

# When the Task Map Folds: Empirical Patterns in Knowledge-Work Skill Composition After Generative AI

*Ingrid Brouwer\**, *Kavya Ramanujan*

Center for AI and Knowledge Work (CAIKW)

Generative Economic Review • May 17, 2026

GER 1.4

---

**JEL Classification:** J21, J23, J24, J31, J63, O33, C81

**Keywords:** generative artificial intelligence, labor demand, knowledge work, online job postings, occupational exposure, skill composition, within-occupation polarization, task-based framework, restructuring hypothesis, post-ChatGPT

---

## Abstract

We document the early empirical patterns in US knowledge-work labor demand over the thirty-three months following the November 2022 public release of large language models, using a panel of approximately 41 million online job postings from January 2023 through September 2025. We classify postings by occupation (SOC 2018 six-digit) and merge with occupation-level AI exposure scores following Eloundou et al. (2023). Three findings are central. First, between 2023Q2 and 2025Q3, postings in the top exposure quintile declined by 19.4 percent while postings in the bottom quintile declined by 4.1 percent, a 15.3 percentage-point differential that survives controls for industry mix, region, and macroeconomic conditions. Second, the within-occupation composition of skill requirements shifted substantially in highly-exposed occupations: the share of postings mentioning routine cognitive tasks (data entry, standard report generation, first-line response handling) fell by 6.8 percentage points; the share mentioning AI-collaboration skills (prompt engineering, AI verification, AI workflow integration) rose by 7.1 percentage points; the share mentioning judgment-intensive skills (architectural design, ambiguous-case judgment, strategic communication) rose by 4.6 percentage points. Third, the within-occupation posted wage distribution polarized: the 75th-percentile posted wage in highly-exposed occupations rose by 12.4 percent while the 25th-percentile fell by 3.2 percent, a 15.6-percentage-point differential robust to

controls. The three margins jointly support the restructuring hypothesis articulated in the contemporary methodology literature (Acemoglu et al., 2022; Eloundou et al., 2023): generative AI is substituting for routine cognitive tasks at the lower end of the within-occupation distribution while complementing judgment-intensive tasks at the higher end. We are explicit that the design is descriptive of a cross-sectional differential under a single common shock; the residual confounds from contemporaneous monetary tightening and post-pandemic sectoral reallocation are documented and partial-out diagnostics reported. We close by drawing implications for occupation classification, workforce education investment, and the projection of long-run skill premia.

---

## **1. Introduction**

Whether and how generative artificial intelligence will reshape labor markets is one of the most consequential open questions in contemporary economics. Public debate oscillates between the prediction that large language models will eliminate vast swaths of white-collar employment and the prediction that, like prior general-purpose technologies, they will create as many jobs as they destroy on a horizon long enough to matter (Acemoglu and Restrepo, 2019; Bresnahan and Trajtenberg, 1995). The empirical record over the thirty-three months following the November 2022 capability shock is now long enough to support careful descriptive analysis of the patterns that have begun to emerge.

### **1.1 The framing hypothesis**

This paper makes one central empirical claim. US knowledge-work labor demand has, in the thirty-three months following the November 2022 public release of large language models, restructured along three margins simultaneously: posting volumes have fallen disproportionately in highly-exposed occupations; within those occupations, the composition of skill requirements has shifted toward AI-collaboration and judgment-intensive content and away from routine cognitive content; and the within-occupation posted wage distribution has polarized, with high-percentile wages rising and low-percentile wages falling. The three margins together correspond to the predictions of the restructuring hypothesis articulated in the contemporary methodology literature, and the joint pattern is empirically distinct from the alternative hypotheses (pure substitution, pure complementarity, reorganization without volume change, or null effect) that the literature has considered.

### **1.2 Four contributions**

The paper makes four substantive contributions to the empirical labor-economics literature on generative AI.

First, we provide the most comprehensive contemporary empirical record of US knowledge-work labor demand in the post-ChatGPT period, using a panel of approximately 41 million online job postings from January 2023 through September 2025. The panel covers all SOC 2018 six-digit occupations with adequate posting frequency, all US metropolitan statistical areas, all NAICS four-digit industries, and the full wage distribution where wages are posted.

Second, we document a 15.3 percentage-point cross-quintile differential in posting volume change between the top and bottom AI-exposure quintiles, controlling for industry mix, region, firm size, and macroeconomic conditions. The differential is concentrated in occupations whose modal tasks include routine cognitive work that large language models can now perform with high reliability.

Third, we document a within-occupation skill composition shift. In the top exposure quintile, the share of postings mentioning routine cognitive skills (data entry, standard report generation, first-line response handling) fell by 6.8 percentage points between 2023Q1 and 2025Q3; AI-collaboration skills (prompt engineering, AI verification, AI workflow integration) rose by 7.1 percentage points; judgment-intensive skills (architectural design, ambiguous-case judgment, strategic communication) rose by 4.6 percentage points. These shifts are absent in the bottom exposure quintile, ruling out interpretations that attribute the change to economy-wide trends rather than to AI exposure specifically.

Fourth, we document a within-occupation wage polarization pattern. The 75th-percentile posted wage in highly-exposed occupations rose by 12.4 percent while the 25th-percentile fell by 3.2 percent, a 15.6-percentage-point spread. The polarization is robust to controls and is concentrated in occupations whose AI-collaboration skill mentions rose most.

### **1.3 Intellectual history of the question**

The question this paper engages reached its current form through three intellectual transitions. Autor et al. (2003) established the task-based framework in which technology, capital, and labor compete to perform tasks; their decomposition of computerization's labor-market effects established the modern empirical approach to automation. Acemoglu and Autor (2011) formalized the framework and used it to interpret the polarization of US wages during the late twentieth century. Eloundou et al. (2023) extended the framework to generative AI specifically, constructing a mapping from O\*NET task descriptions to large-language-model capabilities and estimating economy-wide exposure. The contribution of the present paper is to implement the predicted patterns under the Eloundou-style exposure mapping in the post-ChatGPT period, documenting the three margins of labor-demand response jointly.

### **1.4 What the paper claims**

The paper makes five explicit empirical claims that the reader can evaluate against the evidence in Sections 4 and 5:

1. Posting volume in the top AI-exposure quintile fell by 19.4% between 2023Q2 and

- 2025Q3; in the bottom quintile it fell by 4.1%; the 15.3 pp differential is statistically significant ( $p < 0.01$ ) and robust to controls.
2. Within the top quintile, routine cognitive skill mentions fell by 6.8 pp; AI-collaboration skill mentions rose by 7.1 pp; judgment-intensive skill mentions rose by 4.6 pp; the corresponding changes in the bottom quintile are statistically indistinguishable from zero.
  3. The 75th-percentile posted wage in highly-exposed occupations rose by 12.4% while the 25th-percentile fell by 3.2%, a 15.6 pp polarization spread.
  4. The joint pattern (negative posting volume, positive judgment-intensive composition shift, polarized wages) is the predicted signature of the restructuring hypothesis and is empirically distinguishable from the alternative hypotheses.
  5. The differential survives partial-out diagnostics for contemporaneous monetary tightening (using occupation-level interest-rate beta) and post-pandemic sectoral reallocation (using Dingel and Neiman (2020) telework-feasibility scores).

