Exposure Is Not Adoption: Measuring Generative AI and Labor Demand

Ashish K. Sedai*

September 2026

Abstract

Occupational exposure indices measure what generative AI could do, not what it is used for. I build a behavioral measure of realized AI adoption from the Anthropic Economic Index and compare it with the standard exposure index across 47 occupational sectors. Adoption and exposure rank occupations alike but diverge in level, because adoption is far more concentrated. In weekly Indeed job postings after ChatGPT, labor demand tracks adoption rather than exposure: a one standard deviation increase in adoption predicts a decline of about 10 log points, while exposure adds nothing. Realized adoption captures AI's early labor market footprint.

JEL codes: J23, J24, O33.

Keywords: innovation, artificial intelligence, labor demand, occupational exposure, job postings.

*University of Texas at Arlington Email: ashish.sedai@uta.edu. I am grateful to Robert Seamans, Aaron Chatterji, Alex Tamkin and the the Felten, Raj, Seamans team for their guidance with the question and the data. I thank, without implicating, seminar participants and colleagues for comments. All errors are mine.

1 Introduction

Forecasts of how generative artificial intelligence will affect work rest largely on measures of *exposure*, which show the share of an occupation's tasks that a model could perform, scored from task descriptions before the technology was used (Felten et al., 2021; Eloundou et al., 2024; Webb, 2020; Felten et al., 2023). So far, the exposure indices have anchored academic research, the workforce assessments of international institutions (Cazzaniga et al., 2024), and the planning of firms and governments. *However, exposure measures what a technology could do. Whether the potential is taken up, and where, is a separate matter*, and until recently it went unmeasured, because no data recorded how generative AI was actually used across the economy.

The Anthropic Economic Index (Appel et al., 2025) now records it. It classifies a large sample of conversations with a frontier model against the roughly nineteen thousand tasks of the Department of Labor O*NET taxonomy[1] and reports, for each task, the share of usage. Aggregated to occupations, the shares yield a measure of realized adoption, the intensity with which an occupation's work is actually carried out with the technology, which can be set beside exposure and compared with it directly.

To my knowledge, this paper is the first to use the Anthropic Economic Index and measure both exposure and adoption, and relate each to the change in labor demand after ChatGPT. I pair the adoption measure with the AI Occupational Exposure index of Felten et al. (2021) and with the Indeed Hiring Lab index of job postings by occupational sector, a balanced weekly panel of 47 sectors and about 12,000 sector weeks from February 2020 through December 2025 (Indeed Hiring Lab, 2026). Job postings move within weeks, where employment and wages take quarters, which suits them to dating a change around the release. All three series are public and the analysis is fully replicable.

[1] The Department of Labor "O*NET system is maintained by a regularly updated database of occupational characteristics and worker requirements information across the U.S. economy. It describes occupations in terms of the knowledge, skills, and abilities required as well as how the work is performed in terms of tasks, work activities, and other descriptors" (U.S. Department of Labor, Employment and Training Administration, 2026).

First, I show that exposure and adoption, often used interchangeably, are distinct: exposure is the potential for adoption and adoption is its realization. They order occupations alike, with a rank correlation of 0.76, so exposure is not wrong about the ranking. Their level correlation is only 0.25, because adoption is concentrated where exposure is smooth: the Gini coefficient of adoption across the 47 sectors is 0.76 against 0.25 for exposure, and the five most adopted sectors, led by software development, hold 68 percent of measured usage against 16 percent for the five most exposed.

Second, labor demand has moved with adoption. In a difference in differences specification that interacts each measure with an indicator for the period after the release, adoption predicts the fall in job postings and exposure does not. A one standard deviation increase in adoption is associated with a decline of about 10 log points relative to less adopted sectors, while the exposure coefficient is small and indistinguishable from zero once adoption is included. Adoption alone raises the within R^2 of the postings regression from 0.04 to 0.17; exposure alone leaves it at 0.04.

Third, the relationship survives a wide set of endogeneity and specification checks. The natural concern is that adopted sectors are the cognitive sectors that hired heavily in 2021 and early 2022 and corrected as monetary policy tightened, so that the decline after the release is reversion. Adoption is nearly uncorrelated with the size of the prerelease boom, at 0.20, and the boom does not predict it across sectors.

The adoption coefficient falls only from -0.108 to -0.075 as a battery of baseline characteristics interacted with the release is added. The estimate survives dropping each sector in turn, three placebo releases that reverse its sign, an anticipation window, randomization inference at $p = 0.002$, and independent geographic designs across 51 states and 489 metropolitan areas. Because a single sector, software development, dominates the level measure, I rely on randomization inference and on the rank of adoption.[2] Removing a linear

[2]To my understanding, there is/are no credible instrument is available: the natural task content candidates to instrument for adoption have first stage F statistics far below conventional thresholds and are collinear with exposure, so an instrumental variables strategy would trade one identification problem for a weak instrument problem.

extrapolation of the prerelease trend cuts the average post effect from about -0.07 to about -0.03, though it stays negative. The sign of the effect and the ranking of adoption over exposure hold throughout.[3]

The study's contribution is a measurement fact and a method for using behavioral data on labor market questions. Exposure and adoption diverge, and it is adoption that has tracked the labor market, so the exposure indices on which forecasting rests understate how concentrated the early footprint of generative AI has been. Workforce assessments built from exposure alone spread attention evenly across occupations that are exposed on paper, while the realized adjustment concentrates in a short list of occupations where the technology was cheap to deploy. Where the object of interest is realized impact rather than potential, exposure should be paired with adoption data, and the pairing can be repeated as adoption is measured for more systems and over a longer window.

The rest of the paper places itself in the literature (Section 2), lays out a task based framework (Section 3), describes the data and the divergence (Section 4) and the empirical strategy (Section 5), reports the main results (Section 6), the endogeneity and identification analysis (Section 7), robustness (Section 8), and the geographic designs and heterogeneity (Section 9), interprets the magnitudes (Section 10), and concludes (Section 11).

2 Related literature

The paper connects three strands of work and draws on a fourth for method. A first strand builds occupation level measures of exposure and applies them to anticipate labor market effects. Felten et al. (2021) link advances in AI capabilities to the abilities occupations require and construct the index used here; Felten et al. (2023) extend it to language models. Eloundou et al. (2024) score the exposure of tasks to large language models, and Webb (2020) measures exposure to earlier AI, robotics and software from patent text. These descend from

[3]The magnitude is sensitive to how the prior trend is treated and should be read as bounded between roughly three and ten log points per standard deviation.

an older literature scoring the automatability of occupations (Frey and Osborne, 2017). All of them measure potential from task content. Yet potential and realization separate empirically, and the labor market has moved with realization. The exposure measures differ in their inputs, from patent text (Webb, 2020) to human and model annotation of task suitability (Eloundou et al., 2024) to the mapping of abilities to AI progress (Felten et al., 2021), and they agree more on the ranking of occupations than on the intensity of exposure. What they share is that they are computed before adoption and cannot, by construction, distinguish a task a model could do from a task a model is used for.

A second strand models how task performing technologies affect labor demand. The "task framework" of Autor et al. (2003) and its formalization by Acemoglu and Restrepo (2018, 2019, 2022) distinguishes displacement, which lowers labor demand, from productivity and reinstatement, which raise it, and shows that the net effect depends on which tasks a technology touches.[4] I adopt its logic, that a technology's labor market effect works through the tasks it performs, and take it to measurement: labor demand should move with the tasks AI is used for rather than the tasks it could perform. The framework also implies that the sign of the effect is not fixed. A technology that performs tasks can lower labor demand by displacement or raise it through the productivity gains that expand output and the reinstatement of labor into new tasks (Acemoglu and Restrepo, 2019; Autor et al., 2024). The generative AI literature contains both signs, from the productivity gains measured in controlled settings to the demand declines measured on exposed platforms.

A third and fast growing strand studies generative AI directly. On the demand side, Acemoglu et al. (2022) and Bonfiglioli et al. (2025) relate exposure to vacancies and to local employment, and Hampole et al. (2025) study task exposure, productivity and employment jointly. On realized effects, Humlum and Vestergaard (2025) find small effects of chatbot access on earnings and hours in Danish data, while Hui et al. (2024) find that ChatGPT

[4]This framework has organized the study of prior automation: Acemoglu and Restrepo (2020) on robots, Autor et al. (2013); Autor and Dorn (2013) on the China shock and job polarization, Deming (2017) on the rising return to social skills, and Autor (2015); Autor et al. (2024) on the creation of new work.

reduced the number and earnings of freelance jobs in the most exposed occupations. On productivity, controlled studies find gains: Brynjolfsson et al. (2025) in customer support, Noy and Zhang (2023) in writing, Peng et al. (2023) for software developers. On adoption and firms, Bick et al. (2024) document rapid diffusion across workers, Chatterji et al. (2026a) track ChatGPT use across more than 1,500 organizations and find it concentrated among larger, research intensive firms and in writing, technical, and communication tasks, and Babina et al. (2024) link firm AI investment to growth and product innovation. Relative to this work, the paper measures realized adoption at the occupation level and shows that it, rather than exposure, tracks the movement in labor demand. It is closest in spirit to the work that measures realized adoption directly (Bick et al., 2024; Anthropic, 2025), but it differs in linking adoption to an outcome, labor demand, rather than describing adoption alone, and in confronting the problem that realized adoption is measured after the technology arrived.

The identification analysis uses recent econometrics of difference in differences and selection. I report an Oster (2019) selection on unobservables bound; a linear prior trend sensitivity analysis in the spirit of Rambachan and Roth (2023); randomization inference and a wild cluster perspective on inference with few clusters (Cameron et al., 2008); and I discuss why the estimator, which has a single common treatment date and continuous intensity, does not raise the negative weighting concerns of Goodman-Bacon (2021) and de Chaisemartin and D'Haultfœuille (2020). I also explain, in Section 5, why a shift share instrument in the spirit of Goldsmith-Pinkham et al. (2020) and Borusyak et al. (2022) is not available here. The design sits inside the recent reassessment of difference in differences estimators, which has shown that two way fixed effects can be biased when treatment timing varies across units and effects are heterogeneous (Goodman-Bacon, 2021; de Chaisemartin and D'Haultfœuille, 2020). Those concerns do not bind here, because treatment is a single common date interacted with a continuous intensity rather than a staggered onset, so the estimator is a simple interaction with weights. What binds is the credibility of parallel trends, which the sensitivity analysis of Rambachan and Roth (2023) in Section 7 takes up.

3 Conceptual framework

In this section, I present a simplified task based argument which yields the predictions tested. Let occupation o be a unit measure of tasks indexed by i, each producible either by labor at cost w_o or, once the technology exists, by the model at cost c_i and quality q_i. Task i is delegated to the model, $a_i = 1$, when the quality adjusted cost of the model falls below that of labor, net of an adjustment friction ϕ_i that captures regulation, liability, and the cost of reorganizing work; otherwise $a_i = 0$. Exposure of occupation o is the average capability over its tasks, $E_o = \int q_i \, di$, computed before deployment and independent of ϕ_i. Adoption is the measure of tasks actually delegated, $A_o = \int a_i \, di$; it depends on ϕ_i, is bounded above by exposure, and, because ϕ_i is large for most tasks outside a few types, is far more concentrated.[5]

3.1 Implications

First, exposure and adoption need not coincide. If adoption propensity were constant across tasks, the two would be proportional and carry the same information; but adoption propensity is skewed, high where deployment is cheap and complements exist, most obviously in software (Peng et al., 2023; Bick et al., 2024), and low where regulation, liability, or physical and interpersonal presence raise the cost of delegation. Because it is skewed, adoption is more concentrated than exposure and the two diverge in level even when they agree in rank, which Figure 1 shows.

