a workday, and it does not involve a layoff.

COBIT 2019 names the organizational places where such residual work would have to be owned if a firm chose to govern it: managed human resources (APO07), managed risk (APO12), ensured resource optimization (EDM04), and managed performance (MEA01) [66], [67]. NIST Artificial Intelligence Risk Management Framework (AI RMF) 1.0 and the generative-AI profile likewise treat human oversight and accountability as governance functions [68], [69]. Those publications specify controls. They do not estimate prompting time, verification minutes, or strain.

Classification against P4: the lab experiment is consistent

TABLE V
RQ-A. EVIDENCE ON WORK INTENSITY WHEN NO LAYOFF IS REQUIRED. NO ROW MEASURES A U.S. WORKDAY AFTER A FIXED ROSTER.

Design	What it shows, and what it does not
Randomized lab, Fan et al.	Task time -22% and NASA-TLX -18.2; verification load tracks fatigue across tasks. Not a workday.
Self-report, Lee et al.	Effort often rated lower; activity shifts toward verification. Not hours. Vendor-affiliated.
Randomized email trial, Dillon et al.	Fewer Outlook hours. No change in meetings or document counts. Not net workload.
Danish difference-in-differences	Recorded hours and earnings effects larger than 2% are ruled out. New oversight tasks are descriptive. Not a U.S. total.
Cross-section, Li et al.	Workload and support associated with generative-AI use. Not a cause and not net hours.

with shorter task time coexisting with verification load that tracks fatigue; the survey is consistent with lower perceived cognitive effort and a shift toward verification. Total working hours, role ambiguity after adoption, and strain in firms that did not reduce headcount remain unmeasured. Frequent prompting is not interpreted here as an increase in the workday.

3) RQ-A. Intensification when headcount is unchanged: The question is whether adoption raises workload per employee, task volume, verification, correction, supervision, or performance expectations while the roster stays fixed. Table V keeps the designs apart. A shorter task and a lower task-load score can sit next to verification effort. Neither result is a workday total, and neither requires a layoff. The Danish hours bound is a null on recorded hours, not a finding that verification disappeared. The cross-sectional association of workload with generative-AI use does not identify which one came first.

The National Institute for Occupational Safety and Health, in a 10 April 2024 bulletin, defines work-related psychosocial hazards as factors in the work environment that can cause stress, strain, or interpersonal problems. Its summary of the 2018 General Social Survey Quality of Worklife supplement says that close to 30 percent of workers found their work stressful always or often, almost 70 percent agreed they had to work very fast, and 43 percent said job demand interfered with family life. Those figures are general occupational stress in 2018. They are not an AI-adoption estimate and not a measure of hours after a fixed roster [70].

The 2023–24 Census snapshot reported training, new workflows, and cloud purchases [50]. In the 2026 supplement, 64 percent of AI-using firms report no institutional adjustment. Training staff and developing new workflows are the most common adjustments, each reported by about 15 percent of AI-using firms [38]. Those shares were re-read from CES-26-25. They are firm reports of an adjustment, not hours absorbed by remaining employees. Establishments that adopted earlier AI changed skill requirements [16]. Within-occupation adoption is often below half of workers even where the occupation is exposed [53]. U.S. Workforce Innovation and Opportunity Act (WIOA) training spells from 2012–2024 show high quarterly earnings returns for trainees from low-AI-exposure occupations. Returns for trainees from high-exposure occupations rose from about \$1,000 per quarter before 2020 to about \$3,000 by 2022–2024, compared with matched workers who received only job-search assistance. The authors attribute most of that rise to moves into less AI-exposed occupations, and a smaller part to training in AI-complementary skills [12]. Classification: descriptive co-occurrence for Census training reports; matched comparison for WIOA earnings returns. Neither measures hours absorbed after an AI-related separation, and the training study is not a 2023–2026 generative-AI experiment. The about-\$1,000 and about-\$3,000 figures were re-read from the NBER abstract of the August 2026 revision. An earlier revision used a slightly different pre-2020 rounding, so the abstract's "about" is retained.

Product documentation for computer-using agents describes designated human supervisors, email or inline approval, pauses on timeout, and controls that can disable generative-AI publishing [71], [72]. A review of agentic and multi-agent systems finds the strongest capability evidence in software engineering, web automation, and computer use, and treats those systems as bounded and monitored rather than as independent workers [73].

A local prototype places a human-review gate after inventory, evidence, and governance steps. With semantic evaluation on, system latency was higher (about 126 seconds versus about 102 seconds) and decision-family correctness did not differ (59 of 160 versus 57 of 160). Human calibration and production integration remain pending [74]. That latency contrast is system time inside the prototype, not employee hours, production error rates, labor input, or firm profit.

4) Theoretical expectations: Downsizing-survivor research predicts intensification after cuts. Ghost-work and algorithmic-control research predict uncounted residual labor [24], [25]. NIST's AI Risk Management Framework treats human accountability as an organizational function [68]. These literatures motivate the construct. They are not AI-identified U.S. payroll evidence.

5) Unresolved hypothesis: AI responsibility transfer is defined here as the movement of tasks, accountability, or oversight from eliminated, unfilled, or automated roles onto remaining employees, including supervision of AI systems. No included U.S. study measures hours transferred from laid-off workers to remaining workers after an AI-associated reduction, and no included study measures oversight hours. The construct is not established.

6) RQ-B. Downsizing, remaining employees, and mental health: A general downsizing study is not evidence that an AI-related cut moved hours onto the people who stayed. Survivor and algorithmic-control research supply that context only [24], [25]. Burnout on a survey scale is not a clinical depression diagnosis. Table VI separates those objects. No included study estimates depressive symptoms, or a clinical anxiety diagnosis, after an AI-associated reduction. The Gallup contrasts are within-person associations with perceived displacement, not with a verified separation. The Finnish path is not a U.S. layoff contrast and is not significant as a direct adoption-to-wellbeing link.

The National Institute of Mental Health reports that 8.3 percent of U.S. adults, an estimated 21.0 million people, had a major depressive episode in 2021. The definition is a period of at least two weeks with depressed mood or loss of

TABLE VI
RQ-B. MENTAL-HEALTH AND REMAINING-EMPLOYEE EVIDENCE. GENERAL DOWNSIZING IS CONTEXT, NOT AN AI EFFECT.

Record	Class
Hours moved after an AI-associated cut	No included U.S. study.
Gallup panel	Displacement concern, not a measured layoff: engagement -0.20 SD, burnout $+0.18$ SD. Not clinical depression.
Finnish firm survey	No significant direct path from adoption to wellbeing. Not a U.S. layoff contrast [76].
Insecurity and technostress reviews	Pre-generative-AI theory. Not an AI-downsizing estimate [5], [6].
Announced-cut tallies	Employer attribution. Not completed separations and not a mental-health outcome.

interest and a majority of specified symptoms, from the 2021 National Survey on Drug Use and Health. That is a population prevalence. It is not a workplace sample, and it is not an estimate of depression after an AI-associated reduction [75]. NIOSH's 2024 bulletin links psychosocial hazards, including job insecurity and high demands, to burnout and depression in the occupational-health literature it reviews. That link is not an identification of an AI-related layoff [70]. Bureau of Labor Statistics productivity and hours series, which are Department of Labor statistics, do not isolate AI. No Department of Labor series in this review measures hours transferred to remaining employees after an AI-associated cut.

D. Job insecurity, well-being, and expectations

1) Empirical findings: Pew Research Center's report of 25 February 2025, from a survey of 5,273 employed U.S. adults fielded 7–13 October 2024, finds 52 percent worried about future workplace AI, 36 percent hopeful, 33 percent overwhelmed, and 32 percent expecting fewer long-run job opportunities; 6 percent expect more. Among workers who have used AI chatbots for work, 40 percent say the tools are extremely or very helpful for speed and 29 percent for quality [77]. Classification: self-reported perception. These percentages, the sample of 5,273, and the 7–13 October 2024 field dates were re-read from the 25 February 2025 report page. The canonical report URL is on Pew's social-trends site.

In Gallup panel data for 2023 through early 2026, workers who use AI daily or several times a week are more than twice as likely as infrequent users to fear job elimination within five years. Separately, within the same workers, reporting that the job was at least somewhat likely to be eliminated was associated with 0.20 standard deviations lower engagement, 0.18 higher burnout, 0.19 lower satisfaction, and a 4.1 percentage-point higher probability of active job search, relative to reporting no such concern. Gallup's elimination item refers to new technology, automation, or AI. Respect and perceived organizational care for well-being attenuate the association between frequent use and that fear [78]. These contrasts were re-read from the Gallup article on 8 November 2026. Classification: longitudinal association in a panel. Not a randomized effect of AI exposure, not a direct coefficient of frequent use on burnout, and not an effect of a layoff.

APA's 2025 Work in America survey, fielded by Harris Poll among 2,017 employed adults from 26 March to 4 April 2025, finds that 54 percent of workers say job insecurity has a significant impact on their stress [79]. The survey is not AI-identified. It is contextual occupational-health evidence. The 54 percent figure was re-read from the APA report.

Federal Reserve synthesis of the Atlanta Fed Survey of Business Uncertainty reports that senior leaders project stronger AI productivity gains and lower future employment than workers project [37]. That is a hierarchical perception gap, not a forecast error relative to realized total-factor productivity. Vendor language of substantive productivity gains is a corporate expectation [80], [81].

Highfill and Samuels compare expected and realized AI use, 2023–2026. Adoption was initially slower than expected (2023:Q4 expected relative growth of 19.6 percent versus realized 3.6 percent), then briefly faster (2025:Q1 expected 15.1 percent versus realized 34.9 percent), and by 2025:Q4 nearly aligned nationally (21.4 percent expected versus 21.5 percent actual in 2026:Q2). Forecasts for 2025:Q2–Q3 are not compared with later outcomes because the question changed. Finance and insurance still missed: 5.0 percent expected six-month growth as of 2025:Q4, and 15.5 percent realized by 2026:Q2 [46]. Classification: descriptive comparison of adoption forecasts. The same paper's link from 2016–2018 AI motivations to later industry production accounts is mixed and described by the authors as murky. It is not a test of 2023–2026 generative-AI return forecasts. The adoption percentages above were re-read from BEA Working Paper WP2026-16.

Table VII scores only dimensions with evidence in this corpus. Professional surveys add perception, not audited outcomes. AICPA and CIMA, the American Institute of Certified Public Accountants and the Chartered Institute of Management Accountants, surveyed 1,446 member finance and accounting leaders in August–September 2025, a global member sample rather than a U.S. payroll census. In the AICPA and CIMA release, 88 percent said artificial intelligence will be the most transformative technology trend for accounting and finance over the next 12–24 months, 8 percent said the organization is very well prepared, 21 percent said it is well prepared, and 56 percent identified generative AI as the most prominent skills gap [82]. The National Association of Corporate Directors (NACD) 2025 public-company director responses ($n = 146$) show that 62.33 percent had set aside full-board agenda time for AI, 52.74 percent had asked management how AI could affect future workforce needs, 36.30 percent had adopted an AI governance framework, 27.40 percent had written AI duties into committee charters, 8.90 percent had approved an annual budget for AI projects, and 6.16 percent had established metrics for management reporting [83]. Wharton Human-AI Research and GBK Collective surveyed about 800 U.S. enterprise leaders from 26 June to 11 July 2025: 43 percent saw a risk of declining skill proficiency as use rose, and 89 percent said generative AI enhances employees' skills [84]. Statements in that survey that leaders measure return are self-reports. They are not the independently measured financial return the predecessor did not find [1]. The AICPA percentages were re-read from the 17 December 2025 member release,

the NACD percentages from the 28 July 2025 public table, the Wharton percentages from the 2025 executive summary, and the MIT Sloan–BCG percentages cited below from the 18 November 2025 report page.

The supported statement in Table VII is that adoption forecasts improved while productivity, employment, and return expectations remain open. Autonomy and return on investment are not measured gaps.

2) What the attitude evidence does not show: Positive and negative poles coexist. Some workers report speed help, and a field experiment shows email-time savings [61], [77], [85]. Frequent users also report more fear. Within-person displacement concern tracks lower engagement and higher burnout [78]. No included U.S. study establishes a causal effect of enterprise AI adoption, or of an AI-related layoff, on burnout. The insecurity meta-analysis and the technostress literature predict that fear and new demands can harm attitudes even when objective job loss is limited [5], [6]. That prediction is theory, not a 2022–2026 estimate.