The claims are descriptive of the cross-sectional differential; we do not claim clean causal identification of the AI-specific channel under the single common shock of November 2022.

## 1.5 Roadmap

Section 2 places the analysis within the literatures on routine-biased technical change, the labor-market effects of generative AI, job postings as a measure of labor demand, the within-occupation skill composition literature, the recent econometric literature on difference-in-differences with single common shocks, and the methodological discipline of pre-registration in observational economics. Section 3 describes the data, the AI-exposure mapping, the skill-composition measurement, the regression specifications, and the pre-specified robustness margins. Section 4 reports the central findings. Section 5 discusses interpretations against the four candidate hypotheses, addresses the joint-shocks identification concern, and considers the international evidence and broader implications. Section 6 concludes.

A note on identification is in order. The November 2022 release of large language models is a single common shock affecting the whole US labor market. The post-2022 period coincides with the fastest monetary-tightening cycle since the Volcker disinflation, with post-pandemic sectoral reallocation, and with the 2023 contraction of the technology sector. The cross-sectional differential we document cannot be attributed cleanly to the AI-specific channel; what we offer is a careful description of the patterns conditional on the macroeconomic context of the post-ChatGPT window, with explicit partial-out diagnostics that bound the residual identification gap.

## 2. Literature Review

Six literatures bear directly on the empirical analysis. We treat each in turn and close with a paragraph on the position of the present paper.

### 2.1 The task-based framework and routine-biased technical change

The task-based framework for labor market analysis is the conceptual foundation of the empirical work in this paper. Autor et al. (2003) document that the diffusion of computers from the 1980s through the 1990s reduced demand for routine tasks—both manual and cognitive—and raised demand for non-routine analytical and interpersonal tasks. The framework explained the polarization of US wages during the 1980s–1990s as the result of computerization’s relative ease of substituting for routine cognitive tasks (middle-skill work) while leaving non-routine analytical (high-skill) and non-routine manual (low-skill) tasks largely unaffected.

Acemoglu and Autor (2011) formalize the framework, representing production as the performance of tasks rather than as the application of indivisible factor inputs. Tasks are heterogeneous in their automatability—their susceptibility to substitution by capital, software, or both—and workers are heterogeneous in their comparative advantage at performing different tasks. Technological change shifts the boundary of tasks that capital can perform, redistributing workers across the remaining task space and reshaping the demand for skills.

For generative AI, the task-based framework remains the natural starting point but requires re-tuning. Where computerization substituted for routine tasks, generative AI plausibly substitutes for non-routine cognitive tasks—exactly the tasks the previous framework took to be the protected domain of high-skill labor (Autor, 2024; Acemoglu et al., 2022). If this characterization is correct, the demand effects of generative AI may show patterns that invert the patterns observed during computerization: a compression rather than a polarization of the wage distribution at the cross-occupation level, but a polarization within highly-exposed occupations as routine-cognitive sub-tasks within those occupations are substituted while judgment-intensive sub-tasks are complemented.

### 2.2 The labor market effects of generative AI

The rapidly growing body of work on the labor market effects of generative AI specifically constitutes the literature that the present paper most directly extends. Eloundou et al. (2023) construct a mapping from O\*NET task descriptions to the capabilities of large language models and estimate that approximately 80 percent of US workers could see at least 10 percent of their tasks affected by generative AI. Their occupation-level exposure measure is the conceptual foundation of the exposure measure we apply.

Brynjolfsson et al. (2023) report results from a randomized field experiment in customer support documenting that generative AI tools raise the productivity of less-skilled workers more than that of more-skilled workers, with the productivity effect concentrated in the

bottom three deciles of the worker skill distribution. The finding is consistent with the within-occupation polarization mechanism we document: AI complements high-skill judgment-intensive work while substituting for low-skill routine cognitive work, with the within-occupation wage distribution adjusting accordingly.

Noy and Zhang (2023) document similar productivity gains in writing tasks, with a similar pattern of larger gains for less-skilled workers. Peng et al. (2023) document a substantial reduction in coding task completion times among software developers using AI pair-programming tools, with the reduction concentrated in routine sub-tasks of software development.

Acemoglu (2024) provides an aggregate framework that maps these microeconomic productivity gains to long-run growth implications, with conclusions more conservative than the most aggressive projections in industry reports. Acemoglu argues that the within-task productivity gains documented in microeconomic studies do not aggregate cleanly to economy-wide productivity because only a fraction of work tasks are exposed to AI, the cost of human supervision of AI output is non-trivial, and the deployment of AI substitutes is constrained by complementary investments. The empirical patterns we document support a test of this framework by examining whether documented within-task productivity gains coincide with the predicted within-occupation skill-composition shifts.

Humlum and Vestergaard (2024) use Danish administrative data to estimate the labor demand response to generative AI and find modest aggregate effects in the early period. The Danish and US results have, in the contemporary working-paper literature, been compared without a common methodological specification; one of the contributions of the comparison framework we operate within is to render such cross-country comparisons more disciplined.

### **2.3 Job postings as a measure of labor demand**

The empirical literature on job postings as a measure of labor demand has matured substantially. Hershbein and Kahn (2018) demonstrate that job posting data captures meaningful variation in firm-level skill requirements and that postings exhibit substantial information beyond what is recoverable from administrative employment data. They document the use of posting data in identifying recession-driven changes in firm skill demand and validate the data's representativeness against benchmarks from the Bureau of Labor Statistics.

Deming and Kahn (2018) use job postings to document the rising premium on social skills since 2000, finding that occupations with rising social-skill content also exhibit faster employment growth. Their methodological contribution—a procedure for extracting standardized skill terms from posting text—is the conceptual foundation of the skill-composition measurement we apply.

Acemoglu et al. (2022) use job postings to identify "AI-exposed" firms and document their differential labor-demand response in the pre-generative-AI period. Their methodology of measuring firm AI exposure from postings text is distinct from the occupation-based

exposure measure we apply, but the two are complementary: occupational exposure asks how AI affects the demand for workers in a given occupation, while firm AI exposure asks how AI-using firms differ in their demand for workers regardless of occupation.

The principal limitations of posting data are well documented. Carnevale et al. (2014) note that postings systematically over-represent white-collar work, urban locations, and larger firms, and under-represent skilled trades, rural locations, and smaller employers. Posting duration—the time a posting remains active before being filled—exhibits substantial variation that complicates interpretation of posting volumes (Hershbein and Kahn, 2018). Modestino et al. (2020) and others have documented that posted skill requirements may be “aspirational” in the sense that firms post for skills they do not require of marginal hires; the implication is that within-occupation shifts in posted skills should be interpreted as shifts in stated requirements rather than actual hiring criteria.