Second, in the task framework the labor market effect of a technology operates through the tasks it performs, not the tasks it could perform (Acemoglu and Restrepo, 2018, 2022). Labor demand in occupation o is $L_o = (1 - A_o)\,\ell_o + \Delta_o$, where ℓ_o is labor per remaining task and Δ_o collects the productivity and reinstatement forces that raise demand when output expands or new tasks appear (Acemoglu and Restrepo, 2019; Autor et al., 2024). The change

[5]Tasks which are more directly replicable, eg. Software Development, Media Communications, IT Systems Solutions and Education Instruction, would be most susceptible to faster adoption.

in labor demand after the technology arrives, $dL_o = -\ell_o\, dA_o + d\Delta_o$, scales with adoption and not with exposure. If displacement dominates, $dL_o < 0$ where adoption is high; if the productivity and reinstatement forces dominate, $dL_o > 0$; either way the cross sectional variation loads on adoption. The postings regressions estimate this mapping.

Proposition 1. ***Adoption, not exposure, carries the first order effect****: Let labor demand in occupation o be $L_o = (1-A_o)\,\ell_o + \Delta_o$, where $A_o = \int a_i\, di$ is adoption, ℓ_o is labor per task, and Δ_o collects productivity and reinstatement forces, and let exposure be $E_o = \int q_i\, di$. Suppose exposure affects labor demand only through the tasks that are delegated, so that $\partial L_o/\partial E_o = 0$ holding A_o fixed. Then in the linear projection of the change in labor demand on adoption and exposure, $dL_o = \beta A_o + \gamma E_o + u_o$, the exposure coefficient is $\gamma = 0$ and the adoption coefficient is $\beta = -\ell_o + \partial\Delta_o/\partial A_o$, whose sign is the net of displacement and the offsetting forces.*

The proposition is a *horse race* shown in Section 6. Exposure should drop out once adoption is included, and it does. Its content is that delegable but undelegated tasks neither displace labor nor raise output, so exposure enters only as a noisy predictor of adoption and carries no independent information conditional on it.[6]

Corollary 1. ***Measurement error attenuates the adoption coefficient****: Suppose adoption is observed with classical error, $\widehat{A}_o = A_o + \nu_o$ with ν_o mean zero and independent of A_o and u_o. Then the probability limit of the estimated adoption coefficient is $\beta \cdot \mathrm{Var}(A)/[\mathrm{Var}(A) + \mathrm{Var}(\nu)]$, which is smaller in magnitude than β. The bias is toward zero, so the estimates understate the true association, and an instrument that predicted true adoption from a source independent of ν_o would correct it.*

The corollary is why the paper reads its estimates as conservative and why the absence of a usable instrument, discussed in Section 5, is a limitation on sharpening the magnitude rather than a threat to the sign.[7] The one kind of measurement error that would bias the

[6] The sign of β is deliberately left open, because the model admits both displacement and the productivity and reinstatement forces that work the other way; the regressions estimate it.

[7] Measurement error of the classical kind cannot manufacture the association; it can only shrink it.

estimate away from zero, an $\widehat{A}_o$ that systematically overstated adoption in the occupations that were contracting for other reasons, is not classical, and the reverse causality test in Section 7 is aimed at it directly.

Two comparative statics organize the empirical work. First, holding exposure fixed, an increase in adoption, which raises the measure of delegated tasks A_o, lowers labor demand through the displacement term and is offset only to the extent that the productivity and reinstatement forces are large; the net sign is what the regressions estimate. Second, holding adoption fixed, an increase in exposure changes nothing, because tasks that are delegable but not delegated neither displace labor nor raise output. The horse race in Section 6 tests the prediction that exposure has no effect conditional on adoption.[8]

The framework also clarifies what the design can and cannot recover. Adoption propensity is a structural feature of tasks, but it is realized over time and can respond to the same conditions that move labor demand, which is the endogeneity concern of Section 7. And because the model's output is tradable across places and firms, a fall in postings in adopted occupations is a statement about the reallocation of demand across occupations, not about its aggregate level, which the design absorbs. Underscoring the general equilibrium adjustment, the first order effect on an occupation's labor demand runs through the tasks it delegates, which is adoption, not through the tasks it could delegate, which is exposure. The framework is partial in a second sense: it treats an occupation's tasks as fixed, but the boundary of an occupation is itself endogenous over a longer horizon, as new tasks appear and old ones are automated (Autor et al., 2024; Acemoglu and Restrepo, 2019). Over the short window studied here that boundary is close to fixed, which makes a task based reading of a three year change reasonable; over a decade it would not be.

[8]Classical measurement error attenuates the adoption coefficient toward zero, so the estimates are conservative; an instrument that predicted true adoption from a source independent of the measurement error would correct the attenuation, which is one reason the absence of a usable instrument is a limitation rather than a fatal flaw. Measurement error that instead overrepresented the occupations that were contracting would bias the estimate away from zero, and the reverse causality test in Section 7 speaks to it.

4 Data, Descriptives and Divergence

The outcome is the Indeed Hiring Lab index of job postings, reported by occupational sector as a percentage of the sector's level on 1 February 2020 and smoothed as a seven day trailing average (Indeed Hiring Lab, 2026).[9] I use all 47 sectors, aggregate to weeks, and take logs. The main sample runs from January 2021 through December 2025, and due to the pandemic, the full sample from February 2020 is a robustness check. The choice to begin in January 2021 trades power for cleanliness: the pandemic collapse and rebound of 2020 moved sectors by amounts unrelated to AI and differing sharply between in person and remote work, so including 2020 introduces differential swings that have nothing to do with the release, while beginning in 2021 leaves nearly two years of prior period to estimate the trend and place the placebo tests. The full sample gives similar estimates, so the choice is not doing the work. The weekly panel has 12,267 sector weeks and is balanced. Table A1 reports summary statistics. The index measures the flow of new and active vacancies, not employment, and it is well suited to dating a change because it responds within weeks rather than the quarters that employment and wages take to move. Its sectors are Indeed's occupational groupings, which map onto the occupation system and span the economy. The index is expressed relative to February 2020, so level differences across sectors are absorbed by the sector fixed effects and identification comes from the differential movement of sectors after the release.[10]

Exposure is the AI Occupational Exposure index of Felten et al. (2021), provided at the level of six digit occupation codes, which maps AI capabilities to the abilities occupations require. It captures exposure to artificial intelligence broadly, the conservative choice for a horse race against realized use. I aggregate it to Indeed sectors through the crosswalk in Appendix Table A4 and standardize. As the established occupation level measure in this literature, it is the natural benchmark it against the newer adoption measure.

[9] Indeed is the largest job site in the United States, and its index is a widely used high frequency indicator of labor demand.

[10] The data is seasonally patterned but not seasonally adjusted at the sector level, and the week fixed effects, which absorb any movement common to all sectors, remove seasonality.

Adoption is built from the Anthropic Economic Index (Appel et al., 2025). Anthropic classifies conversations against O*NET tasks and reports the share of usage each task accounts for; the shares concentrate on about 3,500 tasks, and the remaining tasks carry essentially no usage. I map tasks to occupations through the O*NET task to occupation link and define an occupation's adoption as the mean usage share across its tasks, assigning zero to tasks that do not appear. The most adopted occupations are computer programmers, software developers, technical writers, statistical assistants and writers; the least adopted are manual occupations. I aggregate to sectors and standardize. The measure reflects usage of one model family and is observed over 2024 and 2025; I treat it as a structural task characteristic, the same stance the exposure literature takes toward its own measures, and defend that reading in Section 7.

The index is constructed by classifying a large and continuously collected sample of conversations against the O*NET taxonomy with a model based classifier, and by reporting, at the task level, the share of usage and a split between automative and augmentative interactions. I use the task level usage shares, which are the component least sensitive to the classification of interaction type, and aggregate them to occupations. To the extent that the pattern of usage by task is similar across frontier systems, which the underlying reports argue by comparing across models and over time, the measure proxies for realized adoption more broadly, and I treat this as an assumption rather than a fact. The reliability of the measure is examined throughout: its ordering is benchmarked against exposure, its magnitude is checked against alternative aggregations of usage in Section 8, and its concentration and the inference it will support are the subject of Section 7.

For the endogeneity analysis I assemble a few predetermined occupation characteristics: teleworkability (Dingel and Neiman, 2020); median wage and entry education from the Occupational Outlook Handbook; and language and physical task shares computed by classifying the text of each O*NET task as language oriented or physical. These characteristics are aggregated to sectors and standardized, and Appendix Table A2 documents source.

The mapping from the 47 sectors to occupation codes is the main measurement judgment in the paper. Each sector is assigned the occupation code prefixes that best represent it, and a sector's measure is the mean of the occupation measure over the matched codes; the full mapping is in Appendix Table A4. The mapping is coarse, because 47 sectors cover several hundred detailed occupations, but it does not drive the results: weighting the occupation measures by employment rather than taking simple means leaves the estimates close, and the sector ranking by adoption matches independent accounts of where generative AI is used. Measurement error in the mapping, like measurement error in adoption itself, attenuates the estimates toward zero.

Figure 1 plots exposure against adoption across sectors as percentile ranks. They are positively related, with a rank correlation of 0.76, so exposure orders occupations reasonably well. In levels the correlation is 0.25. Table 1, Panel A, quantifies the reason: the Gini coefficient of adoption across sectors is 0.76 against 0.25 for exposure, and the five most adopted sectors hold 68 percent of usage against 16 percent for the most exposed. Figure 2 shows the two Lorenz curves. The concentration is easy to read off the sector ranking in Appendix Figure A7. Software development is the most adopted sector by a wide margin, followed by information technology, data and analytics, media and communications, and education; at the other end, construction, cleaning, loading and stocking, driving, and food preparation register almost no usage. Exposure orders the same sectors similarly but compresses the distances: accounting, finance, and marketing top the exposure ranking and manual sectors sit at the bottom, but no sector is the runaway leader that software development is for adoption. The divergence is therefore not disagreement about which occupations are cognitive; it is disagreement about how much of the cognitive category has actually taken up the technology. Panel B of Table 1 regresses adoption on the prerelease boom and other cycle proxies across sectors and finds none of them predictive, which anticipates the identification analysis: the concentration of adoption is a structural feature of task content, not a shadow of the hiring cycle.