E. Organizational performance after adoption

Performance claims have to name the level (task, worker, establishment, firm, industry, or aggregate productivity) and the domain (operational, commercial, financial, human capital, or hidden cost). A faster writing task is not a firm productivity gain.

1) Boundary from task experiments: The predecessor already synthesized the writing, customer-service, consulting, and coding experiments [1], [63], [85]–[87]. Those speed and quality results are not re-listed here. The employee-relevant boundary is that quality can fall outside the capability frontier [86], and that less-experienced software developers had higher adoption and larger productivity gains in three randomized rollouts of a coding assistant. Combined across Microsoft, Accenture, and a Fortune 100 firm, and across 4,867 developers, completed tasks rose 26.08 percent (standard error 10.3 percent). The predecessor had already reported that percentage. Microsoft-affiliated coauthors and MIT GenAI Initiative funding mean the estimate is not an independent census of developers. A Wharton coauthor does not convert it into a firm-performance result [10]. A Microsoft technical report on early tools remains corporate research and is not independent confirmation [88]. The 26.08 percent figure, the standard error, the developer count, and the 27 February 2026 online publication date were re-read from the *Management Science* abstract (doi:10.1287/mnsc.2025.00535).

2) Aggregate and firm performance: In the private nonfarm business sector, BLS reports labor-productivity growth of 3.0 percent in 2024 and 2.2 percent in 2025, and total-factor-productivity growth of 1.5 percent and 0.8 percent [89]. In the nonfarm business sector in 2026:Q2, labor productivity rose 1.4 percent at an annual rate and 2.2 percent from four quarters earlier. From 2019:Q4 through 2026:Q2, growth has been 2.1 percent annualized, compared with 1.5 percent from 2007:Q4 through 2019:Q4 [90]. BLS does not isolate AI. The 2024 and 2025 rates were re-read from the total-factor-productivity release, and the 2026:Q2 rates from the 6 August 2026 productivity-and-costs release. A later revised productivity release, dated 3 September 2026, still reports the 1.4 percent quarterly and 2.2 percent year-ago labor-productivity rates. A calibration still implies modest decade-scale total-factor-productivity effects [91].

A BEA industry-account paper finds, in a baseline specification, productivity-enhancing and input-saving associations of AI intensity. An alternative timing specification is weaker and still labor-saving [92]. Classification: specification-sensitive association.

Makridis, Highfill, and Samuels combine Gallup frequent worker AI use, pooled 2025:Q2–2026:Q1 at the state-by-industry level, with industry BTOS employer use, and link both to 2016–2024 employment, earnings, and real output. With state-by-industry and industry-by-year fixed effects, cells with higher worker-reported frequent use show stronger post-2020 real-output paths. Employment differences are positive and imprecise. The industry-only BTOS specification does not recover the output pattern. The authors state that the estimates are descriptive: AI use is measured after the outcome window, and an occupation-exposure instrument is too weak to confirm the pattern [47]. Census regressions likewise associate integration breadth with commercial performance and, separately, with employment-decrease reports, without a causal interpretation [38]. Realized worker utilization and employer-reported adoption are not the same object. Neither result supports a claim that AI causes economic growth, and neither identifies the financial performance of firms that combined AI with workforce reduction.

3) Performance after workforce reduction: No included study assigns firms to AI with layoffs versus AI without layoffs and then measures profit, quality, rework, or turnover. The joint package is unidentified [38], [48]. Stock-price movements are not used.

Microsoft's Form 10-K reports approximately 228,000 full-time employees on 30 June 2025 and approximately 223,000 on 30 June 2026 [80], [93]. Earlier year-end counts were not re-read from the 2023 and 2024 filings in this pass and are not restated. Fiscal year 2025 (FY2025) revenue in the annual report was $281.7 billion, up 15 percent [81]. The same report's shareholder letter states that Azure revenue surpassed $75 billion and that Copilot-family products surpassed 100 million monthly active users. Those two product statements are management narrative, not a line in the audited headcount table. Headcount then declined by about 5,000 while the firm described AI as a productivity agenda [80], [93]. Classification: corporate co-occurrence. It does not identify Copilot or agents as the cause of hiring, attrition, or profit. No included independent study estimates U.S. Copilot Studio prevalence, displacement, error rates, or profit effects through September 2026.

TABLE VII
EXPECTATION–REALITY DIMENSIONS WITH EVIDENCE IN THIS CORPUS.

Dimension	Managerial or firm expectation	Realized evidence	Gap	Strength
Adoption	Six-month BTOS plans	Learning curve; near national alignment by late 2025	Narrowing	Moderate
Productivity	Leaders above workers; vendor claims of substantive gains	Task experiments yes; aggregate total-factor productivity not AI-identified	Open at firm and economy	Moderate
Employment reduction	Leaders more pessimistic than workers; AI-named cuts	Postings not generally reduced; firm headcount reports are a separate self-report	Open / mixed	Moderate
Task autonomy / agents	Product architecture for supervision	Oversight required on paper; no outcome census	Insufficient	Low
Enterprise return	Implied by the investment boom	Descriptive worker-use association; employer-use specification does not recover it	Insufficient	Low

F. Classified pathway

Fig. 3 and Table VIII collect the classes assigned above. Table IX records the same relationships with the confidence that the design supports. Neither table is a structural model.

Link	Evidence status
AI adoption	Supported as descriptive use
Task automation or augmentation	Mixed
Restructuring	Insufficient as a cause
Remaining-employee effects	Insufficient
Enterprise outcomes of AI plus cuts	Insufficient
Managerial expectations	Mixed
Organizational readiness	Insufficient

Fig. 3. Analytical pathway. Labels are evidence status, not causal identification. The path from restructuring to remaining employees, and from those employees to enterprise outcomes, is insufficient.

TABLE VIII
CLASSIFICATION OF THE POST-ADOPTION PATHWAY.

Link	Finding and class
Adoption → task change	Exclusive augmentation is the modal firm report. Causal for specified tasks; descriptive for firms.
Task change → postings	Aggregate postings not generally reduced; freelance demand down. Association and quasi-causal mix.
Task change → employment	Headcount reports and task-integration mode are cross-sectional and distinct from postings. Causal evidence insufficient.
Restructuring → load	Oversight is required in the architecture. Measured duty transfer is absent. Insufficient.
Overhead → net time	Task time and task load can fall while verification remains. Lab and self-report; workday hours unmeasured.
Employees → attitudes	Insecurity and engagement move with displacement concern; time savings are also reported. Panel and perception.
Expectations	Adoption forecasts improved. Productivity, employment, and return gaps remain. Descriptive.
AI + cuts → performance	Unidentified. Insufficient.

On RQ7, P2 and organizational consequence remain propositions. Fig. 4 locates the unmeasured links. HARWO in Section V writes those links as terms in a constrained allocation problem. The figure and the specification are organizing devices, not a confirmed map of enterprise transformation and not a computed policy.

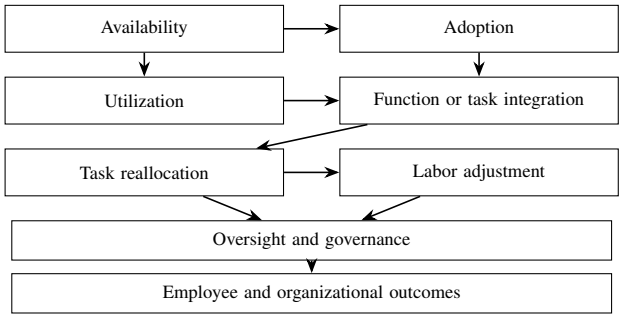


Fig. 4. Enterprise human–agent map. An organizing hypothesis, not a validated theory.
Expectations and readiness are candidate moderators. No arrow is an identified effect.

V. HUMAN–AI RESPONSIBILITY-AWARE WORKFLOW OPTIMIZATION

HARWO is a proposed computational reading of the responsibility-transfer framework. It asks how a workflow should assign work to a fixed set of people and to AI agents when completion time, human workload, operational risk, and cost are scored together. The human set X_H is an input, not a decision. A policy that removes an employee is outside the feasible set, so one optimization does not choose a layoff. A smaller roster can be supplied from outside the model and scored as a second fixed set. That comparison is two externally specified scenarios. It is not a search over headcount. HARWO is not an implemented solver, and no enterprise log was optimized.

A. Workflow graph

A case is one execution of an enterprise workflow. Precedence is an acyclic directed graph $G = (V, E)$. The node set splits into work tasks V_T and decision nodes V_D, with $V = V_T \cup V_D$. An edge $(u, v) \in E$ means v cannot start until u has completed, and no edge returns to an ancestor. A verification loop is not an edge of G. After task v emits an output, a gate may send that output back to v for another prompt, check, or correction, at most m_v times. The loop ends when the gate accepts the output or the count reaches m_v. If the count reaches m_v without acceptance, the case leaves v on the exception path, which is an edge of G and therefore

TABLE IX
EVIDENCE-GAP MAP FOR THE EMPLOYEE AND PERFORMANCE QUESTIONS.

Relationship	Best evidence and conclusion	Conf.
AI → task productivity	Experiments. Supported on specified tasks.	High
AI → task quality	Experiments. Heterogeneous.	High
AI → job postings	Liu–Webber monitoring. Not generally reduced.	Mod.
AI → employment	Census and market designs. Rare in firm reports.	Mod.
AI → announced layoffs	Challenger attribution. Rising stated reason, not a cause.	Mixed
Worker AI use → output	Makridis et al., descriptive positive association.	Mod.
Employer AI → output	Same study, BTOS specification. Not recovered.	Mod.
AI + layoffs → performance	No study in the corpus.	Low
AI → insecurity	Pew and Gallup. Supported association.	Mod.
AI → burnout	Gallup panel. Conditional on displacement concern, not a layoff or direct-use effect.	Low–mod.
AI → responsibility transfer	No U.S. hours study in the corpus.	Low
Agents → firm outcomes	Product documents and a preprint capability review only.	Low

cannot return to v. Actors are $X = X_H \cup X_A$, the humans who remain employed and the AI agents available to the workflow.

A policy is a triple $\pi = (e, g, a)$. The executor is $e(v) \in \text{Eligible}(v) \subseteq X$, the actor that performs task $v \in V_T$. The authorizer is $g(d) \in X_H \cup \{\varnothing\}$, the human who authorizes decision node $d \in V_D$, or $\varnothing$ if that node has no authorization. The accountable human is $a(v) \in X_H$, named for the output of task v whether or not that human executed v. Handoffs are edges whose two ends have different actors. Approval gates are decision nodes with $g(d) \in X_H$. Exception paths are edges taken when a gate rejects an output or a verification loop starts. Execution by an agent and accountability for the output are different assignments: $e(v)$ may lie in X_A while $a(v)$ remains in X_H.

B. Risk-aware allocation

Let $T(\pi)$ be the completion time of one case, $H(\pi)$ the human workload on that case, $R(\pi)$ the operational loss, and $C(\pi)$ the monetary cost. Human workload sums, over tasks and loop iterations, the time a person spends executing, prompting, verifying, correcting, supervising, and handling an exception. Those minutes stay in $H(\pi)$ whether or not they create an incremental cash expense. It is the interaction overhead of Section II-C written as a workflow sum. Monetary cost excludes the human minutes already in $H(\pi)$. It includes applicable model usage, licensing, and other incremental cash costs that do not repeat those minutes in the accounts. Prompting, verification, correction, supervision, and human time on an exception path are terms inside $H(\pi)$, not a second copy of completion time. $T(\pi)$ is the makespan of the case, so a shorter T can sit next to a larger H. Workload and cash are distinct decision criteria, so β and δ may both be positive. The separation is an accounting non-overlap.