#### **2.4 Within-occupation skill composition**

A growing literature has examined the within-occupation evolution of skill requirements as a complement to the more conventional cross-occupation employment analysis. Deming and Kahn (2018) pioneered the systematic extraction of skill terms from posting text and documented the rising premium on social skills. Atalay et al. (2020) extend this approach to identify within-occupation skill-content shifts across multiple decades of US labor demand.

For the post-ChatGPT period, the within-occupation analysis is particularly important because the aggregate cross-occupation framework (under which AI is sometimes characterized as “substituting for cognitive workers”) understates the heterogeneity of the labor-market response. Within a single occupation, generative AI can substitute for routine cognitive sub-tasks (e.g., drafting standard documents) while complementing judgment-intensive sub-tasks (e.g., reviewing AI-generated drafts for accuracy and strategic fit). The within-occupation skill composition we measure is the empirical signature of this within-occupation heterogeneity.

#### **2.5 The recent econometric literature on difference-in-differences**

The empirical design we employ has, at its core, a difference-in-differences comparison between high- and low-AI-exposure occupations over the November 2022 capability shock. The recent econometric literature has documented several identification concerns with such designs.

de Chaisemartin and D’Haultfoeuille (2020) document that two-way fixed-effects estimators in panels with heterogeneous treatment timing produce weighted averages of unit-level effects in which some weights can be negative, raising interpretation concerns even under the parallel trends assumption. For the present setting—a single common shock affecting all units simultaneously—the heterogeneous-timing concerns do not arise. The relevant concern is the cross-sectional one: that the cross-section of exposure may correlate with

unobserved time-varying confounds.

Rambachan and Roth (2023) document that conventional pre-trend tests have low power to detect violations large enough to bias post-event estimates. We report Rambachan-Roth sensitivity bounds on the principal differential coefficient under three nested restrictions on the post-event pre-trend violation, finding that the result survives at the  $\bar{M} = 1$  restriction.

Callaway and Sant'Anna (2021) propose a generalized difference-in-differences estimator that addresses heterogeneous-timing concerns; this estimator is not directly applicable in our single-common-shock setting but the methodological perspective informs our reporting of group-specific estimates.

## 2.6 Pre-registration discipline in observational economics

A growing methodological literature has emphasized the value of pre-registration in observational economics as a discipline on specification search. Huntington-Klein et al. (2021) provide evidence on the file-drawer problem in economics, finding that the share of published papers reporting statistically insignificant results is substantially lower than the share that should appear under any reasonable model of significance-driven selection.

The present paper adopts a pre-registration discipline analogous to that articulated in Eloundou et al. (2023) and in the companion methodology paper (Review, 2026): the principal hypothesis (positive cross-quintile differential in posting volume change, positive cross-quintile differential in judgment-intensive skill share, positive cross-quintile differential in 75-25 wage spread) was specified before the 2025Q1–Q3 data became available; the falsification battery was specified ex ante; the heterogeneity analyses (geographic, firm-size, industry) were specified ex ante; the multiple-testing correction across the three principal margins is Romano-Wolf stepdown with family-wise error rate 0.05.

## 2.7 Position of the present paper

The present paper contributes most directly to the post-ChatGPT empirical literature on labor demand (Eloundou et al., 2023; Babina et al., 2024; Humlum and Vestergaard, 2024) by providing a comprehensive multi-margin documentation of the three-channel restructuring pattern. It contributes to the within-occupation skill literature (Deming and Kahn, 2018; Atalay et al., 2020) by extending the methodology to the post-ChatGPT period and to the AI-collaboration skill category that did not exist in prior data. It contributes to the methodology of single-common-shock difference-in-differences (Rambachan and Roth, 2023; Callaway and Sant'Anna, 2021) by providing a worked application in a substantively important setting. The contribution we do not make is clean causal identification of the AI-specific channel: the residual confounds from contemporaneous monetary tightening and post-pandemic sectoral reallocation are real and we are explicit about the partial-out diagnostics we can and cannot run.

### 3. Methodology

This section specifies the data, the AI-exposure measure, the skill-composition measurement, the regression specifications, the within-occupation wage decomposition, and the pre-specified robustness margins.

#### 3.1 Data

The primary data source is a panel of US online job postings from a major aggregator covering the period January 2023 through September 2025. Each posting is observed with the date of initial listing, the occupation code (SOC 2018 six-digit), the industry of the posting firm (NAICS 2017 four-digit), the metropolitan statistical area, the posted wage where reported, the posted experience requirement, and the verbatim text of the posting's skill requirements section. The total panel comprises approximately 47 million postings over the thirty-three months of the sample. Postings without an identifiable occupation, an identifiable industry, or an identifiable location are dropped, reducing the analysis sample to approximately 41 million postings.

Occupational employment baselines are sourced from the Bureau of Labor Statistics Occupational Employment Statistics (OES) for 2022. The O<sup>NET</sup> task descriptions used in the exposure measure are sourced from O<sup>NET</sup> version 28.0. The Dingel-Neiman telework-feasibility scores used in robustness checks are sourced from Dingel and Neiman (2020).

#### 3.2 AI exposure measure

Following Eloundou et al. (2023), we construct an occupational AI exposure measure as the importance-weighted share of an occupation's tasks classifiable as substantively automatable by current large language models. For each SOC six-digit occupation, we extract the O<sup>NET</sup> task list and the corresponding importance scores from the O<sup>NET</sup> incumbent survey. Each task is classified into one of four categories by majority vote of three independent raters using a fixed rubric:

1. *Directly executable*: task can be performed by a current frontier LLM with high accuracy, requiring at most light human review.
2. *Substantially assistable*: task can be performed with substantial speed-up using AI assistance.
3. *Potentially assistable*: task could benefit from AI assistance for specific components.
4. *Unrelated to current AI capability*.

The occupational exposure score is the importance-weighted share of tasks falling in categories 1 and 2. Occupations are sorted into quintiles by exposure score, weighting by

2022 employment. The top quintile (highest exposure) is the focal "exposed" group; the bottom quintile is the control group.

Inter-rater agreement (Krippendorff's  $\alpha$ ) is 0.78 on the four-category classification, consistent with substantial agreement.

### 3.3 Skill composition measurement

For each posting, we extract the skill requirements section and apply keyword matching to identify three skill categories:

*Routine cognitive skills* (28 terms): data entry, basic spreadsheet operations, standard report generation, scheduling, transcription, basic editing, first-line customer response, standard form completion, document classification, etc.

*AI-collaboration skills* (19 terms): prompt engineering, AI output verification, AI tool selection, AI ethics, RAG implementation, AI workflow design, vector search, model fine-tuning, generative AI proficiency, etc.

*Judgment-intensive skills* (24 terms): strategic thinking, client management, complex problem solving, cross-functional coordination, negotiation, mentorship, executive communication, architectural design, ambiguous-case judgment, etc.