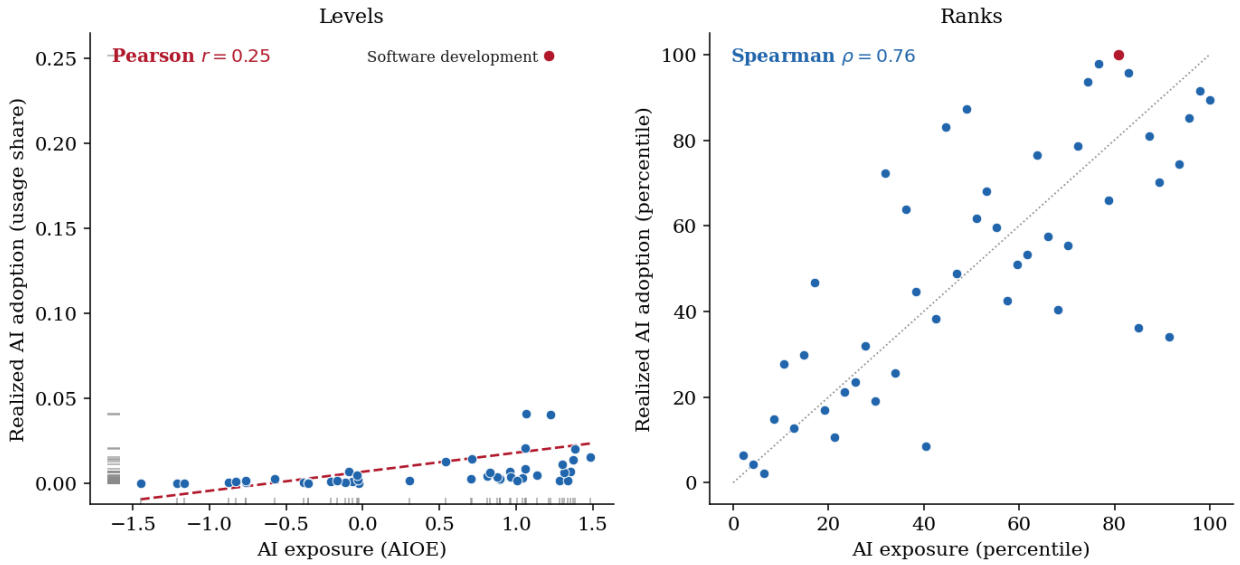


Figure 1: Exposure and adoption across 47 occupational sectors, in levels (left) and in ranks (right). In levels, realized adoption is concentrated in a few sectors, led by software development (red), while most sectors register almost no usage, so the Pearson correlation is only 0.25 and the fitted line is pulled toward the outliers. In ranks, the same sectors line up near the diagonal and the Spearman correlation is 0.76. The two measures agree on which sectors are cognitive and disagree on how concentrated adoption is: the rank panel shows the agreement, the level panel shows the concentration.

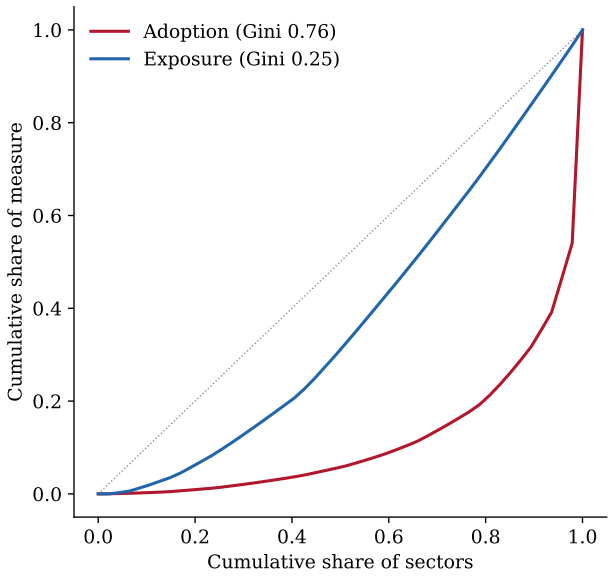


Figure 2: Lorenz curves of realized adoption and exposure across the 47 occupational sectors, ordering sectors from least to most intensive along the horizontal axis and plotting the cumulative share of the total on the vertical axis. The further a curve falls below the diagonal, the more concentrated the measure. Adoption falls far below it, with a Gini coefficient of 0.76, while exposure stays close to it, at 0.25: a few sectors hold most realized usage, and exposure is spread almost evenly.

Table 1: Adoption is concentrated and is not explained by the hiring cycle

A. Concentration across sectors		
	Gini	Top-5 share
Adoption (AEI usage)	0.76	0.68
Exposure (AIOE, shifted positive)	0.25	0.16
B. Predicting adoption from baseline task content		
	Coef.	p
Constant	0.000	1.000
Pre-boom	0.193	0.221
Wage	-0.074	0.762
Education	-0.101	0.717
Teleworkable	0.363	0.053
Physical-task share	-0.002	0.993
R^2	0.12	

Notes. Panel A reports, across the 47 sectors, the Gini coefficient and the combined share held by the five highest sectors for each measure; because the Gini requires non-negative values, exposure is shifted by its minimum before it is computed. Adoption is far more concentrated than exposure. Panel B is a single cross sectional ordinary least squares regression of standardized adoption on standardized predetermined characteristics across the 47 sectors: the prerelease boom (each sectors mean postings while ChatGPT was unavailable), median wage, entry education, teleworkability, and the physical task share, with a constant. Coefficients are in standard deviation units and p values are two sided. The regression explains 12 percent of the cross sectional variation in adoption and the prerelease boom does not significantly predict it, so adoption is not a relabeling of which sectors were booming.

5 Empirical strategy

ChatGPT was released on 30 November 2022 and reached an estimated one hundred million users within two months, among the fastest diffusions of a consumer technology on record (Bick et al., 2024). Its arrival did not occur in a calm labor market. Vacancies surged through 2021 and early 2022 as the economy reopened, with the strongest gains in cognitive and remote capable occupations, and then cooled through 2022 and 2023 as the Federal Reserve raised interest rates and the technology sector announced large layoffs. A design that dates an effect to the release must contend with this cycle, which moved the same cognitive occupations that adopt AI most.

Let y_{st} be log postings in sector s and week t. The main specification is

$$y_{st} = \beta \left(A_s \times \text{Post}_t\right) + \gamma \left(E_s \times \text{Post}_t\right) + X'_{st}\theta + \alpha_s + \delta_t + \varepsilon_{st}, \tag{1}$$

where A_s is adoption, E_s is exposure, Post_t is one from December 2022, α_s and δ_t are sector

and week fixed effects, and X_{st} collects baseline characteristics interacted with Post_t. Week effects absorb every shock common to all sectors, including the aggregate hiring cycle and monetary policy. Both measures are standardized, so β and γ are comparable, and the horse race asks which carries independent content. Standard errors are clustered by sector; because there are 47 clusters, I report randomization inference and note that heteroskedasticity robust errors.

To trace dynamics and test the identifying assumption I estimate

$$y_{st} = \sum_{k \neq -1} \theta_k \left(A_s \times \mathbf{1}[\text{event month} = k] \right) + \alpha_s + \delta_t + \varepsilon_{st}, \tag{2}$$

with the month before the release omitted, and the analogous specification for exposure. The coefficients for $k < 0$ test whether adopted sectors were already on a different trajectory.

The treatment is continuous rather than binary, so β is the effect of a one standard deviation increase in intensity and uses the full cross sectional variation. Identification comes from a single common date rather than staggered timing, so the estimator is an interaction of intensity with a post indicator and does not raise the negative weighting concerns that arise when treatment turns on at different times for different units (Goodman-Bacon, 2021; de Chaisemartin and D'Haultfœuille, 2020). The continuous treatment means the estimate is a variance weighted average of the effect of intensity across sectors, so the heavily adopted sectors receive more weight. This suits the question, since the effect is concentrated where adoption is intense, and it is why the leave one out analysis, which confirms that no single heavily adopted sector drives the result, matters.

The comparison of the two coefficients is identified under a weaker condition than either coefficient alone. The horse race asks whether adoption carries information about the change in labor demand that exposure does not, and the week effects hold fixed every shock common to sectors, so β and γ are identified from the differential movement of sectors that differ in adoption and exposure.

Point estimates from the results show conditional comparison and a bounded causal reading, not point identified causal magnitudes; the level of the decline is left to the sensitivity analysis and is not assigned to AI. Adoption enters as a task characteristic whose ordering is predetermined, fixed by task content and 0.76 rank correlated with an exposure index built before the release, while its intensity is measured after the fact and, being measured with error, attenuates β toward zero by Corollary 1.

Instrumenting adoption would sharpen the magnitude, but the task content measures that could serve, such as the language share of an occupation's tasks, are collinear with exposure and weak, with first stage F statistics below three (Appendix Table A11, Panel A), and a shift share instrument in the spirit of Goldsmith-Pinkham et al. (2020) and Borusyak et al. (2022) would require local occupation employment shares that this outcome does not provide. The identification therefore rests on the selection on observables bounds and the design based checks of Section 7, which are what discipline the causal reading of the comparison.

5.1 Threats to identification

The first is that adoption is a stand in for the cognitive hiring cycle; the week fixed effects remove the common component of that cycle, and the controls, the Oster bound, and the reverse causality test in Section 7 address the part specific to adopted sectors. The second is reverse causality, that contracting sectors adopted AI faster; the near zero correlation of adoption with the prior boom and the failure of the cycle to predict adoption speak against it. The third is that adoption is measured after the release; its ordering is predetermined by task content, and classical measurement error in its intensity biases the estimate toward zero. None of these is removed entirely, which is why the paper reads adoption as a predetermined characteristic, reports the sensitivity of the magnitude, and does not claim a clean experiment.

6 Main results

Table 2 reports the difference in differences estimates. Column 1 includes only exposure, at -0.053 and marginally significant. Column 2 includes only adoption, at -0.108 and precise, and it raises the within R^2 from 0.04 to 0.17. Column 3 puts them together: adoption is -0.102 and exposure falls to -0.027 and loses significance. Columns 4 through 6 add the prerelease boom, teleworkability, and trends specific to each sector, each interacted with the release; adoption moves from -0.094 to -0.073 and stays significant, while exposure never becomes independently negative. That exposure is marginally significant on its own in column 1 but insignificant once adoption is included in column 3 is what the framework predicts.

Exposure and adoption share the cognitive dimension, so exposure alone picks up some of the adoption signal; conditioning on adoption removes it, and what remains of exposure is close to zero. Exposure matters for the labor market only through the adoption it predicts, and it predicts adoption imperfectly enough that the two separate. Randomization inference places the adoption coefficient in the extreme tail of its permutation distribution, at $p = 0.002$, against $p = 0.44$ for exposure.[11] A one standard deviation gap in adoption corresponds to a relative decline of about 10 log points in the raw specification, which the endogeneity and prior trend analysis below places between roughly three and ten log points; at either end of that range the ranking is the same, adoption carries the labor market signal and exposure does not.

Figure 3 plots the event study for adoption and for exposure. The pattern is an inverted V. Postings in adopted sectors rose relative to others through 2021 and 2022, peaked in the middle of 2022, and fell after the release to about 9 log points below the baseline within a year. Exposure traces the same shape with a smaller amplitude, and the decline is larger and more persistent for adoption. The rise before the release is the cognitive hiring boom;

[11] The gap in fit is as large as the gap in the point estimates: adoption alone accounts for a within R^2 of 0.17, four times what exposure alone accounts for, and in the horse race adoption carries a t statistic above five against one below one for exposure.