The four expectations do not share a unit. A weighted sum is therefore applied only after each term is divided by a positive reference scale chosen before any comparison [28]. The scales T_{ref}, H_{ref}, R_{ref}, and C_{ref} are decision parameters. They are not estimated in this paper. With weights $\alpha, \beta, \gamma, \delta \geq 0$ and $\alpha + \beta + \gamma + \delta = 1$, the scalar problem is

$$\min_{\pi \in \Pi} \; \alpha\, \mathbb{E}[T(\pi)]/T_{\text{ref}} + \beta\, \mathbb{E}[H(\pi)]/H_{\text{ref}} + \gamma\, \mathbb{E}[R(\pi)]/R_{\text{ref}} + \delta\, \mathbb{E}[C(\pi)]/C_{\text{ref}}. \tag{1}$$

Equation (1) is the objective that is used. The same weights are not applied to the raw expectations, because those four expectations do not share a unit. Both β and δ may be positive when H and C are the distinct criteria just defined. The special case $C = a + bH$ has no cash component that varies separately from human workload. Its intercept does not change the minimizer, and its variable part can be absorbed into the human-time weight. That case is not a zero-weight restriction on the general model. If any expectation or any reference scale is unobserved, that policy is not scored. The missing term is reported as missing.

The feasible set Π requires all of the following.

1) Precedence: a node starts only after its predecessors in the acyclic graph complete. A verification return is the bounded counter above, not a predecessor edge, and it stops at m_v.
2) Eligibility: $e(v) \in \text{Eligible}(v)$.
3) Capacity: for each human h, assigned human time is at most the available time L_h. Headcount is not a decision variable.
4) Quality: for each task with a stated error tolerance q_v, the expected error after the chosen execution and gate is at most q_v.
5) Safety: every task in the mandatory set V_M has a human authorization before release.
6) Accountability: every released output has exactly one accountable human in X_H.

Mandatory membership is an organizational rule. NIST AI RMF 1.0 and the generative-AI profile describe governance, measurement, and human oversight for higher-impact uses [68], [69]. COBIT 2019 names resource, risk, human-resource, and performance objectives a firm can assign [66], [67]. Those documents do not set V_M, q_v, or the weights. Risk $R(\pi)$ is expected loss on outputs that pass the chosen gate. This paper does not supply a loss table.

C. Decision procedure

Fig. 5 is the procedure. It is a specification for a later implementation. It is not a tested program, and it has no optimality or runtime claim.

1) Build G from the workflow definition or from an event log [27]. Record tasks, dependencies, handoffs, gates, and bounded verification loops.

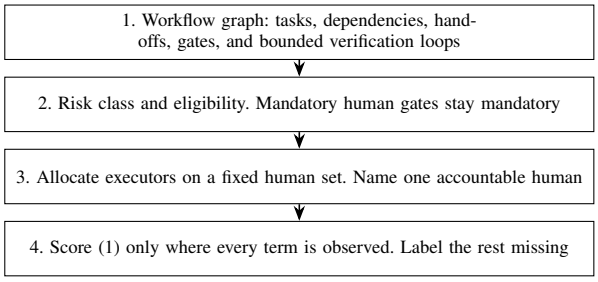


Fig. 5. HARWO decision specification. The diagram orders the procedure. It is not a computed policy or a proved optimum.

2) Mark V_M and $\mathrm{Eligible}(v)$ from the organization's safety and role rules [29], [68].
3) Hold X_H fixed. Assign each task an eligible actor and each mandatory gate a human authorizer. Attach one accountable human to every released output.
4) If durations, error rates, losses, and costs are observed, evaluate (1) on the feasible candidates that were constructed. Otherwise return those candidates with the missing terms labeled.
5) A person selects among the feasible scored policies. The procedure does not authorize deployment.

What HARWO addresses is a narrower gap than task allocation itself. Existing accounts can say who performs a step, how an operator's load should move a task, or that a person must remain able to correct an output. They do not, in the sources compared here, state a feasible policy on a fixed roster in which the executor, the authorizer, and the accountable human are different assignments, and in which prompting, verification, correction, and supervision are human time rather than a shorter case. Net human time is $H(\pi)$. A gross reduction in task duration is $T(\pi)$. Human workload stays in $H(\pi)$ even when those minutes have no incremental cash price, and $C(\pi)$ does not repeat them, so both weights may stay positive. Table X states the contrast from the retrieved abstracts and from the sources already used above. The gaps are what those accounts do not formulate. They are not a claim that the underlying graph, the weighted sum, or human-in-the-loop control is new. Table XI separates what the cited record already informs from what a later test would still have to measure. No such test is reported here. Appendix G records the full notation, the assumptions, and the same procedure.

VI. Discussion

The predecessor showed that boards lack a public evidence chain from AI cost to verified replacement and realized return [1]. This study classifies P1–P4. The record describes task change, the label employers put on announced cuts, and the security workers report when they use AI frequently. It leaves residual hours and post-cut performance unmeasured.

A. Answers to the research questions

On RQ1, firm reports describe a mixture, and exclusive augmentation is the modal task effect among firms that report any task effect (66 percent of that group; 44 percent of AI-using firms report augmentation). The Census supplement's headcount items are a separate self-report, reported with the other employment objects in the results, and those associations are cross-sectional.

On RQ2, three employment objects move differently. Aggregate postings in the Liu–Webber window are not generally lower where AI adoption is higher. A freelance market shows a short-run demand decline on working-paper magnitudes. Challenger's AI reason is a leading year-to-date label for announced plans through August 2026 even when it is not the leading reason in that month. The label is the employer's statement. It does not identify the separation.

On RQ3, the empirical residue is Census reports that about 15 percent of AI-using firms train staff and about 15 percent develop new workflows, while 64 percent report no institutional adjustment; earlier changes in skill requirements; WIOA earnings returns that mostly reflect moves out of exposed occupations; and vendor-affiliated telemetry of less email time without a change in meeting or document counts. Danish employer–employee records show new oversight tasks without a detectable earnings or hours effect. Agent products are documented to pause for human approval, and a preprint capability review describes computer-use agents as monitored systems [73]. None of that measures hours moved from separated U.S. workers onto those who stay. Responsibility transfer remains a hypothesis. The human–agent map in Fig. 4 shows where that hypothesis sits. It is not evidence that the transfer occurred.

On RQ4, worry about workplace AI is widespread in a national worker survey, and frequent use covaries with fear of elimination. Within the same workers, displacement concern tracks lower engagement, higher burnout, lower satisfaction, and a higher chance of active search. The association is not an effect of an identified layoff, and it is not a direct estimate of frequent use on burnout. The burnout meta-theory from the pre-generative-AI literature is not a substitute for that missing design.

On RQ5, the expectation gap is multidimensional. Near-term adoption forecasts moved toward realized use after an early period in which adoption lagged plans. Beliefs about productivity, employment, and financial return are not similarly closed. Leaders' greater pessimism about employment, relative to workers, is a perception gap. Director self-reports show more agenda time than workforce metrics [83].

On RQ6, specified tasks can be completed faster, and sometimes at higher quality inside a capability frontier, with quality losses outside it. That is a task result. Worker-reported frequent use is descriptively associated with stronger post-2020 real-output paths in one BEA design whose authors refuse a causal reading. Employer-reported adoption does not recover that pattern. No design compares performance between AI adopters that cut staff and AI adopters that did not. Aggregate labor productivity continued to rise and is not AI-identified. Task speed, a descriptive output association, and aggregate productivity are three different objects, and none is the financial performance of the AI-plus-reduction package.

On RQ7, P2 and organizational consequence remain propositions. Fig. 4 locates the unmeasured links. The figure is an organizing diagram, not a confirmed map of enterprise transformation.

TABLE X
RELATED-WORK COMPARISON. EACH GAP IS LIMITED TO WHAT THAT SOURCE, AS USED HERE, DOES NOT FORMULATE. HARWO IS A PROPOSED SPECIFICATION, NOT A SOLVED ALLOCATION.

Source	Approach	Human workload	Verification	Accountability	Gap relative to HARWO
Task allocation [2]	Who performs the task	Not a workflow sum	Not a loop cost	Not an assignment	Stops at the performer
Process mining [27]	Graph from an event log	Not scored	A loop is visible, not priced	Not named on release	No accountable human
Weighted sum [28]	Preference across units	Not defined	Not defined	Not defined	No roster or loss table
Human–AI guidelines [29]	Correction by a person	Not a capacity limit	Required, not timed	A person must be intelligible	Not a feasible policy
Technostress [5]	Strain from demands	Not minutes on a case	Not a workflow term	Not an assignment	No allocation
Call-center reallocation [30]	Move tasks from load and stress	Operator load is an input	Not separated from load	Operator stays in the loop	Abstract does not split authorization from accountability
Levels of automation [31]	Shared language for team design	Not a time sum	Not a cost	Named as a design requirement	Abstract has no verification term at a fixed roster
HARWO, this paper	Policy (e, g, a) on a graph	$H(\pi)$ at fixed X_H	Inside H, not inside T	$a(v)$ distinct from $e(v)$ and $g(d)$	Not solved or estimated

B. What the classification adds

The framework in Section II-C is the contribution. The predecessor asked whether a board can connect AI cost to verified replacement and realized return [1]. This paper classifies the four propositions against the same public record. It does not estimate a new return.

P1 is descriptively consistent with a separation between task reports and employment records, and it is not identified. The employment record is four results, set out in the workforce-restructuring results. Census respondents report headcount change on a different item from task change. Challenger records an employer's stated reason for an announcement, and that label became the leading 2026 year-to-date reason without becoming a separation count. Liu and Webber find no general reduction in postings where measured AI adoption is higher. Hui, Reshef, and Zhou find a 2 percent job decline and a 5.2 percent earnings decline in one online freelance market. Collapsing those four results into one employment effect misstates each of them.

TABLE XI
TRACE FROM EACH QUESTION TO A FUTURE MEASUREMENT. CLASSES ARE THOSE ALREADY IN THE CORPUS. THE LAST COLUMN IS NOT A RESULT OF THIS PAPER.

Question	Evidence now	HARWO term	Still untested
RQ1	Descriptive task mix	$e(v)$	Whether task change cuts jobs
RQ2	Attribution, not separations	X_H held fixed	Hours moved after a cut
RQ3	Oversight tasks observed	$g(d)$, $a(v)$	U.S. hours on those tasks
RQ4	Insecurity and burnout associations, not a layoff	Not in the scalar problem	A clinical outcome or a layoff effect
RQ5	Adoption forecasts narrowed; return and employment expectations open	Not a term	A forecast matched to a firm outcome
RQ6	No adopter comparison	$R(\pi)$, $C(\pi)$	Performance with and without cuts
RQ7	Proposition	Full policy π	A scored log
P1	Descriptive, not identified	Breadth is outside π	The conditional employment link
P2	Untested in U.S. payroll	$H(\pi)$ at fixed X_H	Oversight hours at unchanged headcount
P3	Panel association	Not in the scalar problem	A within-person displacement path
P4	Lab and self-report	Loop time inside $H(\pi)$	The same path on a workday

P2 is untested in U.S. payroll data. The mechanism predicts residual human work after task change even when headcount is unchanged [7], [8], [26]. Copilot telemetry shows less email time without a detected change in meeting or document counts. Danish records describe new oversight and integration tasks without a detectable earnings or hours effect. Product controls document approval pauses. None of those observations measures hours transferred onto U.S. workers who remain after a cut, so P2 stays a proposition.

P4 separates gross task time from net employee time when no layoff occurs. In a 60-person Python lab, AI assistance cut completion time by 22 percent and NASA-TLX workload by 18.2 points, while a verification-load index partially mediated rising stress and fatigue across repeated tasks [65]. In a survey of 319 knowledge workers, perceived cognitive effort was lower in most examples, and the reported activity shifted toward verification, integration, and stewardship [64]. The first result is a within-session saving. The second is a self-reported change in the kind of effort, from a study with vendor-affiliated authors. Neither measures the workday, so neither shows that prompting lengthened total hours. COBIT's human-resource, risk, resource-optimization, and performance objectives, and the NIST generative-AI profile, name where accountability for that checking would sit [66]–[69]. They do not measure it.

P3 is consistent with one U.S. panel and is not a layoff effect. Frequent use is associated with fear. Within person, displacement concern is associated with engagement and burnout. A Finnish managerial survey of 207 firms found no significant direct association between AI adoption and employee wellbeing (path estimate -0.111); reported associations ran through task optimization and safety [76]. That sample is outside the U.S. window. It shows why a direct adoption-to-wellbeing path is not assumed here. A multilevel review already separates collaboration, attitudes, algorithmic control, and labor-market effects [9]. The present framework adds the requirement that those themes be scored separately from organizational consequence.