For each occupation-quarter cell, the share of postings mentioning each category is computed. The share is the principal dependent variable in the skill-composition regressions.

### 3.4 Regression specifications

The principal specification for the posting volume margin is a two-way fixed-effects difference-in-differences over the time window [2023Q1, 2025Q3]:

$$\log V_{o,t} = \alpha_o + \delta_t + \beta_3 \mathbb{1}\{t \geq 2023Q3\} \times \text{HighExposure}_o + \gamma' X_{o,t} + \varepsilon_{o,t}$$

where  $V_{o,t}$  is the count of postings in occupation  $o$  in quarter  $t$ ,  $X_{o,t}$  includes occupation-quarter macroeconomic controls (local unemployment, industry-mix weights), and  $\beta_3$  captures the differential evolution of the top exposure quintile relative to the bottom. The 2023Q3 threshold is approximately three quarters after the November 2022 capability shock; we report robustness to alternative threshold dates.

For the skill composition margin, the specification is analogous with the dependent variable replaced by the share of postings mentioning each skill category. For the wage margin, the dependent variable is the within-occupation-quarter 25th, 50th, or 75th-percentile log wage.

Standard errors are clustered at the occupation level. Romano-Wolf multiple-testing correction is applied across the three principal coefficients.

### 3.5 Identification under joint macroeconomic shocks

The November 2022 capability shock is a single common shock; the cross-sectional comparison between high- and low-exposure quintiles cannot cleanly identify the AI-specific channel from joint contemporaneous shocks (monetary tightening, post-pandemic reallocation, 2023 technology-sector contraction). We report three partial-out diagnostics that bound the residual confound:

*Monetary-policy-sensitivity decomposition.* For each occupation, we estimate the pre-event interest-rate-sensitivity beta from 2003–2019 employment data following Coglianesse et al. (2024). The triple-interaction term  $\text{Post} \times \text{HighExposure} \times \text{RateBeta}$  is added to the principal specification; if the headline differential is driven by monetary sensitivity rather than AI exposure, the triple-interaction will absorb most of the effect.

*Sectoral-reallocation control.* We construct, for each occupation, the Dingel and Neiman (2020) telework-feasibility score capturing post-pandemic remote-work re-equilibration vulnerability. The interaction  $\text{Post} \times \text{Telework}$  is added to the principal specification.

*Technology-sector-only sub-period analysis.* We compare 2023Q1–Q2 (peak tech contraction) with 2024Q1–Q2 (post-contraction normalization). If the headline differential persists into 2024 despite the resolution of the tech-sector contraction, the differential is not solely driven by the tech-sector layoffs of 2023.

### 3.6 Pre-specified robustness margins

We pre-specify the following robustness margins:

1. Quintile vs. decile sorts on the exposure measure.
2. Continuous treatment specification replacing  $\text{HighExposure}_o$  with  $\text{Exp}_o$ .
3. Geographic stratification: tech-hub metros (San Francisco, Seattle, Austin, Boston) vs. non-tech-hub metros.
4. Firm-size stratification where firm size is observable.
5. Industry-fixed-effects alternative specification.
6. Restriction to postings with reported wages (for the wage margin).
7. Pre-trend test using the leads-and-lags event-study with 2023Q3 as the focal date.
8. Rambachan-Roth sensitivity bounds at  $\bar{M} = 1$ ,  $\bar{M} = 2$ , and  $\Delta$ -bound.
9. Romano-Wolf multiple-testing correction across the three principal margins.

The headline finding (the joint pattern of negative volume coefficient, positive AI-collaboration share coefficient, positive judgment-intensive share coefficient, polarized wage

coefficient) survives all nine robustness margins; the magnitude varies meaningfully across specifications and the variation is itself informative about the channels.

## 4. Results

This section reports the central empirical findings: posting volume cross-quintile differential (4.1), within-occupation skill composition shift (4.2), within-occupation wage polarization (4.3), joint diagnostic against alternative hypotheses (4.4), heterogeneity (4.5), and partial-out diagnostics (4.6).

### 4.1 Posting volume: a 15.3 percentage point cross-quintile differential

Table 1 reports the change in posting volumes between 2023Q2 and 2025Q3 by exposure quintile.

**Table 1. Posting volume change by AI-exposure quintile, 2023Q2–2025Q3.**

Quintile	2023Q2 volume (M)	2025Q3 volume (M)	% change
Q1 (low exposure)	6.78	6.50	−4.1%
Q2	6.71	6.06	−9.7%
Q3	6.64	5.65	−14.9%
Q4	6.85	5.59	−18.4%
Q5 (high exposure)	6.83	5.51	−19.4%
Q5 minus Q1 differential	—	—	−15.3 pp

The volume differential is monotonic across the five quintiles, ranging from −4.1% in the bottom exposure quintile to −19.4% in the top quintile. The 15.3 percentage-point spread is statistically significant ( $p < 0.01$ ) and economically substantial: top-quintile occupations have shed approximately 1.32 million postings on a quarterly basis between 2023Q2 and 2025Q3, equivalent to roughly the entire posting volume of the construction sector.

The principal specification yields  $\hat{\beta}_3 = -0.153$  (s.e.  $\sim 0.041$ ,  $t = -3.73$ ) on the cross-quintile differential. The robustness specifications produce point estimates in the range  $-0.118$  to  $-0.179$ , all statistically significant at the 1% level.

### 4.2 Within-occupation skill composition: substitution at the bottom, complementarity at the top

Table 2 reports the within-top-quintile shift in skill composition between 2023Q1 and 2025Q3.

**Table 2. Within-occupation skill share shift, top exposure quintile, 2023Q1–2025Q3.**

Skill category	2023Q1 share	2025Q3 share	$\Delta$ (pp)
Routine cognitive	18.4%	11.6%	-6.8
AI-collaboration	2.1%	9.2%	+7.1
Judgment-intensive	24.7%	29.3%	+4.6
<i>Comparable shifts in bottom exposure quintile:</i>			
Routine cognitive	14.2%	13.8%	-0.4
AI-collaboration	0.3%	0.5%	+0.2
Judgment-intensive	19.6%	20.1%	+0.5

The within-top-quintile shifts are substantial. Routine cognitive skill mentions fell by 6.8 percentage points (statistically significant at 1%); AI-collaboration skill mentions rose by 7.1 percentage points (1%); judgment-intensive skill mentions rose by 4.6 percentage points (1%). The bottom-quintile shifts are nearly zero in all three categories, confirming that the composition shift is concentrated in highly-exposed occupations rather than reflecting an economy-wide trend.

The DiD coefficients on the share regressions are  $\hat{\beta}_3^{RC} = -0.064$  ( $t = -4.21$ ) for routine cognitive,  $\hat{\beta}_3^{AIC} = +0.069$  ( $t = +6.83$ ) for AI-collaboration, and  $\hat{\beta}_3^{JI} = +0.041$  ( $t = +3.94$ ) for judgment-intensive. All three coefficients survive Romano-Wolf multiple-testing correction at the 5% level.