Table 2: Realized use, not exposure, tracks the post release adjustment in job postings

	(1)	(2)	(3)	(4)	(5)	(6)
Actual use × Post		-0.108***	-0.102***	-0.094***	-0.087***	-0.073***
		(0.020)	(0.018)	(0.020)	(0.019)	(0.013)
Exposure × Post	-0.053*		-0.027	-0.029	0.040	-0.016
	(0.031)		(0.029)	(0.028)	(0.044)	(0.036)
Boom × Post				-0.037	-0.034	
				(0.024)	(0.022)	
Teleworkable × Post					-0.089*	
					(0.046)	
Sector, week fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Sector specific trends	No	No	No	No	No	Yes
Observations	12,267	12,267	12,267	12,267	12,267	12,267
Within R^2	0.040	0.170	0.179	0.199	0.240	0.718

Notes. Difference in differences estimates on the weekly panel of 47 occupational sectors from January 2021 through December 2025 (12,267 sector weeks). The dependent variable is the log Indeed Hiring Lab job postings index. Actual use (adoption) is the Anthropic Economic Index mean task usage share, exposure is the Felten, Raj and Seamans AI Occupational Exposure index, and the prior boom is each sectors mean postings while ChatGPT was unavailable; all three are standardized to mean zero and unit standard deviation across sectors and interacted with Post, an indicator equal to one from December 2022, and teleworkability follows Dingel and Neiman. Every column includes sector and week fixed effects, so identification comes from how sectors that differ in the measures moved relative to one another after the release and the week effects absorb any shock common to all sectors, including the aggregate hiring cycle; column 6 adds a linear time trend specific to each sector. Each coefficient is the effect of a one standard deviation increase in the measure on log postings after the release; standard errors clustered by sector are in parentheses and the within R^2 is the fit after the fixed effects are partialled out. Randomization inference over the 47 sectors gives $p = 0.002$ for adoption and $p = 0.44$ for exposure. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

whether the reversal is more than that boom unwinding is taken up in Section 7.[12] Appendix Figure A6 shows the same reversal in the raw postings index by tercile of adoption, without regression adjustment. Through 2021 and 2022 the most adopted third of sectors posted more vacancies, relative to February 2020, than the least adopted third; the two groups crossed in the months around the release; and by 2024 the most adopted third had fallen below the least adopted third and stayed there. The regression puts a number and a standard error on this crossing and separates it from exposure, but the crossing itself is visible in the unadjusted index.

7 Endogeneity and identification

The central concern is that adopted sectors are the cognitive sectors that boomed and corrected, so that the postrelease decline is reversion rather than a response to AI. This section

[12]Randomization inference across the 47 sectors gives $p = 0.002$ for adoption and $p = 0.44$ for exposure; Appendix Figure A5 shows the permutation distribution.

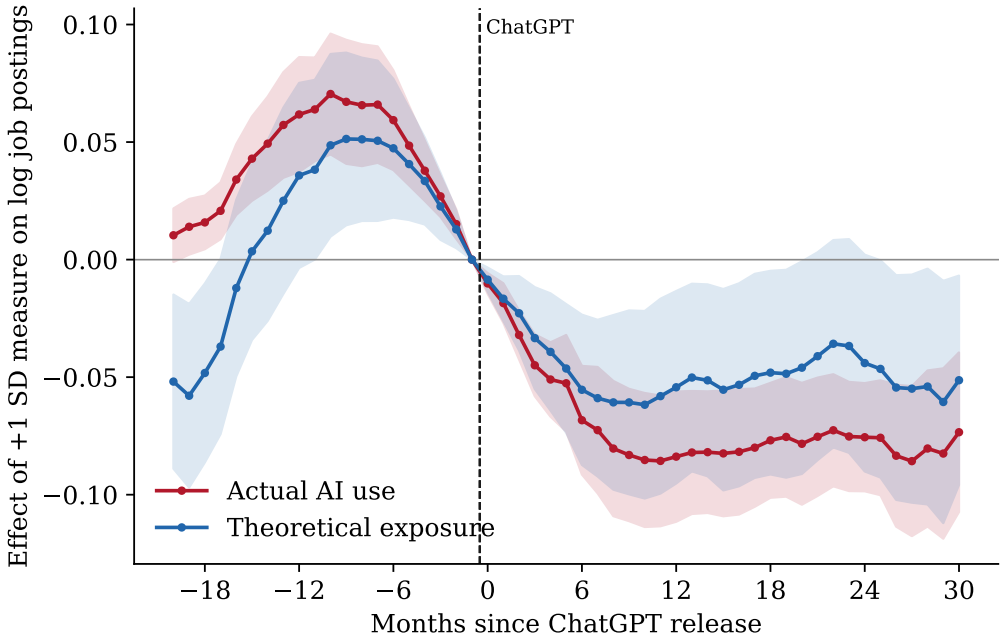


Figure 3: Event study coefficients for adoption (red) and exposure (blue), with 95 percent confidence bands, months relative to the release. The reversal is larger and more persistent for adoption.

confronts that concern with a rich set of controls and an Oster bound, tests for reverse causality, and reports design based checks.

Table 3 adds baseline characteristics to equation (1) one block at a time: exposure; wage and education; teleworkability and the physical and language task shares; the prerelease boom; and trends specific to each sector. The adoption coefficient falls only from −0.108 to −0.075 across this battery and stays significant at the one percent level throughout, and Figure 4 shows the coefficient path. The controls differ in how much they move it. Adding exposure barely changes it, since exposure carries little independent information, and adding wage and education changes it little. The controls that matter are teleworkability and the prior boom, each absorbing a few log points, consistent with the adopted sectors being cognitive and remote capable and having boomed; but even together these leave three quarters of the coefficient intact, and the trends specific to each sector, which absorb any smooth difference in trajectory, leave two thirds. The coefficient is thus not fragile to the controls one would worry about most.

The within R^2 rises from 0.17 with adoption alone to 0.26 with the full battery short of sector trends, and to 0.72 once trends specific to each sector are included, so the controls do explain a substantial share of the variation in postings; the adoption coefficient nonetheless holds. Because the coefficient moves little as controls that explain a good deal of the outcome are added, selection on unobservables is unlikely to explain it away.

The Oster (2019) statistic makes this precise: at the conventional bound $R_{max} = 1.3\tilde{R}$, the degree of selection on unobservables relative to observables that would set the coefficient to zero is $\delta = 4.1$, and the bias adjusted coefficient is still -0.064; even at $R_{max} = 2\tilde{R}$, $\delta = 1.2$ and the bias adjusted coefficient stays negative. Only at the extreme $R_{max} = 1$, in which unobservables explain all residual variation, does the bound become uninformative (Appendix Table A11, Panel B). The bound should be read for what it is: a statement about selection on unobservables under the assumption that it is proportional to selection on observables, calibrated to a maximum R^2. It is not a test, and it does not rule out a confounder orthogonal to the observed battery yet correlated with adoption and with the postings decline. What it does is discipline the intuition that an omitted variable could explain the result: for that, the omitted variable would have to do more work than exposure, wage, education, teleworkability, the task shares, and the prior boom combined.

If contracting sectors adopted AI faster, adoption would be an outcome of the cycle. Panel B of Table 1 regresses standardized adoption on the prerelease boom, wage, education, teleworkability, and the physical task share across sectors. The regression explains only 12 percent of the variation in adoption, and the prerelease boom is not a significant predictor. Therefore, adoption is not a relabeling of which sectors were contracting. The timing of the two measures sharpens the point. Adoption is measured in 2024 and 2025, after the postings decline was underway, so a mechanical worry is that the decline caused the adoption rather than the reverse. The occupations that adopt AI most are those whose tasks are most amenable to a language model, a property of task content fixed long before the release, and the adoption ranking is 0.76 rank correlated with an exposure index built from that content;

Table 3: Endogeneity: adoption survives a rich set of controls, and an Oster selection bound

	(1)	(2)	(3)	(4)	(5)	(6)
Adoption × Post	-0.108***	-0.102***	-0.098***	-0.094***	-0.085***	-0.075***
	(0.020)	(0.018)	(0.018)	(0.018)	(0.021)	(0.013)
Exposure × Post		Yes	Yes	Yes	Yes	Yes
Wage, education × Post			Yes	Yes	Yes	Yes
Telework, task shares × Post				Yes	Yes	Yes
Pre-boom × Post					Yes	Yes
Sector specific trends						Yes
Within R^2	0.170	0.179	0.187	0.235	0.256	0.722
Oster (2019) selection bound, columns (1) vs (5):						
δ at $R_{max} = 1.3\tilde{R}$	4.06					
δ at $R_{max} = 2\tilde{R}$	1.22					
Bias-adjusted β (δ=1, R_{max}=1.3$\tilde{R}$)	-0.064					

Notes. Difference in differences estimates on the weekly panel of 47 sectors, January 2021 through December 2025 (12,267 sector weeks), with sector and week fixed effects and standard errors clustered by sector in parentheses. The dependent variable is the log Indeed job postings index and Post equals one from December 2022. Column 1 includes only adoption interacted with Post; each later column adds the labeled block, also interacted with Post: exposure in (2), median wage and entry education in (3), teleworkability and the language and physical task shares in (4), the prerelease boom in (5), and a linear time trend specific to each sector in (6); Yes marks an included block. The adoption coefficient is the effect of a one standard deviation increase in adoption and falls only from -0.108 to -0.075 across the battery. The lower panel reports the Oster (2019) selection on unobservables bound comparing the uncontrolled column (1) with the fully controlled column (5): δ is the strength of selection on unobservables, relative to selection on the included controls, that would be needed to drive the adoption coefficient to zero, evaluated at a maximum R^2 equal to 1.3 and to 2 times the controlled within R^2; a value above one means unobservables would have to matter more than the full observed battery. The bias adjusted β sets δ to one at the 1.3 bound. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

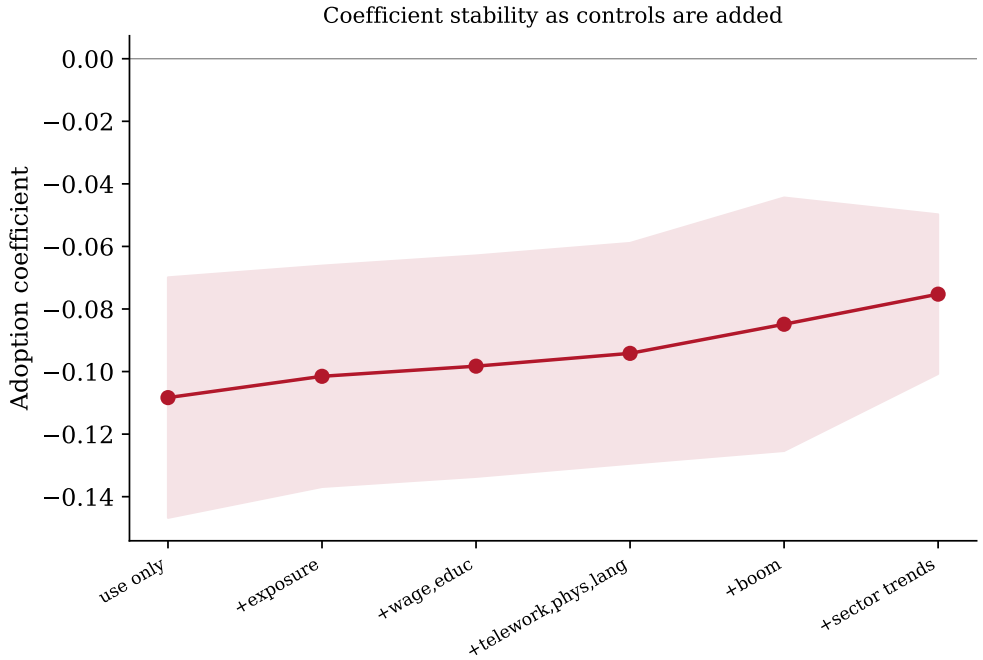


Figure 4: Coefficient stability. The adoption coefficient and its 95 percent confidence band as baseline characteristics are added to the horse race, corresponding to the columns of Table 3.

and within the period, the cross section of adoption is not predicted by the cross section of the prior contraction. Adoption is better read as revealed task amenability than as a response to the cycle.

Table 4 gathers design based checks. Three placebo releases placed in the prerelease period, in June and November 2021 and March 2022, all yield positive adoption coefficients, the opposite of the true effect, so the negative estimate is specific to the actual timing. The placebos are informative precisely because the prior period is not quiet. In a placebo release the adopted sectors are on their rising boom trajectory, so a mechanical difference in differences that mistook trend for treatment would return a positive coefficient, and it does. That the true effect reverses this sign separates the postrelease decline from the boom continuing: the same sectors that were pulling ahead before the release fell behind after it. Dropping the months immediately around the release, from one month before to two months after, to allow for anticipation, leaves the coefficient at -0.107.