Organizational consequence remains unidentified. No design compares AI adopters that reduced headcount with AI adopters that did not. Task speed, a descriptive output association, and aggregate productivity are different objects from

that comparison. The framework is therefore a set of defined constructs and testable propositions with a classified evidence status. HARWO restates P2 and P4 as a constrained allocation problem with human time kept visible when headcount is fixed. It is not a validated model. Collapsing P1–P4 into one adoption effect would misread both this paper and the predecessor [1].

The practical implication follows the same separation. In-frontier task gains are real in the experiments the predecessor reviewed, and out-of-frontier delegation has a measured quality risk [1], [86]. Headcount reductions that are justified by AI still require a task inventory and a mechanism [1]. Integration mode may matter more than a binary adopted-AI flag [38]. Fear among frequent users is an organizational variable with a panel association to displacement concern, and that concern has panel associations with engagement and search. It is not a proven consequence of layoffs [78].

C. Why the human consequences remain difficult to measure

The evidence gap is a mismatch between the outcomes that existing studies measure and the organizational consequences that remain of interest. The first three distinctions are limits the source authors state. The fourth is this review's comparison of U.S. sources that measure different objects.

Brynjolfsson, Li, and Raymond measure productivity in one customer-support deployment, as issues resolved per hour. They state that the findings are medium-run effects in a single firm and that the article is not designed to identify aggregate employment or wage effects. They also state that their data do not allow them to observe changes in wages, overall labor demand, or the skill composition of workers hired at the firm [63]. That is the authors' limit on one workplace study. It is not a finding that enterprise AI leaves employment unchanged. A separate randomized deployment records application-level time, including fewer Outlook hours, with no detected change in Teams meetings, Word documents, or the quantity and composition of tasks [61]. That telemetry is not a measure of net employee workload.

Milanez states, in an OECD working paper, that significant knowledge gaps remain in how firms, workers, and worker representatives adapt to AI. Nearly 100 case studies in manufacturing and finance, drawn from interviews in 2021 and 2022 across eight OECD countries, report job reorganization more often than displacement, skill requirements that were unchanged in many cases and higher in a significant portion, and frequent reports of increased work intensity [103]. The working paper does not present those reports as a U.S.-wide causal estimate for generative AI after November 2022, and this review does not use them as one.

An ILO research brief reports that productivity gains are real but uneven, and that worker-reported time savings have not yet translated into higher measured output, earnings, or employment [104]. That sentence distinguishes reported time savings from those measured outcomes. It is not evidence that AI has no organizational benefit.

The U.S. sources already in this review measure different objects. The Census supplement records firm-reported task effects and firm-reported headcount change as separate cross-sectional items, and its authors do not treat those associations as causal [38]. Challenger records the employer's stated reason for an announced job-cut plan, not a completed separation and not the work left to remaining employees [48]. A worker panel measures perceived displacement risk, engagement, and burnout on survey scales, and it does not link those scores to a verified AI-associated separation in the same organization [78]. Survey burnout is not a clinical diagnosis.

These sources therefore cannot, separately, establish whether AI-generated time savings reduce employees' overall workload, whether responsibilities move to workers who remain, or whether AI-related restructuring improves organizational performance. That comparison across studies is this review's inference. It is not an explanation, given by the cited authors, for why the missing observations were not collected. Testing responsibility transfer requires longitudinal observations that link AI adoption, employee task allocation, prompting and verification time, workforce changes, employee well-being, and organizational performance.

D. Implications for executives and boards

Organizations adopting AI should establish longitudinal, privacy-preserving measurement systems that capture not only task-level productivity but also changes in human workload, verification and oversight time, responsibility allocation, employee well-being, and organizational performance. These measurements should distinguish AI adoption from its actual use and separate AI-associated workforce changes from causally established displacement. Such evidence would enable more rigorous assessments of whether productivity gains translate into reduced employee workload or merely redistribute responsibilities. These recommendations follow from the documented measurement limitations in the reviewed literature and should not be interpreted as empirically validated benefits of the proposed HARWO framework.

These classifications are the decision content. Publications from AICPA and CIMA, NACD, Wharton Human-AI Research, MIT Sloan Management Review, MIT Sloan Executive Education, and ISACA describe readiness, oversight, education, or a proposed audit. They do not measure performance after a staff reduction. A survey, a curriculum, and a proposed method are not evidence that AI improves organizational performance.

For CFOs, the AICPA and CIMA member survey is a readiness and skills signal in a global sample of 1,446 finance and accounting leaders, not a return calculation [82]. Expected transformation does not identify value realization. The performance result in this paper is that task-speed gains and a descriptive output association do not measure the return from combining AI with workforce reduction.

For boards, the NACD responses show discussion running ahead of measurement: agenda time and a workforce question are more common than a reporting metric or a project budget [83]. That pattern matches the unmeasured oversight hours in this corpus. It is director self-report, not a workforce outcome.

For CEOs, a November 2025 MIT Sloan Management Review report with Boston Consulting Group, which BCG

sponsored, says that 76 percent of respondents in its global executive survey view agentic AI as more like a coworker than a tool. In the same survey, agentic-AI adoption is reported at 35 percent, with another 44 percent of organizations planning to deploy it, and the authors describe that adoption as running ahead of strategy [94]. The sample is not a U.S. payroll census, and the report does not identify performance after a staff reduction. Wharton Human-AI Research's U.S. leader survey records both a belief that generative AI enhances skills (89 percent) and concern about declining proficiency (43 percent) [84]. Those are perceptions. The employee evidence in this paper remains the fear and displacement-concern associations, and the absence of measured hours moved onto remaining workers.

MIT Sloan Executive Education's AI Executive Academy is a two-week program for senior executives, offered with the Schwarzman College of Computing and described as bridging AI technology and business leadership. Participants who complete it earn an executive certificate [95]. That page is a curriculum statement. It is not a finding about workforce transformation, employee well-being, or organizational performance, and completion of the program is not evidence that an AI investment has earned a return.

For CIOs and audit committees, the ISACA Journal article proposes an audit that identifies models, scopes risk, names stakeholders, and documents gaps [96]. That is an accountability method for executive decision-making. It does not measure whether employees absorbed extra work or whether the firm earned a return. This paper does not repeat the predecessor's investment test [1]. For HR leaders, the U.S. training comparison and the Danish records are the mechanism evidence: earnings returns and new oversight tasks can be observed without a U.S. demonstration that hours from separated workers landed on those who stayed [11], [12].

For workers, the near-term U.S. pattern in this corpus looks more like task reshuffling and a selected contract-market demand loss than like uniform job extinction [38], [45], [54]. Occupational projections remain forecasts [57].

Governance standards are relevant here only as oversight and workforce-risk context. The predecessor applied an ISACA decision lens to investment and headcount authorization [1]. This paper does not repeat that test and does not classify any firm. COBIT 2019 places managed human resources (APO07), managed risk (APO12), resource optimization (EDM04), and performance monitoring (MEA01) among the objectives a governance system can assign [66], [67]. Assigning an objective is not evidence that employees absorbed extra work.

A design-science architecture proposes platform, risk, maturity, and agentic-oversight components for prioritizing enterprise AI governance, and treats its scoring checks as internal robustness rather than evidence that a governed firm performs better or that employees work less [97]. NIST AI RMF 1.0 organizes trustworthy AI around govern, map, measure, and manage [68]. OMB Memorandum M-24-10, issued on 28 March 2024, required minimum practices for rights- and safety-impacting federal AI and directed agencies to designate Chief AI Officers [98]. A December 2023 GAO review, which predates that memorandum, found that 15 of 20 agency AI inventories contained incomplete or inaccurate information, and that 10 of 23 civilian Chief Financial Officers Act agencies had implemented all then-applicable agency-specific requirements, 12 had implemented some, and one was exempt [99]. OMB Memorandum M-25-21, issued on 3 April 2025, rescinds and replaces M-24-10. It requires a Chief AI Officer, an annual AI use-case inventory, and minimum risk-management practices for high-impact AI [100]. NIST AI 600-1 is a voluntary generative-AI profile and companion to AI RMF 1.0. It does not estimate employment or return on investment [69]. Secure-deployment guidance treats logging, least privilege, and incident response as baseline [101].

A review of enterprise generative-AI security reports that agentic tool use adds exposures, including prompt injection and excessive agency, and that cross-mapping NIST AI RMF with related security frameworks leaves gaps in multi-agent accountability [102]. Those requirements make oversight labor plausible. They do not measure it, and the security review is not independent confirmation of the workforce results above. Until oversight time and post-cut performance are observed, governance cost belongs on the unmeasured side of the performance ledger.

E. Limitations and open designs

The review limitations belong here. The original synthesis had one reviewer and no registered protocol. The historical Boolean strings, export dates, and hit counts were not archived, so the original yield cannot be reproduced and Fig. 2 has no identification or exclusion counts for that earlier corpus. The 8 November 2026 screens in Appendix A are title screens of exported filters, plus one earlier PubMed record and a later PubMed export of 272 records. They are not a multi-database systematic review. The additional four-query export added no included study. A same-day follow-up retrieved Crossref abstracts for the eleven SSRN records that OpenAlex had left without an abstract. All eleven were excluded on the abstract, and no full text from that set was retrieved. The unexported counts of 43,131 and 1,878,646 were not screened. RoB 2, ROBINS-I, and GRADE were not scored domain by domain. Appendix E states that limit. A later version can become a systematic review only if the protocol and the seven questions are registered first, native database exports are archived, a second reviewer screens the same records, and RoB 2 is reserved for the randomized task experiments and ROBINS-I for the quasi-causal labor studies. Surveys, filings, and product documents stay outside those instruments. Effect sizes stay unpooled until the task, population, and contrast match. Table I and Appendix B record the present status. They do not change the review class.

Stelmaszak et al. supply a qualitative platform case, Valtonen et al. a Finnish managerial model, and Bankins et al. a multilevel review that mixes countries and designs [8], [9], [76]. They define constructs for P2 and P3. They do not test those propositions on U.S. payroll data.

Some magnitudes remain limited to the document that was opened. The freelance 2 percent and 5.2 percent figures were re-read from CESifo Working Paper 10601 and

were not re-confirmed on the *Organization Science* PDF. Liu and Webber's firm count, 1.6 percent AI-posting share, and "precisely estimated null" conclusion were re-read from the FEDS Note; the regression table was not re-extracted cell by cell. Microsoft headcount was re-read for 30 June 2025 (about 228,000) and 30 June 2026 (about 223,000). The 2023 and 2024 Form 10-K headcounts were not re-read and are not reported. A May 2023 Challenger reasons table was not re-extracted, so the contemporary 3,900 monthly figure stays out of Table IV. Azure revenue and Copilot-family user counts remain statements in the FY2025 shareholder letter. The 2026:Q3 Federal Reserve Economic Data (FRED) reading of 40.545 percent was published after this review's September 2026 close. International studies are not used as U.S. totals.

The following claims were re-read from the issuing page or PDF and kept: Pew's worker percentages and sample; APA's 54 percent; the AICPA release percentages, with the unsupported 39 percent preparedness figure removed; the NACD table; the Wharton 43 and 89 percent figures; the MIT Sloan–BCG 76, 35, and 44 percent figures; BLS productivity rates for 2024, 2025, and 2026:Q2; the BLS 33.5 and 15.8 percent projections; the 2026:Q2 FRED value of 45.199 percent; the 26.08 percent developer-task estimate; the Copilot 1.4-hour and two-hour email estimates; the WIOA "about $1,000" and "about $3,000" abstract figures; the Danish 2 percent bound; the BEA expectation percentages; the GAO 15-of-20 and 10-of-23 counts; the Census supplement shares used for RQ1; and the May through August 2026 Challenger AI totals.

Author preprints are context for readiness and control. They are not peer-reviewed replications. Duplicate postings of the same preprint were excluded.

HARWO inherits the same limits. Reference scales, weights, error tolerances, and the mandatory-gate set are not identified. No log was scored, so (1) has no numerical value in this paper. A later study can test the specification only with an observed workflow graph, measured human time that includes prompting and verification, an error or loss record, and a fixed human roster. Until then the procedure remains a proposal.

P2 and the organizational-consequence construct require designs that are still absent: matched employer–employee panels linking tool use, hours, occupation, and separations; time-use studies of human supervisors; audits that separate AI-named announcements from demand, cost, and merger shocks; performance comparisons of AI adopters with and without headcount reduction that are independent of the vendor; and a repeated national series for enterprise-agent adoption. The predecessor's calls for audited cost-and-return panels remain open [1].