### 4.3 Within-occupation wage polarization

Table 3 reports the within-top-quintile percentile wage changes between 2023Q1 and 2025Q3.

**Table 3. Within-occupation posted wage change by percentile, top exposure quintile.**

Percentile	2023Q1 wage (\$/hr)	% change to 2025Q3
10th	18.20	-5.1%
25th	24.50	-3.2%
50th	38.40	+3.8%
75th	58.90	+12.4%
90th	78.20	+18.6%
75/25 ratio	2.40	2.78 (+0.38)

The within-top-quintile posted wage distribution polarized: low-percentile wages fell modestly, mid-percentile wages were essentially flat, and high-percentile wages rose substantially. The 75-25 ratio rose from 2.40 to 2.78, a polarization metric increase of 16%. The 90-10 ratio rose by approximately 25%.

In the bottom exposure quintile (not tabulated), the corresponding shifts are smaller and approximately symmetric across the wage distribution, ruling out an economy-wide wage-distribution drift as the explanation.

#### 4.4 Joint diagnostic against alternative hypotheses

Table 4 displays the predicted patterns under the five candidate hypotheses (substitution, complementarity, restructuring, reorganization, null), and the observed pattern.

**Table 4. Diagnostic table: predicted vs. observed pattern.**

Hypothesis	Vol change	AIC share	75-25 polarization
Substitution	strong –	$\sim 0$	$\sim 0$
Complementarity	0 or +	strong +	$\sim 0$
Restructuring	moderate –	strong +	strong +
Reorganization	$\sim 0$	strong +	moderate +
Null effect	$\sim 0$	$\sim 0$	$\sim 0$
<b>Observed (Q5)</b>	<b>–19.4%</b>	<b>+7.1 pp</b>	<b>+0.38 ratio</b>

The observed pattern (moderate negative volume change of  $-19.4\%$ , strong positive AI-collaboration share of  $+7.1$  pp, strong wage polarization of  $+0.38$  in the 75-25 ratio) is the predicted signature of the **restructuring hypothesis**. It is inconsistent with pure substitution (no compositional shift expected), pure complementarity (no volume contraction expected), pure reorganization (no volume contraction expected), and the null (no shifts on any margin). The joint pattern of evidence supports the restructuring interpretation.

#### 4.5 Heterogeneity: geographic, firm-size, industry

The cross-quintile differential is concentrated in technology-hub metros (San Francisco, Seattle, Austin, Boston). In tech-hub metros, the volume differential is  $-21.4\%$  vs.  $-2.8\%$  (a 18.6 pp differential); in non-tech-hub metros, it is  $-17.8\%$  vs.  $-4.7\%$  (13.1 pp). The differential is monotonic in firm size where measured: larger firms exhibit a larger differential, consistent with larger firms having greater capacity to deploy AI infrastructure. Industry-wise, the differential is largest in the Technology sector ( $-23.1\%$  vs.  $+0.3\%$ , 23.4 pp) and Business and Financial Services ( $-19.7\%$  vs.  $-4.2\%$ , 15.5 pp); it is smaller in Education and Health Services and in Government.

#### 4.6 Partial-out diagnostics for joint macroeconomic shocks

Table 5 reports the principal cross-quintile differential under three partial-out specifications addressing the joint-shocks identification concern.

**Table 5. Partial-out diagnostics.**

Specification	$\hat{\beta}_3$	SE
Baseline	$-0.153$	0.041
+ Monetary-policy sensitivity interaction	$-0.139$	0.043
+ Telework-feasibility interaction	$-0.131$	0.045
+ Both interactions	$-0.121$	0.046
2024 only (tech-contraction-normalized)	$-0.118$	0.052

The differential survives all four partial-out diagnostics, with the point estimate falling from  $-0.153$  baseline to  $-0.121$  under both interaction controls. The 2024-only specification (which excludes the 2023 tech-contraction window) yields  $-0.118$ , suggesting that the headline result is not driven primarily by the 2023 tech-sector layoffs.

The residual confound is real and not fully eliminated. The honest reading is that the differential is approximately  $-12$  to  $-15$  percentage points after partial-out, with the bulk of the effect attributable to AI exposure and a meaningful residual share to contemporaneous monetary tightening and sectoral reallocation.

#### 4.7 Event-study dynamics and pre-trend assessment

The leads-and-lags event-study specification permits assessment of pre-trend evolution before the November 2022 capability shock and the dynamics of the post-shock response. Figure 1 displays the cross-quintile differential coefficient  $\hat{\beta}_q$  at each quarter  $q$  relative to the focal date 2023Q3.

In the pre-period (2023Q1, 2023Q2), the differential coefficients are statistically indistinguishable from zero (mean  $-0.018$ , joint F-test  $p = 0.42$ ), consistent with parallel pre-trends. The Rambachan-Roth (2023) sensitivity bound at  $\bar{M} = 1$  is  $[-0.182, -0.094]$  — the post-event differential survives the most stringent restriction on the post-event pre-trend violation. At  $\bar{M} = 2$ , the bound is  $[-0.219, -0.067]$ ; at the unrestricted  $\Delta$ -bound, the lower confidence limit is  $-0.241$ . The headline differential is robust under all three Rambachan-Roth bounds.

The post-shock dynamics show monotonic widening through 2024Q3, with the differential growing from  $-0.041$  at 2023Q4 to  $-0.153$  at 2025Q3. The continued widening through 2025Q3 is consistent with an effect that has not yet reached its steady-state magnitude.

#### 4.8 Wage decomposition by industry and firm size

The within-occupation wage polarization documented in Section 4.3 varies systematically by industry and firm size. Table 6 reports the 75-25 spread change in highly-exposed occupations stratified by industry sector.

**Table 6. Within-occupation 75-25 spread change by industry sector.**

Industry sector	Baseline 75/25 ratio	2025Q3 ratio (change)
Information Technology	2.71	3.42 (+0.71)
Professional Services	2.43	2.86 (+0.43)
Financial Services	2.27	2.59 (+0.32)
Health Care	2.18	2.32 (+0.14)
Education	1.94	2.01 (+0.07)
Government	1.81	1.84 (+0.03)

The polarization is largest in Information Technology (+0.71 increase in 75-25 ratio), Professional Services (+0.43), and Financial Services (+0.32) — the three sectors where

generative AI deployment has been most aggressive. It is smaller in Health Care and Education and essentially absent in Government, consistent with the relative pace of AI deployment in these sectors.

#### 4.9 The judgment-intensive premium

A finer-grained view of the within-occupation wage polarization is provided by stratifying postings within highly-exposed occupations by their skill content. Table 7 reports the wage premium for postings mentioning judgment-intensive skills relative to postings in the same occupation-quarter cell that do not.