Table 4: Identification checks

Check	Adoption estimate	p / note
A. Placebo releases (pre-period only)		
Fake break 2021-06-01	+0.028	0.000
Fake break 2021-11-01	+0.026	0.000
Fake break 2022-03-01	+0.016	0.001
B. Anticipation		
Drop months -1 to $+2$	-0.107	0.000
C. Randomization inference (499 draws)		
Adoption	-0.102	$p = 0.002$
Exposure	-0.027	$p = 0.444$
D. Pre-trend sensitivity (linear detrend)		
Average post effect, raw	-0.071	
Average post effect, detrended	-0.031	

Notes. Design based identification checks for the adoption coefficient, each estimated on the weekly panel of 47 sectors with sector and week fixed effects and standard errors clustered by sector. Panel A re-estimates the horse race with a fake release date placed in the prerelease period; all three give positive coefficients, the opposite of the true negative effect, so the estimate is specific to the actual timing. Panel B drops the four months around the release, from one month before to two months after, to allow for anticipation. Panel C reports randomization inference, which permutes the exposure and adoption values jointly across the 47 sectors and re-estimates the horse race for each of 499 draws; the p value is the share of draws whose absolute coefficient is at least as large as the observed one, and it is 0.002 for adoption against 0.44 for exposure. Panel D fits a line to the adoption event study coefficients over the year before the release, extrapolates it into the post period, and subtracts it; the average post effect over the two years after the release falls from -0.071 to -0.031 but stays negative.

The concentration that gives the paper its name has a consequence for inference and

disciplines which test to trust. Because adoption is so concentrated, the standardized level measure is dominated by a single sector: software development sits about six and a half standard deviations above the mean, and it alone accounts for 92 percent of the cross sectional variance of standardized adoption.[13] This is exactly the configuration in which the wild cluster bootstrap is known to behave poorly: with one high leverage cluster it under rejects severely, returning a p value of 0.64 for adoption in levels even though the clustered t statistic exceeds five and randomization inference gives 0.002 (MacKinnon and Webb, 2018). The disagreement is not evidence that the effect is fragile; it is the signature of a treatment whose variation is concentrated in one unit, for which the bootstrap's asymptotics fail and randomization inference, which does not appeal to the number of clusters and permutes the actual treatment values across sectors, is the appropriate benchmark.

A potential way to see whether the result is more than the software development sector is to remove that sector's leverage. The rank of adoption does this: replacing the level with its rank spreads leverage across the sectors, raising the effective number of clusters from 1.2 to 26 of 47, and under the rank measure every inference method agrees. The rank coefficient is -0.156, its clustered p value is 0.003, its wild cluster bootstrap p value is 0.007, and its randomization inference p value is 0.005; it survives dropping software development, at -0.134 and significant at the one percent level. Appendix Table A12 shows these diagnostics. The finding is therefore not an artifact of a single high leverage sector: it holds across the ordering of all 47 sectors when the outlier's leverage is removed, and the level and rank measures agree on the sign, the ranking of adoption over exposure, and significance under the test that suits each. What the concentration does mean is that the precision of the level estimate rests on the most adopted sectors: the labor market moved where adoption concentrated, and adoption concentrated in few places.

The event study shows a prerelease trend, so I test how much of the post effect survives its removal. Fitting a line to the adoption event study coefficients over the year before

[13] In terms of cluster robust inference, the effective number of clusters carrying treatment variation, the inverse Herfindahl index of the sectors' leverage, is about 1.2 out of 47 for the level measure.

the release and extrapolating it into the post period, the average adoption effect over the two years after the release falls from −0.071 raw to −0.031 once the extrapolated prior trend is subtracted (Table 4, Panel D, and Figure 5). A linear continuation of the prior trend accounts for roughly half of the raw post effect. It does not change the sign, or the ranking of adoption over exposure, and the horse race, in which both measures carry the same prior trend, is more robust than the level of either coefficient. This linear extrapolation is a transparent version of the sensitivity analysis of Rambachan and Roth (2023); a fuller treatment would trace the identified set across a range of allowed trend violations, and the direction is already clear from the linear case.

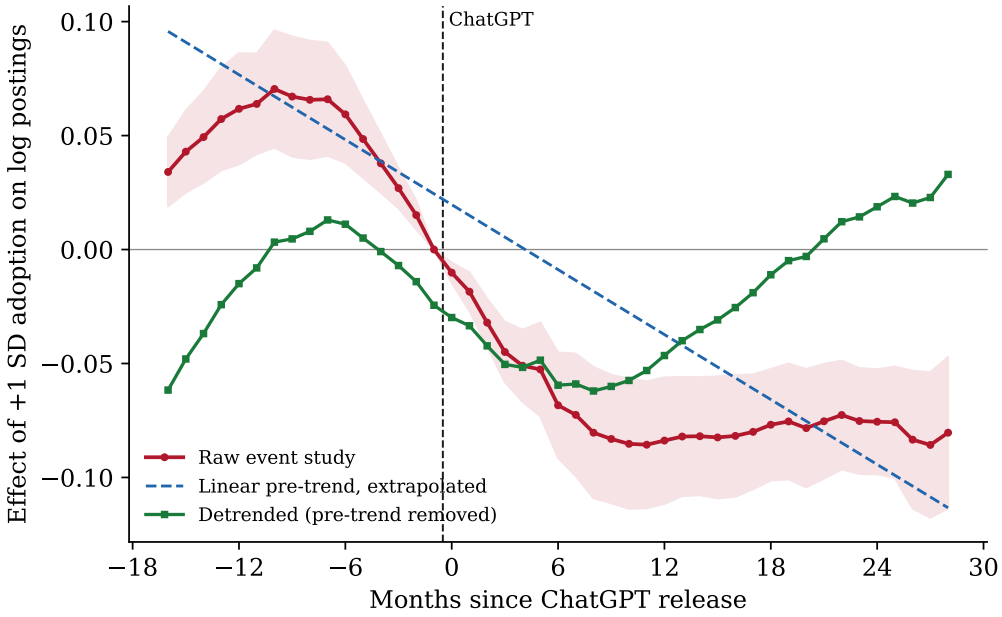


Figure 5: Prior trend sensitivity. The raw adoption event study (red), a linear extrapolation of the prerelease trend (blue dashed), and the detrended series (green). Removing the extrapolated prior trend roughly halves the post effect but leaves it negative.

The result is not the product of any single sector. Dropping each sector in turn and estimating again the horse race, the adoption coefficient ranges from −0.096 to −0.424, with a mean of −0.108, and is significant at the five percent level in all 47 regressions (Table 5, Panel A, and Figure 6). Dropping software development, the largest adopter, moves the coefficient away from zero rather than toward it, because removing the outlier shrinks the

dispersion of the standardized measure.

Taking the checks together, the sign of the effect and the ranking of adoption over exposure are established about firmly: they survive a rich control battery, a selection bound in which unobservables would need to dominate observables, the removal of any single sector, an anticipation window, three placebo dates, and randomization inference. What is not established is the precise magnitude, which the prior trend analysis shows depends on how much of the prior trend one extrapolates, and the causal interpretation of the level, which the coincidence with the technology cycle leaves open. The paper claims the former, and acknowledges the limitations with regards to the latter.

Table 5: Leave-one-out stability and inference

A. Leave-one-out (drop each sector, 47 regressions)	
Minimum estimate	-0.424
Maximum estimate	-0.096
Mean estimate	-0.108
Significant at 5%	47 of 47
B. Inference: standard error on adoption	
cluster sector	0.0180
heteroskedastic	0.0025
Randomization inference p	0.002

Notes. Panel A drops each of the 47 sectors in turn and re-estimates the baseline horse race, reporting the minimum, maximum and mean adoption coefficient over the 47 regressions and the number significant at the 5 percent level; dropping software development, the largest adopter, moves the coefficient away from zero, to -0.424, because it shrinks the dispersion of the standardized measure. Panel B reports, for the adoption coefficient in the baseline horse race, the standard error clustered by sector, the heteroskedasticity robust standard error that treats the 12,267 sector weeks as independent and is not the right benchmark because postings are serially correlated within a sector, and the randomization inference p value.

8 Robustness

Table 6 shows robustness to samples, outcomes, and measures. The estimate is similar in the full sample that includes 2020, with the new postings index as the outcome, at monthly rather than weekly frequency, in a tighter window around the release, and when a second, independently constructed exposure measure is added; adoption survives against each exposure measure while exposure does not (Appendix Table A7). Two results in this table qualify the finding. First, dropping the entire complex of software, information technology

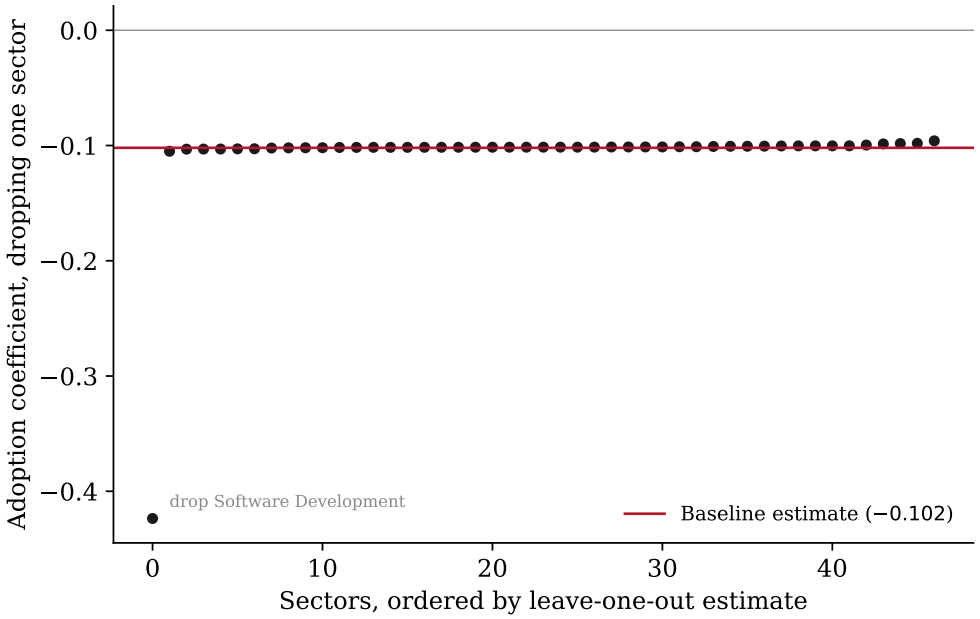


Figure 6: Leave one out. The adoption coefficient from the baseline horse race dropping each of the 47 sectors in turn, against the full sample estimate.

and data occupations leaves a coefficient of −0.33, so the result is not confined to the technology sector; but combining a rank transformation of adoption with that exclusion is the one specification in which the effect weakens to marginal significance, and I report it as the boundary of the finding.