VII. Conclusion

The responsibility-transfer framework is the contribution beyond *Enterprise AI at the Decision Point*. Automation and augmentation allocate tasks. Technostress and insecurity explain strain. Human–AI collaboration explains an interaction that still depends on people. The framework defines residual responsibility and organizational consequence as separate constructs and states P1–P4 so that later work can test them. HARWO is the proposed workflow form of that claim: a directed graph, a human-time term at fixed headcount, a risk term, and one accountable human on every release. Execution, authorization, and residual accountability are different assignments. The scalar problem is a preference statement with normalized units. It is not a solved program, not a claim that weighted-sum optimization is new, and not an estimated enterprise result.

The U.S. record through September 2026 classifies those propositions. P1 is consistent with a separation between how firms describe task change and how employment is recorded. Firms that report a task effect usually report augmentation, and an AI label on an announced cut is an attribution. P3 is consistent with one panel: frequent use tracks fear, and displacement concern tracks engagement, survey burnout, satisfaction, and search. That association is not a clinical diagnosis and is not linked to a verified separation. P2, interaction overhead inside jobs that continue, and the performance of AI combined with headcount reduction, are unmeasured as workday outcomes. A shorter task time is not, by itself, a shorter workday, and an AI label on an announcement is not a completed separation. Until residual hours, net of verification, and that performance comparison are observed, the framework remains a theoretical and classificatory contribution rather than a validated account of enterprise transformation. Organizations need longitudinal, privacy-preserving data linking AI adoption, changes in human workload, verification and oversight time, workforce changes, employee well-being, and organizational performance before those consequences can be assessed.

References

[1] Rah, A.; Johnson, D.W. Enterprise AI at the Decision Point: A U.S.-Focused Systematic Review of Productivity, ROI, Workforce Change, and Governance. SSRN 2026. Preprint, not peer reviewed. https://doi.org/10.2139/ssrn.7517919

[2] Autor, D.H. Why Are There Still So Many Jobs? The History and Future of Workplace Automation. J. Econ. Perspect. 2015, 29(3), 3–30.

[3] Acemoglu, D.; Restrepo, P. Automation and New Tasks: How Technology Displaces and Reinstates Labor. J. Econ. Perspect. 2019, 33(2), 3–30. https://doi.org/10.1257/jep.33.2.3

[4] Raisch, S.; Krakowski, S. Artificial Intelligence and Management: The Automation–Augmentation Paradox. Acad. Manage. Rev. 2021, 46(1), 192–210.

[5] Tarafdar, M.; Tu, Q.; Ragu-Nathan, B.S.; Ragu-Nathan, T.S. The Impact of Technostress on Role Stress and Productivity. J. Manage. Inf. Syst. 2007, 24(1), 301–328.

[6] Sverke, M.; Hellgren, J.; Naswall, K. No Security: A Meta-Analysis and Review of Job Insecurity and Its Consequences. J. Occup. Health Psychol. 2002, 7(3), 242–264.

[7] Anthony, C.; Bechky, B.A.; Fayard, A-L. "Collaborating" with AI: Taking a System View to Explore the Future of Work. Org. Sci. 2023, 34(5), 1672–1694. https://doi.org/10.1287/orsc.2022.1651

[8] Stelmaszak, M.; Mohlmann, M.; Sorensen, C. When Algorithms Delegate to Humans: Exploring Human-Algorithm Interaction at Uber. MIS Quart. 2025, 49(1), 305–330. https://doi.org/10.25300/MISQ/2024/17911

[9] Bankins, S.; Ocampo, A.C.; Marrone, M.; Restubog, S.L.D.; Woo, S.E. A Multilevel Review of Artificial Intelligence in Organizations: Implications for Organizational Behavior Research and Practice. J. Organ. Behav. 2024, 45(2), 159–182. https://doi.org/10.1002/job.2735

[10] Cui, Z.; Demirer, M.; Jaffe, S.; Musolff, L.; Peng, S.; Salz, T. The Effects of Generative AI on High-Skilled Work: Evidence from Three Field Experiments with Software Developers. Manage. Sci. 2026. https://doi.org/10.1287/mnsc.2025.00535

[11] Humlum, A.; Vestergaard, E. Still Waters, Rapid Currents: Early Labor Market Transformation under Generative AI. National Bureau of Economic Research 2025, 33777.

[12] Hyman, B.G.; Lahey, B.; Ni, K.; Pilossoph, L. How Retrainable Are AI-Exposed Workers?. National Bureau of Economic Research 2025, 34174. https://doi.org/10.3386/w34174

[13] Autor, D.H.; Levy, F.; Murnane, R.J. The Skill Content of Recent Technological Change: An Empirical Exploration. Quart. J. Econ. 2003, 118(4), 1279–1333.

[14] Acemoglu, D.; Kong, F.; Restrepo, P. Tasks at Work: Comparative Advantage, Technology and Labor Demand. National Bureau of Economic Research 2024, 32872. https://www.nber.org/papers/w32872

[15] Acemoglu, D.; Restrepo, P. Robots and Jobs: Evidence from US Labor Markets. J. Polit. Econ. 2020, 128(6), 2188–2244.

[16] Acemoglu, D.; Autor, D.; Hazell, J.; Restrepo, P. Artificial Intelligence and Jobs: Evidence from Online Vacancies. J. Labor Econ. 2022, 40(S1), S293–S340. https://doi.org/10.1086/718327

[17] Eloundou, T.; Manning, S.; Mishkin, P.; Rock, D. GPTs Are GPTs: Labor Market Impact Potential of LLMs. Science 2024, 384(6702), 1306–1308. https://doi.org/10.1126/science.adj0998

[18] Gmyrek, P.; Berg, J.; Kami'nski, K.; Konopczy'nski, F.; Ladna, A.; Nafradi, B.; Roslaniec, K.; Troszy'nski, M. Generative AI and Jobs: A Refined Global Index of Occupational Exposure. International Labour Organization 2025, 140. https://doi.org/10.54394/HETP0387

[19] Felten, E.; Raj, M.; Seamans, R. Occupational, Industry, and Geographic Exposure to Artificial Intelligence: A Novel Dataset and Its Potential Uses. Strategic Manage. J. 2021, 42(12), 2195–2217.

[20] Orlikowski, W.J. The Duality of Technology: Rethinking the Concept of Technology in Organizations. Org. Sci. 1992, 3(3), 398–427.

[21] Brynjolfsson, E.; Hitt, L.M. Beyond Computation: Information Technology, Organizational Transformation and Business Performance. J. Econ. Perspect. 2000, 14(4), 23–48.

[22] Venkatesh, V.; Morris, M.G.; Davis, G.B.; Davis, F.D. User Acceptance of Information Technology: Toward a Unified View. MIS Quart. 2003, 27(3), 425–478.

[23] Zuboff, S. In the Age of the Smart Machine: The Future of Work and Power. Basic Books 1988.

[24] Gray, M.L.; Suri, S. Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass. Houghton Mifflin Harcourt 2019.

[25] Kellogg, K.C.; Valentine, M.A.; Christin, A. Algorithms at Work: The New Contested Terrain of Control. Acad. Manage. Ann. 2020, 14(1), 366–410.

[26] Berente, N.; Gu, B.; Recker, J.; Santhanam, R. Managing Artificial Intelligence. MIS Quart. 2021, 45(3), 1433–1450. https://doi.org/10.25300/MISQ/2021/16274

[27] van der Aalst, W.M.P. Process Mining: Data Science in Action. 2 ed.; Springer 2016. https://doi.org/10.1007/978-3-662-49851-4

[28] Marler, R.T.; Arora, J.S. Survey of Multi-Objective Optimization Methods for Engineering. Struct. Multidiscip. Optim. 2004, 26, 369–395. https://doi.org/10.1007/s00158-003-0368-6

[29] Amershi, S.; Weld, D.; Vorvoreanu, M.; Fourney, A.; Nushi, B.; Collisson, P.; Suh, J.; Iqbal, S.; Bennett, P.N.; Inkpen, K.; Teevan, J.; Kikin-Gil, R.; Horvitz, E. Guidelines for Human-AI Interaction. Proc. CHI Conf. Human Factors Comput. Syst. 2019, 1–13. https://doi.org/10.1145/3290605.3300233

[30] Donat-Bouillud, L.; Amokrane-Ferka, K. Adaptable and Hybrid Automation for Human–AI Task Allocation: Application to Call Center Supervision. Electronics 2026, 15(11), 2452. https://doi.org/10.3390/electronics15112452

[31] van Tuijn, R.; van Dijk, J.; Bootsma, R.; Holleman, B.; Neerincx, M. Designing by Degree: Levels of Automation and Team Design Patterns as a Shared Language. AHFE International 2026, 237. https://doi.org/10.54941/ahfe1008221

[32] Weiner, B.J. A Theory of Organizational Readiness for Change. Implementation Sci. 2009, 4(1), 67.

[33] Rah, A. Digital Transformation Beyond Technology: Governance, Human Resistance, and Cybersecurity. Preprints 2026, 2026051423. https://doi.org/10.20944/preprints202605.1423.v1

[34] Dietvorst, B.J.; Simmons, J.P.; Massey, C. Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err. J. Exp. Psychol. Gen. 2015, 144(1), 114–126.

[35] Rah, A. Agent-to-Agent Business Ecosystems: Delegated Commercial Agency, Governance, and a Reference Architecture for Autonomous Enterprise Commerce. SSRN 2026. Preprint, not peer reviewed. https://doi.org/10.2139/ssrn.7125338

[36] Bick, A.; Blandin, A.; Deming, D.J. The Rapid Adoption of Generative AI. Manage. Sci. 2026, 72(10), 8440–8456. https://doi.org/10.1287/mnsc.2025.02523

[37] Allen, J.S. Monitoring AI Adoption in the U.S. Economy. FEDS Notes, Board of Governors of the Federal Reserve System 2026. https://www.federalreserve.gov/econres/notes/feds-notes/monitoring-ai-adoption-in-the-u-s-economy-20260403.html

[38] Bonney, K.; Breaux, C.; Dinlersoz, E.; Foster, L.; Haltiwanger, J.; Pande, A. The Microstructure of AI Diffusion: Evidence from Firms, Business Functions, and Worker Tasks. U.S. Census Bureau, Center for Economic Studies 2026, CES-26-25. https://www2.census.gov/ces/wp/2026/CES-WP-26-25.pdf

[39] Pare, G.; Trudel, M-C.; Jaana, M.; Kitsiou, S. Synthesizing Information Systems Knowledge: A Typology of Literature Reviews. Inf. Manage. 2015, 52(2), 183–199. https://doi.org/10.1016/j.im.2014.08.008

[40] Snyder, H. Literature Review as a Research Methodology: An Overview and Guidelines. J. Bus. Res. 2019, 104, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

[41] Webster, J.; Watson, R.T. Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quart. 2002, 26(2), xiii–xxiii.

[42] Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; Chou, R.; Glanville, J.; Grimshaw, J.M.; Hr'objartsson, A.; Lalu, M.M.; Li, T.; Loder, E.W.; Mayo-Wilson, E.; McDonald, S.; McGuinness, L.A.; Stewart, L.A.; Thomas, J.; Tricco, A.C.; Welch, V.A.; Whiting, P.; Moher, D. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. BMJ 2021, 372, n71.

[43] Kitchenham, B.; Charters, S. Guidelines for Performing Systematic Literature Reviews in Software Engineering. Keele University and University of Durham 2007, EBSE-2007-01.

[44] Okoli, C. A Guide to Conducting a Standalone Systematic Literature Review. Commun. Assoc. Inf. Syst. 2015, 37, 879–910.