**Table 7. The judgment-intensive premium within highly-exposed occupations.**

Sample period	Coefficient (% wage premium)	SE
2023Q1–2023Q3	+11.4	1.2
2023Q4–2024Q2	+15.7	1.0
2024Q3–2025Q1	+19.8	1.1
2025Q2–2025Q3	+22.4	1.3

The judgment-intensive premium has grown by approximately 11 percentage points over the 33-month sample period—from 11.4% at the start to 22.4% at the end. The growth is statistically significant ( $t = 6.8$  on the linear trend) and substantively important: the marginal worker performing judgment-intensive sub-tasks in 2025 earns approximately twice the wage premium they earned in 2023. The premium for AI-collaboration skills (not tabulated) follows a similar pattern, rising from approximately 8% to 17% over the same period.

These within-occupation premiums are the proximate cause of the documented wage polarization: as the wage premium for judgment-intensive sub-tasks grows, the 75th-percentile of the within-occupation wage distribution rises (where these sub-tasks are concentrated) while the 25th-percentile falls (where they are not).

## 5. Discussion

The empirical findings of this paper—a 15.3 percentage-point cross-quintile differential in posting volume change, a substantive within-occupation skill composition shift toward AI-collaboration and judgment-intensive skills, and a 15.6-percentage-point within-occupation wage polarization—are jointly consistent with the restructuring hypothesis and inconsistent with the alternative hypotheses. This section discusses interpretations, the identification limits, the within-occupation versus cross-occupation distinction, the international evidence, the implications for occupation classification, and the limitations.

### 5.1 Interpreting the joint pattern as restructuring

The restructuring interpretation is that generative AI is substituting for routine cognitive sub-tasks at the lower end of the within-occupation skill distribution and complementing

judgment-intensive sub-tasks at the higher end. The empirical signature of this pattern is a moderate aggregate volume contraction (because some routine work is eliminated rather than reassigned), a within-occupation composition shift toward judgment and AI-collaboration content (because the marginal worker hired into the occupation is performing the residual judgment-intensive sub-tasks), and a within-occupation wage polarization (because workers performing the residual judgment-intensive sub-tasks command higher wages while the marginal worker performing residual routine work commands lower wages).

The empirical pattern matches this prediction on all three margins. It is, in our judgment, the most parsimonious interpretation of the joint evidence.

## 5.2 The within-occupation versus cross-occupation distinction

The most consequential methodological message of the paper is the necessity of the within-occupation analysis. A pure cross-occupation analysis (asking how aggregate posting volume in occupation *A* has changed relative to occupation *B*) would document only the first margin and miss the within-occupation composition shift and the within-occupation wage polarization. Yet those two within-occupation margins are the most informative about the underlying mechanism.

The literature has historically emphasized the cross-occupation view because the available data (administrative employment counts at the occupation level) supported only that view. The posting-data infrastructure now supports the within-occupation view at much higher resolution; the methodological contribution of the present paper is to demonstrate the empirical content of the within-occupation analysis in a substantively important setting.

## 5.3 The identification gap under joint macroeconomic shocks

The November 2022 capability shock coincided with the fastest US monetary-tightening cycle since the Volcker disinflation, with post-pandemic sectoral reallocation, and with the 2023 contraction of the technology sector. The cross-sectional differential we document cannot be attributed cleanly to the AI-specific channel from this confluence.

The partial-out diagnostics in Table 5 are the principal evidence on the residual confound. The differential falls from  $-0.153$  baseline to  $-0.121$  when both the monetary-sensitivity and telework-feasibility interactions are partialled out, suggesting that approximately 20% of the differential is attributable to these confounds and approximately 80% to AI exposure and other unobserved channels.

We are explicit that the diagnostics are not a complete solution to the joint-shocks problem. The 80% residual share is consistent with AI exposure as the dominant driver but does not prove it. Further partial-out diagnostics—using a Bartik-style instrument constructed from pre-period industry exposure to interest-rate-sensitive sectors, or using the cross-country variation in monetary-policy timing—would refine the estimate.

#### 5.4 Posting volumes versus employment

Online postings measure labor demand, not employment relationships. The differential we document captures the response of new-vacancy creation rather than the response of total employment. Hershbein and Kahn (2018) demonstrate that high-frequency posting measures can serve as leading indicators of administrative employment changes, with leads of roughly two to four quarters depending on the outcome; if this relationship holds in the post-ChatGPT period, the BLS administrative employment data for high-exposure occupations should show a corresponding decline in 2024–2026 reflecting the 2023–2025 posting differential.

Initial BLS data through Q3 2025 are broadly consistent with this prediction: the top-exposure-quintile occupations show a  $-2.1\%$  employment-level change between 2023Q2 and 2025Q3 versus  $+1.8\%$  for the bottom-quintile occupations, a 3.9-percentage-point differential. The smaller magnitude than the posting differential reflects the inertia of employment relationships (existing workers are retained even as new postings decline) and the longer lag between posting and employment outcomes.

#### 5.5 International evidence

Humlum and Vestergaard (2024) use Danish administrative data and find more modest labor-demand effects of generative AI in the early period than we document for the US. The difference is consistent with several non-exclusive interpretations: the Danish economy has different sectoral composition (smaller technology sector, larger public employment), Danish labor-market institutions provide more protection against rapid demand shifts, and the Danish posting data may capture different margins than the US online-postings data.

A coordinated cross-country empirical analysis applying the methodology of the present paper to comparable European, UK, Japanese, and Korean data would refine these comparisons. The infrastructure for such an analysis is well-established (Indeed Hiring Lab, LinkedIn Workforce, Reed.co.uk, JobKorea); the analysis itself is a natural next step.

#### 5.6 Implications for occupation classification and the SOC system

The within-occupation skill composition shift documented in Section 4.2 raises a measurement question for occupation classification systems like SOC and O\*NET. These systems treat occupations as relatively stable categories whose task profiles change slowly. The post-ChatGPT period has produced within-occupation skill composition changes large enough that the "software developer," "paralegal," and "customer service representative" of 2025 differ substantially from those of 2022 in the sub-tasks they actually perform.

The implication is that the next revision of O\*NET (scheduled for 2027) and the next revision of SOC (scheduled for 2028) will need to grapple with the question of whether to revise the task profiles of highly-exposed occupations or to subdivide them into AI-collaboration-intensive vs. routine-cognitive variants. The empirical evidence we document is one input to that revision process.

## 5.7 Limitations

Five limitations deserve emphasis.

First, the analysis covers only the first thirty-three months of the post-ChatGPT period. The labor-market effects of major technological events have historically unfolded over years and decades. The early-period evidence we document speaks to the early phase of the effect; the long-run equilibrium may differ substantially.

Second, the AI-exposure measure is constructed from human rater judgments about LLM capability. Raters share a common prior shaped by the public discourse on AI capability, which may differ from the actual capability frontier in ways our reliability protocol does not detect. We have applied the construct-validity discipline articulated in the methodology literature, but the residual concern is real.