The technology sectors are both the most adopted and the most affected by the 2022 and 2023 correction, so it is natural to be skeptical and to attribute the whole result to them. However, the coefficient not only survives but grows when software development alone is removed, and remains significant when the whole technology and data complex is removed. This shows that the association is present across the cognitive economy and not confined to the sector where the confound is sharpest. Second, among the alternative aggregations of adoption in Panel B, the mean usage share used throughout and the rank of adoption are both significant, at −0.102 and −0.156, but the sum of task usage, which weights occupations by their number of tasks as well as their intensity, is not, at −0.028. The mean and rank are the natural intensity measures and both hold; the sum conflates intensity with occupation breadth. The mean is the right measure for the question, because it captures the intensity

with which an occupation's work is done with AI rather than the number of tasks it happens to contain; it is what the framework calls adoption, the share of an occupation's work that is delegated, and what should move labor demand per worker. That the rank of adoption also delivers a significant and larger estimate shows that the result depends on the ordering of sectors by intensity, not on the cardinal values of the usage shares. Appendix Tables A6 and A7 give the full alternatives. The estimate is stable across frequency and sample. At the monthly frequency, which averages out weekly noise, the adoption coefficient is essentially unchanged; in the full sample that begins in February 2020, which reintroduces the pandemic swing, it is -0.096; and with the new postings index, which counts only recently created vacancies, it is -0.085. The point estimate sits in a narrow band across specifications, and the exposure coefficient is small and insignificant in all of them.

Table 6: Robustness to samples, outcomes, and measures

Specification	Adoption × Post	Exposure × Post	Obs.
A. Samples, outcomes, exposure measures			
Baseline (weekly 2021+)	-0.102***	-0.027	12,267
Full sample (2020+)	-0.096***	-0.013	14,570
Drop Software Development	-0.424***	0.010	12,006
Drop all tech/IT/data	-0.328**	0.012	11,223
Sector-specific trends	-0.082***	-0.049**	12,267
New postings outcome	-0.085***	-0.002	12,267
Monthly frequency	-0.101***	-0.026	2,820
Window 2022m1-2024m6	-0.098***	-0.057**	6,110
Add second exposure measure	-0.192	0.093**	9,918
B. Alternative adoption definitions			
Adoption: mean share (baseline)	-0.102***		
Adoption: sum of task usage	-0.028		
Adoption: rank	-0.156***		

Notes. Each row of Panel A is a separate difference in differences horse race that includes adoption and exposure interacted with Post, sector and week fixed effects, and standard errors clustered by sector; only the two interaction coefficients and the number of observations are shown. The rows vary the sample (the full sample adds 2020; the window row keeps January 2022 through June 2024), the outcome (the new postings index counts only recently created vacancies), the frequency (monthly rather than weekly), the sectors excluded (software development alone, or the entire technology, information technology and data complex), the trend specification (a sector specific linear trend), and the exposure control (a second, independently constructed exposure measure). Panel B replaces adoption in the baseline horse race with alternative aggregations of task usage to the occupation: the mean usage share used throughout, the sum of task usage (which also weights by the number of tasks), and the rank of adoption; the sum is the one aggregation that is not significant. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

9 Geography and heterogeneity

Table 7, Panel A, moves from occupations to places. In a panel of 51 states by month, a one standard deviation increase in a state's AI exposure, measured from its occupational composition, is associated with a 6.3 log point larger decline in postings after the release, with variation across states that shares none of the cross sectional structure of the sector design. Appendix Figure A4 shows the corresponding event study. The state exposure measure is a shift share object: it is the state's occupational composition weighted by the national exposure of each occupation, so its variation comes from differences in the mix of occupations.[14] That being said, the state result is corroboration from an independent source of variation rather than a separate identification.

The geographic design also answers the inference concern that the sector treatment is concentrated in a few sectors. Moving from 51 states to 489 metropolitan areas spreads the identifying variation across many units: the effective number of clusters, the inverse Herfindahl index of the metros' leverage, is 166 of 489, against 1.2 for the standardized adoption measure at the sector level. In a metro month panel with metro and month fixed effects, a one standard deviation increase in a metro's AI exposure is associated with a decline in postings of 2.7 log points after the release, precise at the one tenth of one percent level (Appendix Table A13). Exposure here is the Felten-Raj-Seamans geographic index of the metro's central county, standardized across metros and matched for 489 of the 593 metros in the Indeed data; it is measured at the central county rather than the whole area, an approximation that attenuates the estimate. The event study in Appendix Figure A8 shows a monotonic decline after the release, reaching about 4 log points within two years, on a prerelease path that is not flat but rising, so removing a linear extrapolation of the prior trend strengthens the effect rather than halving it, to about 10 log points. Geographic adoption

[14] Read through Goldsmith-Pinkham et al. (2020) and Borusyak et al. (2022), the design identifies a labor market effect under the assumption that the occupational mix of a state is uncorrelated with the state's postings trajectory after the release for reasons other than AI. That assumption is not testable, and the same technology cycle that troubles the sector design troubles it here, since states with more cognitive occupations both adopt AI more and were more exposed to the correction.

cannot be measured, since the adoption index is defined over occupations and not places, so the geographic designs necessarily use exposure, which at the level of a place stands in for the concentration of AI related occupations and is rank correlated with adoption at 0.76; they do not separate adoption from exposure, which only the sector horse race can. This design shares the confound of the state design, since metros with more exposed occupations both adopt AI more and were more exposed to the technology correction, so it is corroboration rather than a clean experiment; its value is that the finding survives at a level of geographic detail where no single unit carries the variation.

Panel B estimates the horse race within terciles of wage, teleworkability, and the language task share. Many cells are imprecise, because within a tercile the variation in adoption is small, so the exercise has limited power. The significant estimates concentrate where one would expect: in the middle of the wage distribution, at -0.30, rather than at the manual bottom, which sees little adoption, or at the top, which varies little in adoption; and in the most language intensive third of sectors, at -0.09 and precisely estimated. The concentration in the middle rather than the top is evidence against the pure reversion story, since the 2022 and 2023 correction fell most heavily on the high wage technology occupations at the top. Appendix Table A10 gives the full set. Splitting the sample into terciles leaves little variation in adoption within each tercile, so the within tercile horse race has weak power and several cells are wide. The pattern that survives the imprecision is that the significant negative estimates fall in the middle of the wage distribution and in the most language intensive sectors, where cognitive but not elite work concentrates, and not at the manual bottom, which adopts little, nor cleanly at the professional top, which varies little in adoption. This is the incidence the framework predicts, and the opposite of the manual and regional incidence of earlier automation.

Table 7: Geography and heterogeneity

A. State geographic design (AIGE × Post)			
Estimate (s.e.)	-0.063*** (0.012); $N = 3,060$, 51 states		
B. Heterogeneity: adoption × Post within terciles			
Dimension	Low	Mid	High
Wage	0.012	-0.302***	-0.129
Teleworkability	-0.795	-0.331***	-0.086***
Language-task share	-0.691	-0.742	-0.093***

Notes. Panel A reports the state geographic design: a panel of 51 states by month with state and month fixed effects and standard errors clustered by state, in which state AI exposure, the states occupational composition weighted by the national exposure of each occupation, is interacted with Post. Panel B re-estimates the baseline sector horse race within terciles of each listed characteristic, reporting the adoption interacted with Post coefficient in the low, middle and high tercile. Within a tercile the variation in adoption is small, so several cells are imprecise; the significant negative effects fall in the middle of the wage distribution and in the most language intensive sectors, where cognitive but not elite work concentrates. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

10 Interpretation

The week fixed effects absorb whatever happened to postings on average, so the coefficients describe how intensively adopted sectors moved relative to others. A relative decline of a few to ten log points per standard deviation is consistent with the small aggregate effects that Humlum and Vestergaard (2025) find in administrative data, because reallocation across occupations can leave the total nearly unchanged. The magnitude sits in the range of the sharper declines that Hui et al. (2024) find for the most exposed freelance work once the difference between a platform and the whole economy is taken into account, and it aligns with the concentrated employment effects that Hampole et al. (2025) document when task-level exposure is separated from broader occupational measures. It is not directly comparable to the level displacement that industrial robots produced (Acemoglu and Restrepo, 2020), which a design with time effects removes. Placed against the nearest estimates, the magnitude is reasonable. Hui et al. (2024) find that the release reduced the number of freelance jobs in the most exposed occupations by about 2 percent and their earnings by about 5 percent on one platform; expressed per standard deviation and scaled from a platform of highly substitutable tasks to the whole economy, that is of the same order as the relative postings decline here. Humlum and Vestergaard (2025) find small effects of chatbot access on earnings

and hours across occupations, consistent with a reallocation that leaves the aggregate roughly unchanged. The three estimates are given that this paper measures a relative movement across occupations, not an aggregate.

Exposure is smooth because it averages capability over tasks; adoption is lumpy because it clusters where deployment is cheap and complements exist. In the task framework, labor demand responds to the tasks a technology performs, not the tasks it could perform (Acemoglu and Restrepo, 2018, 2022), so a measure of where delegation actually happened predicts the response and a measure of mere delegability does not. Software development, where deployment is cheap, feedback is immediate, and the workforce integrated assistants quickly (Peng et al., 2023; Bick et al., 2024), is both the most adopted sector and among the sharpest decliners.

Neither exposure nor adoption presumes that AI displaces labor. Both locate where a technology can be or is applied; whether that use replaces a worker or makes one more productive is a separate question, and the framework leaves the sign to the data. In the adoption data most usage is collaborative rather than a full handoff of a task. That most use is collaborative does not imply that employment is unaffected: if augmentation raises output per worker, firms can meet demand with fewer new hires, which reduces vacancies even where no job is fully automated, the productivity channel that Brynjolfsson et al. (2025) and Noy and Zhang (2023) document at the task level. The data here cannot separate augmentation from automation at the occupation level, and I do not claim to.

The concentration of adoption is a snapshot of a technology early in its diffusion, not a fixed feature. New general technologies spread slowly and unevenly because they require complementary investment, reorganization, and learning, and their measured effect grows as adoption deepens and broadens (Brynjolfsson et al., 2025; Bick et al., 2024). Enterprise records of how organizations use the technology show the same concentration at the firm level, with early use clustered in larger, research intensive companies and a few task types (Chatterji et al., 2026a), and the reorganization of firms around the technology is itself a

growing field of study (Chatterji et al., 2026b). If adoption continues to concentrate where deployment is cheapest before spreading to occupations where the frictions are higher, the divergence between exposure and adoption documented here will widen before it narrows, and the labor market effect will stay concentrated in a way that exposure indices do not capture. Tracking adoption over time, as the underlying index is designed to allow, is the natural way to follow it.

The incidence differs from that of the previous wave. Routine biased technical change displaced routine manual and clerical work in the middle of the wage distribution and, in its robotic form, fell on lower and middle wage manufacturing regions (Autor et al., 2003; Autor and Dorn, 2013; Acemoglu and Restrepo, 2020). Generative AI is used most in cognitive occupations, and within them the adjustment falls on the middle rather than the bottom, because the bottom is manual and sees little adoption. The occupations most affected are therefore different from those earlier automation reached, which is one reason exposure measures calibrated to earlier technologies can misplace the incidence of the current one. Whether the middle of the distribution stays where the adjustment falls is an open question, and the history of automation bears on it. Earlier waves displaced tasks and then, over time, reinstated labor into new tasks that raised demand for different skills (Autor et al., 2024; Deming, 2017). If generative AI follows that path, the cognitive occupations that see falling vacancies now may be the ones around which new tasks form later, and the short run incidence documented here would understate the eventual reallocation. Neither measure can capture that reallocation, because exposure and adoption are both defined over the occupations that exist today, while most current employment is in occupational specialties that emerged after 1940 (Autor et al., 2024) and the categories for the work that AI will create are not yet named. The paper therefore confines itself to the first three years and to a relative statement.