[45] Liu, J.; Webber, D. AI Adoption and Firms' Job-Posting Behavior. FEDS Notes, Board of Governors of the Federal Reserve System 2026. https://www.federalreserve.gov/econres/notes/feds-notes/ai-adoption-and-firms-job-posting-behavior-20260327.html

[46] Highfill, T.; Samuels, J.D. AI Expectations and Outcomes. U.S. Bureau of Economic Analysis 2026, WP2026-16. https://doi.org/10.66137/ESDV4288

[47] Makridis, C.A.; Highfill, T.; Samuels, J.D. AI Utilization and Economic Performance. U.S. Bureau of Economic Analysis 2026, WP2026-18. https://www.bea.gov/sites/default/files/papers/bea-wp2026-18.pdf

[48] Challenger, Gray & Christmas. Challenger Report: August 2026. Challenger, Gray & Christmas, Inc. 2026. https://www.challengergray.com/wp-content/uploads/2026/09/Challenger-Report-August-2026.pdf

[49] Li, Z.; Sun, H.; Yang, T.; Zeng, K. Generative AI in the workplace: how job demands and resources influence employee innovative performance – a JD-R theory perspective. European Journal of Innovation Management 2026. https://doi.org/10.1108/ejim-11-2025-1496

[50] Bonney, K.; Breaux, C.; Buffington, C.; Dinlersoz, E.; Foster, L.; Goldschlag, N.; Haltiwanger, J.; Kroff, Z.; Savage, K. Tracking Firm Use of AI in Real Time: A Snapshot from the Business Trends and Outlook Survey. U.S. Census Bureau, Center for Economic Studies 2024, CES-24-16R. https://www.census.gov/library/working-papers/2024/adrm/CES-WP-24-16.html

[51] Bick, A.; Blandin, A.; Deming, D. Generative Artificial Intelligence, Adoption Rate for Work: Employed Adults (RPSGENAIUSAGESHAREWORK). FRED, Federal Reserve Bank of St. Louis 2026. https://alfred.stlouisfed.org/series?seid=RPSGENAIUSAGESHAREWORK

[52] Altig, D.; Barrero, J.M.; Bloom, N.; Davis, S.J.; Meyer, B.; Parker, N. Surveying Business Uncertainty. J. Econometrics 2022, 231(1), 282–303.

[53] Bick, A.; Blandin, A.; Deming, D.J.; Schumacher, T.R. What Work Does Generative AI Do?. National Bureau of Economic Research 2026, 35677. https://www.nber.org/papers/w35677

[54] Hui, X.; Reshef, O.; Zhou, L. The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market. Org. Sci. 2024, 35(6), 1977–1989.

[55] Hui, X.; Reshef, O.; Zhou, L. The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market. CESifo 2023, 10601. https://www.ifo.de/DocDL/cesifo1_wp10601.pdf

[56] Challenger, Gray & Christmas. 2025 Year-End Challenger Report. Challenger, Gray & Christmas, Inc. 2026. https://www.challengergray.com/blog/2025-year-end-challenger-report-highest-q4-layoffs-since-2008-lowest-ytd-hiring-since-2010/

[57] U.S. Bureau of Labor Statistics. Artificial Intelligence, Information Technology, and Employment, 2024–34. U.S. Department of Labor 2026. https://www.bls.gov/opub/ted/2026/artificial-intelligence-information-technology-and-employment-2024-34.htm

[58] Challenger, Gray & Christmas. Challenger Report: May 2026. Challenger, Gray & Christmas, Inc. 2026. https://www.challengergray.com/wp-content/uploads/2026/06/Challenger-Report-May-2026.pdf

[59] Challenger, Gray & Christmas. Challenger Report: June 2026. Challenger, Gray & Christmas, Inc. 2026. https://www.challengergray.com/wp-content/uploads/2026/07/Challenger-Report-June2600986996.pdf

[60] Challenger, Gray & Christmas. Challenger Report: Layoffs Fall, Hiring Picks Up; AI Leads for Fifth Straight Month. Challenger, Gray & Christmas, Inc. 2026. https://www.challengergray.com/blog/challenger-report-layoffs-fall-hiring-picks-up-ai-leads-for-fifth-straight-month/

[61] Dillon, E.W.; Jaffe, S.; Immorlica, N.; Stanton, C.T. Shifting Work Patterns with Generative AI. National Bureau of Economic Research 2025, 33795. https://www.nber.org/papers/w33795

[62] Dillon, E.; Jaffe, S.; Peng, S.; Cambon, A. Early Impacts of M365 Copilot. Microsoft Research 2025, MSR-TR-2025-18. https://www.microsoft.com/en-us/research/publication/early-impacts-of-m365-copilot/

[63] Brynjolfsson, E.; Li, D.; Raymond, L. Generative AI at Work. Quart. J. Econ. 2025, 140(2), 889–942. https://doi.org/10.1093/qje/qjae044

[64] Lee, H-P.; Sarkar, A.; Tankelevitch, L.; Drosos, I.; Rintel, S.; Banks, R.; Wilson, N. The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects from a Survey of Knowledge Workers. Proc. CHI Conf. Human Factors Comput. Syst. 2025, 1–23. https://doi.org/10.1145/3706598.3713778

[65] Fan, G.; Liu, D.; Pan, L.; Zhang, R. When Help Hurts: Verification Load and Fatigue with AI Coding Assistants. Proc. CHI Conf. Human Factors Comput. Syst. 2026. https://doi.org/10.1145/3772318.3791176

[66] ISACA. COBIT 2019 Framework: Governance and Management Objectives. ISACA 2018.

[67] ISACA. COBIT 2019 Framework: Introduction and Methodology. ISACA 2018.

[68] Tabassi, E. Artificial Intelligence Risk Management Framework (AI RMF 1.0). National Institute of Standards and Technology 2023, NIST AI 100-1. https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf

[69] National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. National Institute of Standards and Technology 2024, NIST AI 600-1. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf

[70] National Institute for Occupational Safety and Health. An Urgent Call to Address Work-Related Psychosocial Hazards and Improve Worker Well-Being. NIOSH Science Bulletin, Centers for Disease Control and Prevention 2024. https://www.cdc.gov/niosh/bulletin/2024/workplace-psychosocial-hazards.html

[71] Microsoft. Human Supervision for Computer Use. Microsoft Copilot Studio documentation 2026. https://learn.microsoft.com/en-us/microsoft-copilot-studio/human-supervision-computer-use

[72] Microsoft. Security and Governance. Microsoft Copilot Studio documentation 2026. https://learn.microsoft.com/en-us/microsoft-copilot-studio/security-and-governance

[73] Rah, A. Agentic and Multi-Agent Systems: A Systematic Review of Tool Use, Benchmarks, and Governance. Research Square 2026. Preprint, not peer reviewed. https://doi.org/10.21203/rs.3.rs-9839978/v1

[74] Rah, A. Semantic Tracing in LLM-Based Multi-Agent Systems Using LangChain, LangGraph, and LangSmith for AI Governance. Research Square 2026. Preprint, not peer reviewed. https://doi.org/10.21203/rs.3.rs-10485157/v1

[75] National Institute of Mental Health. Major Depression. National Institute of Mental Health 2023. https://www.nimh.nih.gov/health/statistics/major-depression

[76] Valtonen, A.; Saunila, M.; Ukko, J.; Treves, L.; Ritala, P. AI and Employee Wellbeing in the Workplace: An Empirical Study. J. Bus. Res. 2025, 199, 115584. https://doi.org/10.1016/j.jbusres.2025.115584

[77] Pew Research Center. U.S. Workers Are More Worried Than Hopeful About Future AI Use in the Workplace. Pew Research Center 2025. https://www.pewresearch.org/social-trends/2025/02/25/u-s-workers-are-more-worried-than-hopeful-about-future-ai-use-in-the-workplace/

[78] Pendell, R. Using AI More Does Not Reassure Workers, Managers Do. Gallup 2026. https://www.gallup.com/workplace/713231/ai-not-reassure-workers-managers-do.aspx

[79] American Psychological Association. Work in America Survey 2025. American Psychological Association 2025. https://www.apa.org/pubs/reports/work-in-america/2025

[80] Microsoft Corporation. Form 10-K for the Fiscal Year Ended June 30, 2025. U.S. Securities and Exchange Commission 2025. https://www.sec.gov/Archives/edgar/data/789019/000095017025100235/msft-20250630.htm

[81] Microsoft Corporation. Annual Report 2025. Microsoft Corporation 2025. https://www.microsoft.com/investor/reports/ar25/

[82] Association of International Certified Professional Accountants. Future-Ready Finance: Technology, Productivity, and Skills Survey Report. AICPA and CIMA 2025. https://assets.ctfassets.net/rb9cdnjh59cm/4zWh0vKiPza9EvMcuekdbw/778cf838b1601270d0de71ffc53bc091/Future_Ready_Finance_Technology_Productivity_and_Skills_Survey_Report.pdf

[83] National Association of Corporate Directors. Which of the Following Activities Relating to AI Oversight Has Your Board Performed?. 2025. https://www.nacdonline.org/all-governance/governance-resources/governance-surveys/surveys-benchmarking/bpo-tables/ai-oversight-activities-conducted-by-the-board/

[84] Korst, J.; Puntoni, S.; Tambe, P. Accountable Acceleration: Gen AI Fast-Tracks into the Enterprise. Wharton Human-AI Research and GBK Collective 2025. https://ai.wharton.upenn.edu/wp-content/uploads/2025/10/2025-Wharton-GBK-AI-Adoption-Report_Full-Report.pdf

[85] Noy, S.; Zhang, W. Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence. Science 2023, 381(6654), 187–192. https://doi.org/10.1126/science.adh2586

[86] Dell'Acqua, F.; McFowland, E.; Mollick, E.; Lifshitz-Assaf, H.; Kellogg, K.C.; Rajendran, S.; Krayer, L.; Candelon, F.; Lakhani, K.R. Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. Harvard Business School 2023, 24-013. https://www.hbs.edu/faculty/Pages/item.aspx?num=64700

[87] Peng, S.; Kalliamvakou, E.; Cihon, P.; Demirer, M. The Impact of AI on Developer Productivity: Evidence from GitHub Copilot. arXiv:2302.06590 2023. https://arxiv.org/abs/2302.06590

[88] Cambon, A.; Hecht, B.; Edelman, B.; Ngwe, D.; Jaffe, S.; Heger, A.; Vorvoreanu, M.; Peng, S.; Hofman, J.; Farach, A.; others. Early LLM-Based Tools for Enterprise Information Workers Likely Provide Meaningful Boosts to Productivity. Microsoft 2023. https://www.microsoft.com/en-us/research/publication/early-llm-based-tools-for-enterprise-information-workers-likely-provide-meaningful-boosts-to-productivity/

[89] U.S. Bureau of Labor Statistics. Total Factor Productivity, 2025. U.S. Department of Labor 2026. https://www.bls.gov/news.release/prod3.nr0.htm

[90] U.S. Bureau of Labor Statistics. Productivity and Costs, Second Quarter 2026. U.S. Department of Labor 2026. https://www.bls.gov/news.release/prod2.htm

[91] Acemoglu, D. The Simple Macroeconomics of AI. Econ. Policy 2025, 40(121), 13–58.