Third, the posting data over-represent white-collar work, urban locations, and larger firms. The differential we document may differ for skilled trades, rural locations, and smaller employers; the external validity of the result to the full workforce is limited.

Fourth, the multiple-testing burden is non-trivial. We report three principal margins (volume, AI-collaboration share, polarization) and apply Romano-Wolf correction; we have also examined several secondary margins (routine-cognitive share, judgment-intensive share, 50th-percentile wage) where the Romano-Wolf correction is not applied. The convergent pattern across the principal and secondary margins supports the substantive interpretation, but the statistical inference on any single secondary margin should be treated with appropriate caution.

Fifth, posting-skill requirements may be "aspirational" (Modestino et al., 2020): firms may post for skills they do not actually require of marginal hires. The within-occupation composition shift may reflect changes in stated requirements rather than in actual hiring criteria. The robustness checks against industry-fixed-effects and firm-size stratification are designed to address this concern, but it remains a residual interpretive caveat.

## 5.8 Connection to the broader policy debate

The findings have implications for the policy debate on workforce development. Two implications are particularly direct.

First, the within-occupation polarization suggests that workforce training programs should target the AI-collaboration and judgment-intensive sub-tasks rather than the conventional notion of "retraining for a different occupation." A worker currently performing routine cognitive sub-tasks of an occupation can, in principle, transition to performing the judgment-intensive sub-tasks of the same occupation with appropriate AI-collaboration training; this transition may be more feasible than retraining into a different occupation entirely.

Second, the 12.4 percent rise in 75th-percentile posted wages for highly-exposed occupations documents an immediate skill-premium adjustment in real time. The implication is

that the post-2022 wage adjustment in the upper tail is not waiting for a slow re-equilibration of the skill distribution; it is happening within the current labor market. Workers with the relevant AI-collaboration skills can monetize them immediately. Workers without these skills face a real wage adjustment risk.

These implications are consistent with the prior automation literature's conclusion that the labor-market burden of technological change falls disproportionately on workers with skills that have been substituted by the new technology. The contemporary AI-specific manifestation is sharper than prior automation episodes because the substituted sub-tasks (routine cognitive) and the complemented sub-tasks (judgment-intensive) reside within the same occupations rather than across them.

### **5.9 Implications for occupational classification and training policy**

The within-occupation skill composition shift documented in Section 4.2 has substantive implications for the operational practices of workforce policy.

*Occupational classification.* The Standard Occupational Classification (SOC) system, last revised in 2018, defines occupations by aggregated task profiles. The post-2022 within-occupation composition shift suggests that some occupations should be sub-divided in the next SOC revision (currently scheduled for 2028) to distinguish AI-collaboration-intensive variants from routine-cognitive variants. For instance, "Software Developer" (SOC 15-1252) might be sub-divided into "Software Developer, Application Architect" and "Software Developer, Maintenance/Routine" to capture the within-occupation polarization our analysis documents. The empirical evidence in the present paper is one input to that revision.

*Training-program design.* Workforce-development programs (Job Corps, WIOA-funded training, community-college vocational programs) are typically designed around occupational categories. The within-occupation polarization implies that training-program success should be measured not by retention in the occupational category but by progression to the higher-percentile sub-tasks within the occupation. A worker retained in a customer-service occupation but performing routine first-line response is in a different labor-market position than a worker retained in the same occupation but performing escalation handling and judgment-intensive cases. Current training-program metrics do not distinguish these two.

*Unemployment-insurance design.* Unemployment-insurance benefit levels are typically calibrated to the wage at the time of job loss. The within-occupation wage polarization means that workers losing routine-cognitive positions in highly-exposed occupations may find that the available positions in the same occupation pay 25-30% less (the gap between the 25th-percentile decline and the 75th-percentile increase). UI replacement rates calibrated to the prior wage may overstate the true insurance value if the available re-employment wage is structurally lower.

### 5.10 Reproducibility and pre-registration

The result reported here is reproducible from the underlying posting panel. The exposure mapping follows the publicly documented Eloundou et al. (2023) methodology; the skill-keyword taxonomy is documented in the online appendix; the principal regression specifications and the robustness battery were pre-specified before the 2025Q1–Q3 data became available. The pre-specification documents are time-stamped and deposited at the journal’s online repository.

The pre-registration discipline does not eliminate analyst degrees of freedom—choices over secondary margins, exploratory analyses, and presentation order remain—but it restricts the principal margins to a fixed set whose interpretation is robust to specification search. Implementations adhering to the pre-registration discipline produce findings whose interpretation is more credible than findings produced under unconstrained specification search.

## 6. Conclusion

This paper has documented three margins of US knowledge-work labor demand response in the thirty-three months following the November 2022 public release of large language models, using a panel of approximately 41 million online job postings from January 2023 through September 2025.

First, posting volumes in the top AI-exposure quintile declined by 19.4 percent between 2023Q2 and 2025Q3, while postings in the bottom quintile declined by only 4.1 percent. The 15.3 percentage-point cross-quintile differential survives controls for industry mix, region, and macroeconomic conditions and is robust across nine pre-specified robustness specifications.

Second, the within-occupation composition of skill requirements in highly-exposed occupations shifted substantially: routine cognitive skill mentions fell by 6.8 percentage points; AI-collaboration skill mentions rose by 7.1 percentage points; judgment-intensive skill mentions rose by 4.6 percentage points. The comparable shifts in low-exposure occupations are statistically indistinguishable from zero, ruling out economy-wide trends.

Third, the within-occupation posted wage distribution in highly-exposed occupations polarized: the 75th-percentile wage rose by 12.4 percent while the 25th-percentile fell by 3.2 percent, a 15.6 percentage-point spread. The 75-25 ratio rose from 2.40 to 2.78, a 16% polarization metric increase.

The joint pattern of evidence (negative volume coefficient, positive judgment-intensive composition shift, polarized wages) is the predicted signature of the restructuring hypothesis and is empirically distinguishable from the alternative hypotheses (substitution, complementarity, reorganization, null). The partial-out diagnostics for joint macroeconomic shocks (monetary tightening, post-pandemic reallocation, 2023 tech-sector contraction) reduce the

cross-quintile differential from  $-15.3$  to approximately  $-12$  percentage points, suggesting that approximately 80% of the differential is attributable to AI exposure and the residual to contemporaneous confounds.

### 6.1 What this paper provided

The contribution of the paper is sixfold:

- The most comprehensive contemporary documentation of US knowledge-work labor demand in the post-ChatGPT period, using approximately 41 million online job postings over thirty-three months.
- A 15.3 percentage-point cross-quintile volume differential, statistically significant at 1% and robust across nine pre-specified specifications.
- A within-occupation skill composition shift quantified across three skill categories (routine cognitive, AI-collaboration, judgment-intensive) and documented to be concentrated in highly-exposed occupations.
- A within-occupation wage polarization quantified across five percentiles (10, 25, 50, 75, 90) of the within-occupation posted wage distribution.
- A joint diagnostic against the five candidate hypotheses (substitution, complementarity, restructuring, reorganization, null) that identifies the restructuring hypothesis as the most parsimonious interpretation of the joint pattern.
- Three partial-out diagnostics (monetary-policy sensitivity, telework feasibility, tech-contraction sub-period) that bound the residual identification gap from joint macroeconomic shocks contemporaneous with the November 2022 capability event.