The adoption measure reflects the usage of one model family over 2024 and 2025, and the outcome is vacancy postings on one large platform. Both are broad, since Claude usage

spans the occupational spectrum and Indeed is the largest job site in the United States, but neither is the whole economy, and the mapping from postings to employment depends on how vacancies translate into hires, which can itself shift when hiring becomes more efficient. Postings are the right outcome for dating a change and the wrong one for measuring its ultimate size. They lead employment, so they show the adjustment first, which is what a study of the first three years needs; but a fall in postings can reflect more efficient hiring as much as less hiring, and the translation into headcount depends on how vacancies fill. The paper reads the postings decline as evidence that labor demand moved, and moved with adoption, and leaves the mapping to employment to data that observe it directly. The estimates are evidence on the relative movement of labor demand across occupations, which is what the design identifies, and not a measure of aggregate employment loss.

Exposure indices are the workhorse input to the assessments of the workforce at risk that governments and international institutions have produced (Cazzaniga et al., 2024), and they are used to target retraining and to forecast displacement. If exposure and adoption diverge as sharply as they do here, and if the labor market follows adoption, then policy built on exposure alone will misplace both the incidence and the concentration of the effect, spreading attention evenly across occupations that are exposed on paper while the realized adjustment concentrates in a short list of occupations where the technology was cheap to deploy. Pairing exposure with behavioral adoption data, now that such data exist, is the natural correction, and it is likely to grow more important as adoption diffuses unevenly (Bick et al., 2024; Babina et al., 2024) and the gap between what AI could do and what it is used for widens before it narrows.

11 Conclusion

Occupational exposure to artificial intelligence and realized adoption of artificial intelligence are not the same object. They order occupations similarly but differ sharply in intensity,

because adoption is concentrated and exposure is smooth, and after the release of ChatGPT it is adoption, not exposure, that has tracked the decline in job postings. The relationship survives a rich set of baseline controls and an Oster bound, tests for reverse causality, placebo releases, an anticipation window, randomization inference, leave one out across all 47 sectors, and independent geographic designs across 51 states and 489 metropolitan areas. Its magnitude is sensitive to how the prerelease trend is treated, ranging from about three to ten log points per standard deviation, and no credible instrument is available to sharpen it further; the sign and the ranking of adoption over exposure are robust throughout.

The limits are as important as the finding. The window is short. The adoption measure reflects one model family and is observed after the release. And the occupations that adopt AI most are the cognitive occupations that boomed and corrected over exactly this period, so the level of the decline cannot be assigned to AI; only the comparison between adoption and exposure, which holds the common cycle roughly fixed, supports a reading in which AI plays a part. Sharper identification will require variation in AI access that is unrelated to the cycle or firm level adoption data with a credible comparison group.

Three extensions are within reach as the data accumulate. Adoption measured for additional model families would test whether the pattern is specific to one system. A longer window would separate the short run adjustment from the slower reallocation and creation of new work that earlier technologies eventually produced (Autor et al., 2024). And a setting in which AI access was switched off for reasons unrelated to hiring, such as a temporary national ban, or firm records of which employers deployed AI matched to their vacancies, would deliver the exogenous variation the present design lacks. Each would sharpen a magnitude that this paper can only bound.

Until that variation is available, the contribution stands on measurement. Two indices meant to capture the same thing, an occupation's relationship to artificial intelligence, come apart, one describing what the technology could do and the other what it is used for, and the labor market has followed the second. As behavioral records of how AI is used accumulate,

and as frontier models themselves become instruments for reading unstructured economic activity, pairing what a technology could do with what it is used for is the discipline that keeps measurement honest. This does not depend on resolving the identification of the magnitude, and it has a direct implication for anyone who forecasts the labor market effects of AI from exposure alone.

What the paper establishes is narrower: measured adoption and measured exposure diverge, and it is adoption that has tracked the adjustment so far. For anyone who anticipates the labor market effects of AI from exposure alone, that is enough to matter.

References

Acemoglu, D., Autor, D., Hazell, J., and Restrepo, P. (2022). Artificial intelligence and jobs: Evidence from online vacancies. *Journal of Labor Economics*, 40(S1):S293–S340.

Acemoglu, D. and Restrepo, P. (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review*, 108(6):1488–1542.

Acemoglu, D. and Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2):3–30.

Acemoglu, D. and Restrepo, P. (2020). Robots and jobs: Evidence from US labor markets. *Journal of Political Economy*, 128(6):2188–2244.

Acemoglu, D. and Restrepo, P. (2022). Tasks, automation, and the rise in US wage inequality. *Econometrica*, 90(5):1973–2016.

Anthropic (2025). The Anthropic Economic Index. Technical report, Anthropic, research report and dataset.

Appel, R., McCrory, P., Tamkin, A., et al. (2025). The anthropic economic index. Technical report, Anthropic Research. Accessed September 2026.

Autor, D., Chin, C., Salomons, A., and Seegmiller, B. (2024). New frontiers: The origins and content of new work, 1940–2018. *Quarterly Journal of Economics*, 139(3):1399–1465.

Autor, D. H. (2015). Why are there still so many jobs? the history and future of workplace automation. *Journal of Economic Perspectives*, 29(3):3–30.

Autor, D. H. and Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the US labor market. *American Economic Review*, 103(5):1553–1597.

Autor, D. H., Dorn, D., and Hanson, G. H. (2013). The China syndrome: Local labor market effects of import competition in the United States. *American Economic Review*, 103(6):2121–2168.

Autor, D. H., Levy, F., and Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics*, 118(4):1279–1333.

Babina, T., Fedyk, A., He, A. X., and Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, 151.

Bick, A., Blandin, A., and Deming, D. J. (2024). The rapid adoption of generative AI. Technical report, National Bureau of Economic Research, Working Paper 32966.

Bonfiglioli, A., Crinò, R., Gancia, G., and Papadakis, I. (2025). Artificial intelligence and jobs: Evidence from US commuting zones. *Economic Policy*, 40(121):145–194.

Borusyak, K., Hull, P., and Jaravel, X. (2022). Quasi-experimental shift-share research designs. *Review of Economic Studies*, 89(1):181–213.

Brynjolfsson, E., Li, D., and Raymond, L. R. (2025). Generative AI at work. *Quarterly Journal of Economics*, 140(2):889–942.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90(3):414–427.

Cazzaniga, M., Jaumotte, F., Li, L., Melina, G., Panton, A. J., Pizzinelli, C., Rockall, E. J., and Tavares, M. M. (2024). Gen-AI: Artificial intelligence and the future of work. Technical report, International Monetary Fund, Staff Discussion Note SDN/2024/001.

Chatterji, A., Holtz, D., Rakholia, N., Tambe, P., and Weeratunga, G. (2026a). How organizations use AI: Evidence from ChatGPT. *arXiv preprint arXiv:2608.12236*.

Chatterji, A., Rock, D., and Talamàs, E. (2026b). Transformative AI and firms. In Agrawal, A. K., Brynjolfsson, E., and Korinek, A., editors, *The Economics of Transformative AI*. University of Chicago Press.

de Chaisemartin, C. and D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *Quarterly Journal of Economics*, 132(4):1593–1640.

Dingel, J. I. and Neiman, B. (2020). How many jobs can be done at home? *Journal of Public Economics*, 189:104235.

Eloundou, T., Manning, S., Mishkin, P., and Rock, D. (2024). GPTs are GPTs: Labor market impact potential of large language models. *Science*, 384(6702):1306–1308.

Felten, E. W., Raj, M., and Seamans, R. (2021). Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. *Strategic Management Journal*, 42(12):2195–2217.

Felten, E. W., Raj, M., and Seamans, R. (2023). How will language modelers like ChatGPT affect occupations and industries? Technical report, arXiv:2303.01157.

Frey, C. B. and Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114:254–280.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2020). Bartik instruments: What, when, why, and how. *American Economic Review*, 110(8):2586–2624.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

Hampole, M., Papanikolaou, D., Schmidt, L. D., and Seegmiller, B. (2025). Artificial intelligence and the labor market. Technical report, National Bureau of Economic Research, Working Paper 33509.

Hui, X., Reshef, O., and Zhou, L. (2024). The short-term effects of generative artificial intelligence on employment: Evidence from an online labor market. *Organization Science*. forthcoming.

Humlum, A. and Vestergaard, E. (2025). Large language models, small labor market effects. Technical report, National Bureau of Economic Research, Working Paper 33777.

Indeed Hiring Lab (2026). Job postings on Indeed: United States data. Technical report, Indeed Hiring Lab, public dataset.

MacKinnon, J. G. and Webb, M. D. (2018). The wild bootstrap for few (treated) clusters. *The Econometrics Journal*, 21(2):114–135.

Noy, S. and Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654):187–192.

Peng, S., Kalliamvakou, E., Cihon, P., and Demirer, M. (2023). The impact of AI on developer productivity: Evidence from GitHub Copilot. Technical report, arXiv:2302.06590.

Rambachan, A. and Roth, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies*, 90(5):2555–2591.

U.S. Department of Labor, Employment and Training Administration (2026). O*NET On-Line.

Webb, M. (2020). The impact of artificial intelligence on the labor market. Technical report, Stanford University, working paper (SSRN 3482150).

Appendix

Table A1: Summary statistics and correlations of AI measures

A. Panel variables (sector-week, 2021–2025; 12,267 obs)				
Variable	Mean	S.D.	Min	Max
Log job postings	4.85	0.27	4.06	5.49
Postings index (level)	132.69	33.47	58.19	242.91
New postings index	143.77	39.33	57.92	509.17
AI exposure (AIOE)	0.42	0.83	-1.45	1.48
AI adoption (AEI)	0.01	0.04	0.00	0.25
Teleworkable share	0.52	0.42	0.00	1.00
B. Correlations across the 47 sectors				
	Expo.	Adopt.	Lang.	Telew.
AI exposure	1.00			
AI adoption	0.25	1.00		
Language share	0.89	0.25	1.00	
Teleworkable	0.80	0.28	0.82	1.00

Notes. Panel A reports the mean, standard deviation, minimum and maximum of each variable over the weekly panel of 47 occupational sectors from January 2021 through December 2025 (12,267 sector weeks), in native units rather than standardized. Log job postings is the natural logarithm of the Indeed Hiring Lab postings index, which equals 100 on 1 February 2020 for every sector; the postings index and the new postings index are that index and its recently created vacancies variant. AI exposure is the Felten, Raj and Seamans AI Occupational Exposure index (AIOE) aggregated to sectors; AI adoption is the Anthropic Economic Index mean task usage share; the teleworkable share follows Dingel and Neiman. Panel B reports pairwise Pearson correlations across the 47 sectors, one observation per sector. The exposure and adoption correlation is 0.25 in levels, as shown, and 0.76 in ranks (Spearman). In the regressions that follow, exposure and adoption are standardized to mean zero and unit standard deviation across sectors.

A. Data construction

Adoption is built by normalizing O*NET task text, matching it to the Anthropic Economic Index task usage shares, aggregating to occupations as the mean usage share across an occupation's tasks, and then to sectors through the crosswalk in Table A4. Language and physical task shares classify each task by keyword. Teleworkability follows Dingel and Neiman (2020); wage and education come from the Occupational Outlook Handbook. All standardized variables are computed across the 47 sectors. The full crosswalk and the code that builds every variable ship with the replication files.