[92] Highfill, T.; Samuels, J.D. Early Estimates of the Impact of AI within BEA's Industry Economic Accounts. U.S. Bureau of Economic Analysis 2026, BEA-WP2026-3. https://bea.gov/sites/default/files/papers/BEA-WP2026-3.pdf

[93] Microsoft Corporation. Form 10-K for the Fiscal Year Ended June 30, 2026. U.S. Securities and Exchange Commission 2026. https://www.sec.gov/Archives/edgar/data/789019/000119312526323660/msft-20260630.htm

[94] Ransbotham, S.; Kiron, D.; Khodabandeh, S.; Iyer, B.; Das, A. The Emerging Agentic Enterprise: How Leaders Must Navigate a New Age of AI. MIT Sloan Management Review and Boston Consulting Group 2025. https://sloanreview.mit.edu/projects/the-emerging-agentic-enterprise-how-leaders-must-navigate-a-new-age-of-ai/

[95] MIT Sloan Executive Education. AI Executive Academy. Program description 2026. https://executive.mit.edu/ai-executive-academy.html

[96] Piazzi, D. A Proposed High Level Approach to AI Audit. ISACA J. 2024, 2. https://www.isaca.org/resources/isaca-journal/issues/2024/volume-2/a-proposed-high-level-approach-to-ai-audit

[97] Rah, A.; Hahues, S. Retrieval-Augmented Generation (RAG), Generative AI, and Agentic AI Governance: An Integrated Enterprise Governance Prioritization Architecture. Research Square 2026. Preprint, not peer reviewed. https://doi.org/10.21203/rs.3.rs-10012178/v1

[98] U.S. Office of Management and Budget. Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence. Executive Office of the President 2024, M-24-10. https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf

[99] U.S. Government Accountability Office. Artificial Intelligence: Agencies Have Begun Implementation but Need to Complete Key Requirements. U.S. Government Accountability Office 2023, GAO-24-105980. https://www.gao.gov/products/gao-24-105980

[100] U.S. Office of Management and Budget. Accelerating Federal Use of AI through Innovation, Governance, and Public Trust. Executive Office of the President 2025, M-25-21. https://www.whitehouse.gov/wp-content/uploads/2025/02/M-25-21-Accelerating-Federal-Use-of-AI-through-Innovation-Governance-and-Public-Trust.pdf

[101] Cybersecurity and Infrastructure Security Agency; National Security Agency; Federal Bureau of Investigation. Deploying AI Systems Securely. U.S. Department of Homeland Security 2024. https://www.ic3.gov/CSA/2024/240415.pdf

[102] Rah, A. Enterprise Generative AI: A Systematic Review of Security Risks and Governance. Research Square 2026. Preprint, not peer reviewed. https://doi.org/10.21203/rs.3.rs-9890887/v1

[103] Milanez, A. The Impact of AI on the Workplace: Evidence from OECD Case Studies of AI Implementation. OECD Social, Employment and Migration Working Papers 2023, 289. Working paper; not an official OECD view. https://doi.org/10.1787/2247ce58-en

[104] Merola, R.; Ernst, E.; Samaan, D.; del Rio-Chanona, M.; Teutloff, O. The Impact of GenAI on Jobs, Productivity and Work Organization: A Review of the Empirical Evidence. International Labour Organization 2026. Research brief, 1 June 2026. https://doi.org/10.54394/00034628

Appendix A. Complete Search Strategies

The protocol for this update was written on 8 November 2026, after the original corpus had been assembled. It was not registered. It keeps the seven questions RQ1–RQ7, the window November 2022 through September 2026, English-language sources, the inclusion and exclusion rules in the Methods, and a narrative synthesis. Primary outcomes are employment change, task reallocation, human time including verification, job insecurity and strain, and organizational performance. Secondary outcomes are training, oversight arrangements, and governance duties. The protocol does not authorize pooling.

The Boolean strings, export dates, and hit counts used to build the corpus before 8 November 2026 were not archived. Scopus, Web of Science, IEEE Xplore, and the ACM Digital Library were not queried in this revision. No institutional session for those indexes was available. Visiting a publisher page is not recorded here as a search of that index. PubMed was queried once through the public E-utilities interface, as recorded below.

One index was both queried and exported. On 8 November 2026 three OpenAlex works calls were exported in full. Scopus, Web of Science, IEEE Xplore, and the ACM Digital Library were not queried. A Crossref bibliographic query for human–AI task allocation, verification, and accountability, limited to 2018 through 30 September 2026, returned 1,878,646 works and was not exported. The same day, PubMed E-utilities returned one record for the title/abstract query ("generative artificial intelligence" OR ChatGPT) AND (employment OR workload OR burnout OR layoff) AND (worker OR employee OR firm), with publication dates from 1 November 2022 through 30 September 2026. That record is Dave, Martin, David, Kumar, and Chakraborty, a scoping review of healthcare-worker mental health, *International Journal of Medical Informatics*, 2026, volume 205, article 106122, doi:10.1016/j.ijmedinf.2025.106122 (PMID 41037981). The title was excluded because the setting is clinical care, not an enterprise workforce estimate. It was not cited.

A separate export on the same date is not part of those three calls. OpenAlex title-and-abstract search for enterprise generative artificial intelligence and employment, 1 November 2022 through 30 September 2026, returned 39 and the file contains 39. The same dates for generative artificial intelligence, employee, burnout, and workload returned 2 and the file contains 2. Human, artificial intelligence, responsibility, accountability, and workflow, 1 January 2018 through 30 September 2026, returned 330 and the file contains 330. PubMed E-utilities, title or abstract ("generative artificial intelligence" OR ChatGPT) AND (employment OR workload OR burnout), publication dates 1 November 2022 through 30 September 2026, returned 272 and the file contains 272. Scopus, Web of Science, IEEE Xplore, and the ACM Digital Library were not queried. APA PsycInfo was not queried. Two later strings, for employee workload and verification and for layoffs and burnout, returned 1 and 2 OpenAlex records and 0 PubMed records. All three OpenAlex records were excluded on title or abstract, as stated in the Methods.

The first OpenAlex call used a title-and-abstract filter on enterprise, generative, artificial, intelligence, and workload, for publication dates from 1 November 2022 through 30 September 2026. It returned 68 works and the export contained 68. Screening of that export is unchanged: one duplicate, 67 titles, one study added to the synthesis. An earlier relevance count of 43,131 works was not exported.

The second call used a title-and-abstract filter on human-in-the-loop, task, allocation, and artificial intelligence, for publication dates from 1 January 2018 through 30 September 2026. It returned 90 works and the export contained 90. One duplicate title was removed. Of the 89 unique titles, 85 were excluded on title because they concerned cybersecurity teaming, robots or aircraft, clinical or military systems, infrastructure, commentaries, indexes, or another topic that did not state an enterprise workforce estimate or an accountability-aware workflow. Abstracts were retrieved for four records. One abstract was empty. One Zenodo abstract stated outcome improvements, but the retrieved text was truncated and those figures were not used. Two abstracts were retained for the technical comparison only.

The third call used a title-and-abstract filter on generative AI, verification, workload, and workers, for the review window. It returned 3 works. All three were excluded on title: an editorial, a Zenodo essay, and an occupational-medicine opinion. None was added.

Appendix B. PRISMA 2020 Documentation

The flow below uses only the exported screen. It does not describe the earlier corpus, whose identification and exclusion counts were not retained. Fig. 2 in the main text remains the path for that earlier corpus and still has no identification count.

OpenAlex export 1, 8 November 2026: 68 identified and exported. Duplicate removed: 1. Titles screened: 67. Excluded on title: 64. Abstracts retrieved: 3. Excluded after abstract: 2. Added to the synthesis: 1.

OpenAlex export 2, same date: 90 identified and exported. Duplicate removed: 1. Titles screened: 89. Excluded on title: 85. Abstracts retrieved: 4. Not used after abstract: 2. Added to the technical comparison only: 2. Added to the outcome synthesis: 0.

OpenAlex export 3, same date: 3 identified and exported. Excluded on title: 3. Added: 0.

Crossref bibliographic query: 1,878,646 reported, not exported, not screened.

PubMed, same date: 1 record retrieved and excluded on setting. Scopus, Web of Science, IEEE Xplore, and the ACM Digital Library were not searched.

Earlier corpus: still the previously verified sources. Its upstream yield is unknown. These exports did not produce that corpus.

Later export, different strings, same date: OpenAlex 39, 2, and 330, plus a wider PubMed export of 272. Duplicate removed: 1. Titles screened: 642. Excluded at title: 589. Abstracts sought: 53. OpenAlex abstracts missing: 11, all SSRN. Crossref abstracts then retrieved for those 11. Abstracts assessed: 53. Excluded after abstract: 53. Full texts sought: 0. Added to the synthesis: 0. The 104 cited works are outside this count and are not added to 643.

Checklist, completed only where the record supports the item. Title identifies a critical integrative review, not a systematic review (page 1). The seven questions, eligibility, and narrative synthesis are in the Methods. Information sources name the indexes that were searched and the indexes that were not. The 8 November 2026 strings are copied in Appendix A. The historical search strategy is unavailable, so PRISMA-S cannot be completed for the earlier corpus and a PRESS review of those strings was not possible. No information specialist peer-reviewed the new strings. Selection was one reviewer, which does not meet the JBI Manual for Evidence

Synthesis requirement of two reviewers. Data items are those in Appendix D. Study risk of bias was not scored with RoB 2, ROBINS-I, or a JBI design checklist (Appendix E). Results were not pooled (Appendix F). Certainty was not graded. The review was not registered. Cited originals used in the synthesis are public. Record-level decisions for the exported files are retained with the source and summarized here rather than printed as several hundred rows.

The new export reconciles as follows. Identified: OpenAlex 39 + 2 + 330 and PubMed 272, total 643. Duplicate removed: 1. Titles screened: 642. Excluded at title: 589. Abstracts sought: 53. OpenAlex had no abstract for 11 SSRN records. Crossref supplied an abstract for each of those 11 on 8 November 2026. Abstracts assessed: 53. Excluded after the abstract: 53. Full texts sought: 0, because no abstract met the rule for a new outcome study. Studies added: 0. Check: 1 + 589 + 53 = 643. The earlier OpenAlex exports of 68, 90, and 3 are a separate flow in the box above and are not added into the 643. The 104 cited works are outside both flows. Their database yield was not archived, so it is not imputed. Full-text retrieval of already-cited public PDFs, same date, is a separate step and is not an identification count. Eight public PDFs were saved. A partial RoB 2 note was made for NBER 33795 only. No study was added or removed. See Appendix E. The 104 cited works remain outside the 643.

Appendix C. Eligibility and Screening

The same inclusion and exclusion rules used for every other source were applied to the 68-record export, including author-written postings. A record was eligible for addition only when the title, and where retrieved the abstract, stated a checkable design and an employee, workload, or organizational outcome inside the review questions. Infrastructure uses of the word workload, security architectures, and commentaries without a design were excluded.

Of 67 unique titles, 64 were excluded at title: 28 because workload meant computing infrastructure; 14 because no workforce or performance estimate was stated; 8 because the record was a security or architecture description without that estimate; 6 because the record was a dataset, workshop, dissertation, book, commentary, or market note; 5 because the setting was clinical, educational, inspection, or a school rather than the enterprise workforce question; 2 because the title did not state a design; and 1 because the paper was a method for analyzing scientific literature. Abstracts were then retrieved for three records. Two were not added. A Discover Artificial Intelligence abstract proposes dynamic human intervention, but the retrieved abstract was truncated and its evaluation figures were not used. A Journal of Information Systems Engineering and Management abstract states managerial time savings and adoption shares without a checkable sample design, and those figures were not used. One record was added: Li, Sun, Yang, and Zeng, European Journal of Innovation Management, 2026, a cross-sectional study of 276 corporate research-and-development employees. Volume, issue, and pages were not on the record used here. The DOI is 10.1108/ejim-11-2025-1496.

The later 643-record export was screened with the same rules. One duplicate title was removed. Of 642 titles, 589 were excluded at the first pass: 254 because the setting was clinical, educational, military, agricultural, or another off-scope domain, and 335 because the title did not state a workforce or performance estimate. Abstracts were requested for the remaining 53. Forty-two abstracts were on OpenAlex and were read. Eleven SSRN records had no OpenAlex abstract. Crossref abstracts for those eleven were read the same day. The exclusions are: doi:10.2139/ssrn.7537464, a multi-country working paper, not a U.S. payroll contrast, coefficients not taken from a re-read table; doi:10.2139/ssrn.7512166, a security-operations simulation outside the enterprise question, full text not retrieved; doi:10.2139/ssrn.6862938, a secondary synthesis of institutional reports; and eight conceptual or governance papers (doi:10.2139/ssrn.7546940, 10.2139/ssrn.7628020, 10.2139/ssrn.7575218, 10.2139/ssrn.7251121, 10.2139/ssrn.7485400, 10.2139/ssrn.7537584, 10.2139/ssrn.7262481, and 10.2139/ssrn.6328579). None was added. Full texts of the 53 were not retrieved. None received a RoB 2 or ROBINS-I score. The 104 cited works were sorted from the stored bibliography and from the role already assigned in the manuscript. That sort separates outcome evidence from theory, methods, official series, surveys, governance, and author preprints. It is not a new full-text appraisal, and those 102 are not added to the 643.

Appendix D. Study Characteristics and Evidence Extraction

Table XII lists the designs that carry the main contrasts in the results. Each figure is one already checked against the cited source in this manuscript. The table is not a claim that these studies were the output of the 68-record screen.

Appendix E. Methodological Quality and Risk of Bias

RoB 2, ROBINS-I, and GRADE were not scored. A domain judgment that was not produced by reading each full text against the instrument is not printed here. The five-question checklist in the Methods remains the appraisal that was used: venue class, design, geography, explicit numerators and periods, and independence from the vendor whose product is evaluated. It does not generate a risk-of-bias score. Journal quartile and citation counts were not used as quality scores.