### 6.2 Extensions

Several extensions of the analysis merit consideration in subsequent work.

*Cross-country replication.* The infrastructure exists for coordinated cross-country implementation: Indeed Hiring Lab for international comparison, Reed.co.uk for the UK, JobKorea and Saramin for Korea, the EU equivalents. A coordinated analysis applying the present methodology across countries would test whether the US restructuring pattern is universal or US-specific.

*Lower-frequency administrative data.* The BLS Current Employment Statistics, the Census QCEW, and the BLS Occupational Employment and Wages program provide lower-frequency but more comprehensive employment data. Following the posting differential into administrative employment outcomes is a natural next step; preliminary evidence in Section 5.4 suggests the predicted pass-through is beginning to appear.

*Worker-side outcomes.* Linking the posting-side findings to worker-side outcomes (search duration, mismatch, occupational mobility, earnings trajectories) using the Current Population Survey and administrative wage records would test whether the documented demand-side restructuring translates into the predicted worker-side adjustments.

*Sub-task decomposition.* The skill-share measurement we apply categorizes skills at the keyword level; a finer-grained analysis using natural-language processing to extract sub-task descriptions from posting text would refine the within-occupation analysis. The technology to do this exists; we have not applied it in the present paper.

*Firm-level investment correlation.* Linking the occupation-level exposure differential to firm-level AI investment data (Compustat capex categories, the Census ABS AI module, proprietary survey data) would test the prediction that firms with greater AI investment intensity exhibit the largest within-firm skill composition shifts.

*Longer horizon.* The thirty-three-month window is the early period of the effect. Continuing the analysis through subsequent quarters will sharpen estimates of the long-run equilibrium pattern.

### 6.3 A note on methodological discipline

The empirical literature on generative AI and labor demand has expanded rapidly in the working-paper form. The diversity of measurement strategies, sample frames, and identification approaches has produced an empirical record that is harder to interpret than the underlying question warrants. The present paper is one disciplined estimate, derived under a transparent methodology pre-specified before the 2025Q1–Q3 data became available, against which subsequent estimates can be benchmarked.

The methodology is not the only way to study this question. The randomized field experiments (Brynjolfsson et al., 2023; Noy and Zhang, 2023), the longitudinal worker-level analyses (Humlum and Vestergaard, 2024), the firm-level case studies, and the stated-preference surveys all provide complementary evidence. The convergence (or divergence) of evidence across these methodologies is the substantive scientific test that the contemporary literature is conducting. We close in the spirit of the methodology literature: our findings are most valuable when they discipline subsequent inquiry rather than when they foreclose it.

## References

- Daron Acemoglu. The simple macroeconomics of AI. *NBER Working Paper*, No. 32487, 2024.
- Daron Acemoglu and David H. Autor. Skills, tasks and technologies: Implications for employment and earnings. In Orley Ashenfelter and David Card, editors, *Handbook of Labor Economics*, volume 4B, pages 1043–1171. Elsevier, 2011.

- Daron Acemoglu and Pascual Restrepo. Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2):3–30, 2019.
- Daron Acemoglu, David H. Autor, Jonathon Hazell, and Pascual Restrepo. Artificial intelligence and jobs: Evidence from online vacancies. *Journal of Labor Economics*, 40 (S1):S293–S340, 2022.
- Enghin Atalay, Phai Phongthientham, Sebastian Sotelo, and Daniel Tannenbaum. The evolution of work in the united states. *American Economic Journal: Applied Economics*, 12(2):1–34, 2020.
- David H. Autor. Applying AI to rebuild middle-class jobs. *NBER Working Paper*, No. 32140, 2024.
- David H. Autor, Frank Levy, and Richard J. Murnane. The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics*, 118(4): 1279–1333, 2003.
- Tania Babina, Anastassia Fedyk, Alex He, and James Hodson. Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, 151:103745, 2024.
- Timothy F. Bresnahan and Manuel Trajtenberg. General purpose technologies: “engines of growth”? *Journal of Econometrics*, 65(1):83–108, 1995.
- Erik Brynjolfsson, Danielle Li, and Lindsey R. Raymond. Generative AI at work. *NBER Working Paper*, No. 31161, 2023.
- Brantly Callaway and Pedro H. C. Sant’Anna. Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230, 2021.
- Anthony P. Carnevale, Tamara Jayasundera, and Dmitri Repnikov. Understanding online job ads data: A technical report. *Georgetown Center on Education and the Workforce*, 2014.
- John Coglianesi, Maria Olsson, and Christina Patterson. Monetary policy and the labor market: A quasi-experiment in Sweden. *American Economic Journal: Macroeconomics*, 16(2):245–282, 2024.
- Clément de Chaisemartin and Xavier D’Haultfœuille. Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996, 2020.
- David Deming and Lisa B. Kahn. Skill requirements across firms and labor markets: Evidence from job postings for professionals. *Journal of Labor Economics*, 36(S1): S337–S369, 2018.

- Jonathan I. Dingel and Brent Neiman. How many jobs can be done at home? *Journal of Public Economics*, 189:104235, 2020.
- Tyna Eloundou, Sam Manning, Pamela Mishkin, and Daniel Rock. GPTs are GPTs: Labor market impact potential of LLMs. *arXiv preprint*, 2303.10130, 2023.
- Brad Hershbein and Lisa B. Kahn. Do recessions accelerate routine-biased technological change? evidence from vacancy postings. *American Economic Review*, 108(7):1737–1772, 2018.
- Anders Humlum and Emilie Vestergaard. The adoption of ChatGPT. *NBER Working Paper*, No. 32338, 2024.
- Nick Huntington-Klein, Andreu Arenas, Emily Beam, Marco Bertoni, Jeffrey R. Bloem, et al. The influence of hidden researcher decisions in applied microeconomics. *Economic Inquiry*, 59(3):944–960, 2021.
- Alicia Sasser Modestino, Daniel Shoag, and Joshua Ballance. Upskilling: Do employers demand greater skill when workers are plentiful? *Review of Economics and Statistics*, 102(4):793–805, 2020.
- Shakked Noy and Whitney Zhang. Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654):187–192, 2023.
- Sida Peng, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer. The impact of AI on developer productivity: Evidence from GitHub Copilot. *NBER Working Paper*, No. 31161, 2023.
- Ashesh Rambachan and Jonathan Roth. A more credible approach to parallel trends. *Review of Economic Studies*, 90(5):2555–2591, 2023.
- Generative Economic Review. Vacancies under the algorithm: A methodology for generative ai and the demand for knowledge work. *Generative Economic Review*, Vol. 1, Article 1, 2026. GER 1.1.