The instrument that would apply, if the full text were re-opened, is named below. The judgment column is intentionally blank. These rows are flagged for author verification. They are not AI-assigned risk-of-bias ratings.

Design already in the results	Instrument that would apply	Judgment
Fan et al., randomized lab tasks	RoB 2	Not scored. Full text not re-opened.
Cui et al., randomized developer rollout	RoB 2	Not scored. Full text not re-opened.
Dillon et al., randomized Copilot rollout	RoB 2	Not scored. Full text not re-opened.
Humlum and Vestergaard, difference-in-differences	ROBINS-I	Not scored. Full text not re-opened.
Hyman et al., matched trainees	ROBINS-I	Not scored. Full text not re-opened.
Census, BLS, BEA, Gallup, Pew, Li et al.	None of those tools. Official series or a cross-section.	Not a domain score.

TABLE XII
DESIGNS THAT CARRY THE MAIN CONTRASTS. EACH FIGURE WAS ALREADY CHECKED IN THE MANUSCRIPT. THIS TABLE IS NOT THE OUTPUT OF THE 68-RECORD SCREEN.

Source	Design and sample	Outcome used here	Limit
Bonney et al., Census BTOS, CES-26-25	U.S. firms, Nov. 2025–Jan. 2026	44% of AI-using firms report augmentation; headcount decrease reported by 2.0% firm-weighted	Firm self-report. Not causal. The 2.0% figure stays in that subsection only.
Bick, Blandin, and Deming, Manage. Sci. 2026	U.S. adults, late 2024	Work use in the prior week 27% of employed adults; reported hours saved 1.4%	Gross self-report, not net time.
Dillon et al., NBER 33795	Vendor-affiliated Copilot rollout	Outlook time -1.4 hours/week (ITT)	Email time, not total workload.
Cui et al., Manage. Sci.	4,867 developers, three firms	Completed tasks $+26.08\%$ (SE 10.3%)	Vendor-affiliated. Not firm performance.
Humlum and Vestergaard, NBER 33777	Danish difference-in-differences	Rules out earnings and hours effects larger than 2%	Not a U.S. total. New oversight tasks are descriptive.
Hyman et al., NBER 34174	Matched U.S. trainees	High-exposure quarterly returns about \$1,000 to about \$3,000	Mostly moves into less-exposed occupations.
Lee et al., CHI 2025	319 knowledge workers, 936 examples	Effort often rated lower; work shifts toward verification	Self-report. Vendor-affiliated. Not a time diary.
Fan et al., CHI 2026	60 participants, three tasks	Time -22%; verification load tracks fatigue	One lab session, not a workday.
Pendell, Gallup	Panel, 2023–early 2026	Displacement concern: engagement -0.20 SD, burnout $+0.18$ SD	Not a layoff effect and not a use-to-burnout coefficient.
Li et al., Eur. J. Innov. Manage. 2026	276 R&D employees, cross-section	Workload and support associated with generative-AI use; use with innovative performance	Not U.S. payroll and not net hours.
Challenger, Gray & Christmas	U.S. announced-cut tallies	AI-cited plans, including 116,175 year-to-date through August 2026	Employer attribution, not completed separations.

The instruments that would apply later, and only after a registered protocol and a second reviewer, are RoB 2 for the randomized task experiments and ROBINS-I for the quasi-causal labor studies. Census and BLS series are official descriptive statistics, not subjects of those instruments. Challenger counts are corporate attributions. NIST, COBIT, OMB, and product documents are governance or architecture sources. Author preprints are context. The two abstracts that were retrieved and not added were not quality-scored, because they were not included. The 64 title exclusions were not quality-scored for the same reason. The 42 OpenAlex abstracts and the 11 Crossref abstracts were excluded before appraisal. A JBI checklist was not applied. Domain scores remain blank unless a public full text was opened and the domain is supported by that text. Human author verification is required before any row is treated as a finished rating. There was no second reviewer.

On 8 November 2026, public PDFs already linked from the bibliography were downloaded. This was not a new database search and it is not added to the 643-record flow. Eight files were saved: Census CES-WP-26-25 (79 pages), NBER 33795 (38 pages, revised November 2025), NBER 33777 (131 pages, revised March 2026), NBER 34174 (75 pages), NBER 35677, BEA WP2026-16 (20 pages), CESifo Working Paper 10601, and ILO Working Paper 140. RoB 2 is used only for the randomized Copilot access experiment. ROBINS-I is the tool that would apply to the Danish adopter contrast. Neither tool is applied to the Census survey, the BEA expectations paper, the ILO exposure index, or the other downloaded files whose identification sections were not re-read.

Full text opened	Tool	What was judged
Dillon et al., NBER 33795	RoB 2	D1, some concerns: licenses were randomized inside a firm-supplied list, and the paper states that the highest-priority users were likely excluded. D2–D5 were not given an overall label. No pre-analysis registration was found in this PDF. Overall risk was not assigned.
Humlum and Vestergaard, NBER 33777	ROBINS-I	Confounding is a concern because chatbot use is not assigned. The design is difference-in-differences. The other domains were not scored. The bound that rules out hours effects larger than 2 percent was re-read on page 5 and is unchanged. Denmark, not a U.S. total.
Bonney et al., CES-26-25	Neither tool	Firm survey. The 2.0 percent firm-weighted decrease was re-read on page 5 and stays in that subsection.
Hyman et al.; Bick et al. NBER 35677; Hui et al. CESifo 10601; Highfill and Samuels; ILO WP 140	Not scored	The PDF was retrieved. The identification section was not appraised in this pass, so no rating is printed.

OpenAlex, the same day, reported no open PDF for the journal versions of Bick, Blandin, and Deming, *Management Science*, and of Li et al. Lee et al. and Fan et al. resolve to ACM links that were not downloaded. Cui et al. and Noy and Zhang have separate SSRN copies marked open; those copies were not downloaded, and the journal PDFs were not opened. Valtonen et al. resolves to the journal landing page and was not downloaded. For those studies the earlier manuscript limits stand, and no new risk-of-bias rating is assigned. No study was added or removed.

APPENDIX F. EVIDENCE SYNTHESIS AND CERTAINTY

No meta-analysis was conducted. The outcomes are not on a common scale. Certainty was not graded. The classes below are the ones used in the results.

RQ1. Official descriptive evidence: augmentation is the modal reported task effect. That description does not identify job elimination. RQ2. Announced-cut attributions and Census headcount reports measure different objects. Neither is a

completed-separation effect. RQ3. Danish records and product documents show oversight tasks and approval gates. They do not measure hours moved onto remaining U.S. workers. RQ4. Survey associations link displacement concern to lower engagement and higher burnout. They are not a clinical diagnosis and not a layoff effect. Gross task-time savings are a separate object, under P4. RQ5. Near-term adoption forecasts moved toward realized use. Productivity, employment, and return expectations are not closed. An exposure index is not this comparison. RQ6. No design compares AI adopters that reduced headcount with AI adopters that did not. RQ7. P2 and organizational consequence remain propositions.

P1 is descriptively consistent and not identified. P2 is untested in U.S. payroll data. P3 matches one panel association and is not a layoff effect. P4 is consistent with a shorter lab task and with verification load that tracks fatigue, and with a self-report that effort shifted toward verification. Table XI in the main text is the map from these statements onto HARWO. The map is not a confirmation.

Appendix G. Full HARWO Technical Specification

HARWO is a proposed computational reading of the responsibility-transfer framework. It is not an implemented solver, not a trained model, and not a scored enterprise log. No optimality or runtime claim is made.

Precedence is an acyclic graph $G = (V, E)$ with work tasks V_T and decision nodes V_D. Actors are the fixed human set X_H and the AI agents X_A. A policy $\pi = (e, g, a)$ assigns an eligible executor $e(v)$, a human authorizer $g(d)$ or none, and exactly one accountable human $a(v)$ for every release. Execution and accountability are different assignments. A verification return is not an edge of G. It may send the output of task v back to v at most m_v times, and it ends on acceptance or at that cap. At the cap without acceptance, the case uses an exception edge of G, which cannot return to v. A policy that removes an employee is outside the feasible set. A smaller roster may be supplied from outside and scored as another fixed set. One optimization does not choose the layoff.

Completion time T is the makespan of the case. Human workload H sums execution, prompting, verification, correction, supervision, and human time on an exception path, whether or not those minutes create an incremental cash expense. Operational loss R is expected loss on outputs that pass the chosen gate. Monetary cost C excludes the minutes already in H. It includes applicable model usage, licensing, and other incremental cash costs with no accounting overlap with H. Both β and δ may be positive. If a reported cost series is exactly $a + bH$, its intercept does not change the minimizer and its variable part can be absorbed into the human-time weight. That special case is not imposed on the general model. The raw weighted sum is not a second objective. The scalar problem is scored only after each expectation is divided by a positive reference scale chosen in advance:

$$\begin{aligned}&\min_{\pi \in \Pi} \\ &\alpha\, \mathbb{E}[T(\pi)]/T_{\mathrm{ref}} + \beta\, \mathbb{E}[H(\pi)]/H_{\mathrm{ref}} \\ &+ \gamma\, \mathbb{E}[R(\pi)]/R_{\mathrm{ref}} + \delta\, \mathbb{E}[C(\pi)]/C_{\mathrm{ref}},\end{aligned}$$

with nonnegative weights summing to 1. A missing expectation or a missing scale leaves the policy unscored. Feasibility requires precedence, eligibility, human capacity L_h at the fixed roster, any stated error tolerance q_v, human authorization for every task in V_M, and one accountable human on every release. NIST AI RMF 1.0, NIST AI 600-1, and COBIT 2019 name governance duties. They do not set the weights, the tolerances, or V_M.

The decision procedure is a specification, not a tested program. Build the graph. Mark mandatory gates and eligibility. Hold the human set fixed and name executors, authorizers, and one accountable human. Score the scalar problem only where every term is observed. A person selects among feasible scored policies. The procedure does not authorize deployment. P2 and P4 are the propositions this object can express, as human time at fixed headcount and as loop time inside H. P1 and P3 are not implied by a feasible policy. Table X in the main text is the comparison. The two added rows use abstracts, not full texts. Donat-Bouillud and Amokrane-Ferka reallocate call-center supervision from operator load, stress, and activity. van Tuijn and colleagues treat levels of automation as a language for team design and name accountability as a requirement. Neither abstract formulates $e(v)$, $g(d)$, and $a(v)$ as different assignments, or prices verification apart from completion time, on a roster that cannot be reduced. HARWO states that joint constraint. It does not solve it. A later test needs an observed graph, measured human time, an error or loss record, and a fixed roster. None of those inputs was scored here.

Appendix H. Reference and Citation Audit

The reference list is numbered in order of first citation after this revision's bibliography rebuild. Printed author, title, source, year, volume, issue, and pages are the fields stored for this manuscript. A DOI link is printed only when a DOI was already stored. A URL is printed when no DOI was stored and a URL was. Entries with neither are identified by venue, volume, and pages, or by institution and report number. Those identifiers were not invented. Two unread Microsoft Form 10-K records for fiscal 2023 and fiscal 2024 remain in the working bibliography and are not cited. Duplicate postings of the same preprint are cited once. Preprints are labeled as not peer reviewed. This audit did not repeat a live HTTP check of every DOI.

Seven author-written works are in the list: the predecessor review and six preprints. They are cited as boundary, governance, or architecture context, not as proof of employee outcomes or organizational productivity. Four further author postings were screened with the same rules and were not added. A Research Square posting on digital transformation and cybersecurity (DOI 10.21203/rs.3.rs-9734788/v1) and a ResearchGate record on governance, human resistance, and cybersecurity (DOI 10.13140/RG.2.2.19233.38964) repeat the readiness argument already cited from the Preprints version. A digital-twin networking benchmark (SSRN DOI 10.2139/ssrn.7260325) and a preprint on quantum trading agents (Research Square DOI 10.21203/rs.3.rs-10253799/v1) do not measure enterprise workload or the quantities HARWO

leaves unmeasured. The ResearchGate DOI was taken from the author's reference list and did not resolve in Crossref. The count of distinct author works is unchanged by these screens. The study added to the synthesis, and the two papers added to the comparison, are not author preprints.