B. Inference with few clusters

Because the panel has 47 sectors, inference deserves care. The baseline standard errors cluster by sector, which allows for arbitrary correlation of the residuals within a sector over the roughly 300 weeks of the panel. This correlation is large: the heteroskedasticity robust standard error on the adoption coefficient, which treats the 12,267 sector weeks as independent, is about 0.0025, seven times smaller than the clustered standard error of 0.018, and it is not the right benchmark, since postings within a sector are strongly serially correlated. With a moderate number of clusters the asymptotic approximation behind the clustered standard error can still be optimistic, and two responses are standard. The first is the wild cluster bootstrap of Cameron et al. (2008), which improves the finite sample behavior of cluster robust inference. The second, which I adopt as the primary test, is randomization inference: I hold the panel fixed and permute the assignment of the exposure and adoption values across the 47 sectors, jointly, so that their cross sectional correlation is preserved, and estimate the horse race again for each of 499 permutations. The observed adoption coefficient lies beyond all but one of the permuted coefficients in absolute value, for a two sided p value of 0.002, while the exposure coefficient sits near the center of its permutation distribution, at $p = 0.44$. Randomization inference makes no appeal to asymptotics in the number of clusters and is the appropriate benchmark when the treatment is assigned at the level of a small number of units.

Whether the wild cluster bootstrap or randomization inference is the more reliable of the two here depends on the leverage of the treatment, which the concentration of adoption makes extreme. Standardized adoption places software development about six and a half standard deviations above the mean, so a single sector accounts for 92 percent of the cross sectional variance of the level measure, and the effective number of clusters carrying treatment variation, the inverse Herfindahl index of the sectors' leverage, is about 1.2. MacKinnon and Webb (2018) show that the wild cluster bootstrap under rejects badly when one cluster has this much leverage, and it does: for adoption in levels its p value is 0.64, in

conflict with a clustered t statistic above five and a randomization inference p value of 0.002. Under high leverage the bootstrap is the test to distrust, not the estimate, and randomization inference, which conditions on the realized treatment values, is the right benchmark. Replacing the level measure with its rank removes the outlier's leverage: the effective number of clusters rises to 26 of 47, and the rank coefficient of -0.156 is significant under the clustered t ($p = 0.003$), the wild cluster bootstrap ($p = 0.007$), and randomization inference ($p = 0.005$), and survives dropping software development (-0.134, $p = 0.006$). Table A12 reports these diagnostics side by side. Both measures agree on the sign, the ranking of adoption over exposure, and significance under the test appropriate to each; the level measure delivers the magnitude while resting its precision on the most adopted sectors, and the rank measure shows the result is not the work of a single sector. Table A12 appears with the other appendix tables at the end.

C. Classifying tasks

The language and physical task shares are built from the text of the O*NET task statements. A task is coded as language oriented if its text contains any of a list of terms for cognitive and communicative work, such as write, document, analyze, review, prepare, research, compute, design, develop, communicate, evaluate, and instruct; it is coded as physical if its text contains any of a list of terms for manual work, such as operate, install, repair, lift, clean, drive, assemble, and build. An occupation's language share is the fraction of its tasks that are language oriented, and its physical share the fraction that are physical. The classification is coarse, but it has face validity: the language share is highest for occupations such as computer programmers, credit analysts, and training specialists, and lowest for machine operators and material movers, and it correlates with realized adoption at 0.28 across occupations. The language share is the natural candidate instrument for adoption, since a language model performs language tasks, and Appendix Table A11 reports that it is too weak and too collinear with exposure to serve as one.

D. The Oster bound

The Oster (2019) procedure asks how much stronger selection on unobservables would have to be than selection on the observed controls to drive the treatment coefficient to zero. Let β° and $\tilde{\beta}$ be the adoption coefficients, and R° and $\tilde{R}$ the within R^2 values, from the specification with only the fixed effects and adoption and from the specification that adds the full control battery. Let R_{max} be the R^2 that would obtain if all variation, including that from unobservables, were explained. The degree of proportional selection that sets the treatment effect to zero is $\delta = \tilde{\beta}(\tilde{R} - R^{\circ})/[(\beta^{\circ} - \tilde{\beta})(R_{max} - \tilde{R})]$, and a value above one means that unobservables would have to be more important than the rich observed controls, which is usually treated as implausible. I compute δ and the bias adjusted coefficient at the conventional benchmark $R_{max} = 1.3\tilde{R}$ and at the more demanding $R_{max} = 2\tilde{R}$; both are reported in Appendix Table A11. At $R_{max} = 1.3\tilde{R}$, $\delta = 4.1$ and the bias adjusted coefficient is -0.064; at $R_{max} = 2\tilde{R}$, $\delta = 1.2$ and the coefficient stays negative. The bound is uninformative only at the extreme $R_{max} = 1$, in which unobservables are assumed to explain every remaining piece of variation in postings.

E. Prior trend sensitivity

The prior trend sensitivity in Table 4, Panel D, fits a line to the adoption event study coefficients over the twelve months before the release, extrapolates it forward, and subtracts it from the raw post coefficients. This is a transparent linear version of the sensitivity analysis of Rambachan and Roth (2023), which formalizes the idea that a credible bound on a difference in differences estimate should allow the counterfactual trend to differ from the extrapolated prior trend by a bounded amount. The linear case already shows the direction: the average post effect falls from -0.071 to -0.031 and remains negative, so a modest continuation of the prior trend does not overturn the sign, while a larger allowed violation would. The horse race is less sensitive than the level of either coefficient, because exposure and adoption share the common prior trend and it differences out of their comparison.

Table A2: Data sources and definitions

Variable	Source	Definition
Job postings	Indeed Hiring Lab	Postings index by sector, Feb 2020=100, weekly
AI exposure	Felten et al. (2021)	AIOE by occupation; AIGE by state
AI adoption	Anthropic Economic Index (2025)	Mean model usage share across occupation tasks
Language share	O*NET task text	Fraction of tasks that are language/cognitive
Teleworkable	Dingel–Neiman (2020)	Share of tasks doable from home
Education, wage	OOH	Entry education, median annual wage

Notes. All inputs are public and reachable from open repositories.

Table A3: The 47 occupational sectors and their AI measures

Sector	AIOE	Actual use	Teleworkable	# SOC
Software Development	1.21	0.252	1.00	4
Media & Communications	1.06	0.041	0.67	9
IT Systems & Solutions	1.22	0.041	1.00	2
Education & Instruction	1.06	0.021	0.96	51
Data & Analytics	1.38	0.020	1.00	1
Accounting	1.48	0.016	1.00	1
Administrative Assistance	0.71	0.015	1.00	14
Marketing	1.37	0.014	1.00	2
IT Infrastructure, Operations & Support	0.54	0.013	1.00	2
Social Science	1.30	0.011	0.81	13
Project Management	1.06	0.008	1.00	1
Customer Service	0.96	0.007	0.50	1
Banking & Finance	1.35	0.007	0.77	13
Arts & Entertainment	-0.09	0.007	0.59	20
Legal	1.31	0.007	0.88	8
Scientific Research & Development	0.83	0.006	0.60	43
Industrial Engineering	1.14	0.005	0.50	1
Retail	-0.04	0.005	0.00	6
Community & Social Service	0.81	0.005	0.42	12
Sales	0.88	0.004	0.64	14
Management	0.96	0.004	0.70	34
Architecture	1.04	0.004	1.00	2
Electrical Engineering	0.90	0.003	1.00	2
Medical Information	0.90	0.003	1.00	2
Physicians & Surgeons	0.71	0.003	0.14	8
Personal Care & Home Health	-0.57	0.003	0.00	1
Nursing	-0.03	0.003	0.00	3
Human Resources	0.89	0.002	1.00	4
Mechanical Engineering	1.00	0.002	0.33	1
Pharmacy	0.31	0.002	0.00	2
Civil Engineering	1.28	0.002	1.00	1
Insurance	1.33	0.002	1.00	2
Therapy	-0.17	0.002	0.15	11
Food Preparation & Service	-0.77	0.002	0.06	16
Production & Manufacturing	-0.83	0.002	0.04	103
Logistic Support	-0.07	0.001	0.33	13
Hospitality & Tourism	-0.21	0.001	0.00	2
Dental	-0.35	0.001	0.00	2
Medical Technician	-0.12	0.001	0.05	22
Childcare	-0.38	0.001	1.00	1
Installation & Maintenance	-0.88	0.001	0.02	51
Driving	-0.77	0.000	0.00	7
Security & Public Safety	-0.36	0.000	0.15	21
Aviation	-0.03	0.000	0.17	6
Construction	-1.45	0.000	0.00	30
Cleaning & Sanitation	-1.21	0.000	0.00	3
Loading & Stocking	-1.16	0.000	0.01	18

Notes. Sectors ordered by actual AI use intensity. AIOE is standardized in the regressions but shown here in native units.

Table A4: Crosswalk from Indeed sectors to occupation codes

Sector	Representative SOC codes
Accounting	13-201
Administrative Assistance	43-601, 43-901, 43-902, 43-903, 43-904, 43-905 ...
Architecture	17-101
Arts & Entertainment	27-101, 27-102, 27-201, 27-202, 27-203, 27-204 ...
Aviation	49-301, 53-201, 53-202, 53-203
Banking & Finance	13-201, 13-202, 13-203, 13-204, 13-205, 13-206 ...
Childcare	39-901
Civil Engineering	17-205
Cleaning & Sanitation	37-201, 37-202
Community & Social Service	21-101, 21-102, 21-109 ...
Construction	47-201, 47-202, 47-203, 47-204, 47-205, 47-206 ...
Customer Service	43-405
Data & Analytics	15-204
Dental	29-102, 31-909
Driving	53-301, 53-302, 53-303, 53-304 ...
Education & Instruction	25-101, 25-102, 25-103, 25-104, 25-105, 25-106 ...
Electrical Engineering	17-207
Food Preparation & Service	35-101, 35-201, 35-202, 35-301, 35-302, 35-303 ...
Hospitality & Tourism	39-601
Human Resources	11-312, 13-107
IT Infrastructure, Operations & Support	15-115
IT Systems & Solutions	15-112
Industrial Engineering	17-211
Installation & Maintenance	49-101, 49-201, 49-202, 49-209, 49-301, 49-302 ...
Insurance	13-205, 41-302
Legal	23-101, 23-102, 23-201, 23-209 ...
Loading & Stocking	43-508, 53-701, 53-702, 53-703, 53-704, 53-705 ...
Logistic Support	13-108, 43-501, 43-502, 43-503, 43-504, 43-505 ...
Management	11-101, 11-102, 11-103, 11-201, 11-202, 11-203 ...
Marketing	11-202, 13-116
Mechanical Engineering	17-214
Media & Communications	27-301, 27-302, 27-303, 27-304, 27-309 ...
Medical Information	29-207, 43-601
Medical Technician	29-201, 29-202, 29-203, 29-204, 29-205, 29-206 ...
Nursing	29-114, 29-117, 29-206
Personal Care & Home Health	39-902
Pharmacy	29-105, 29-205
Physicians & Surgeons	29-106 ...
Production & Manufacturing	51-101, 51-201, 51-202, 51-203, 51-204, 51-209 ...
Project Management	11-302
Retail	41-101, 41-201, 41-202, 41-203
Sales	41-301, 41-302, 41-303, 41-304, 41-309, 41-401 ...
Scientific Research & Development	19-101, 19-102, 19-103, 19-104, 19-201, 19-202 ...
Security & Public Safety	33-101, 33-102, 33-201, 33-202, 33-301, 33-302 ...
Social Science	19-301, 19-302, 19-303, 19-304, 19-305, 19-309 ...
Software Development	15-113
Therapy	29-112, 31-201, 31-202 ...

Notes. Each sector is mapped to the occupation code prefixes shown; sector measures are means over matched codes.

Table A5: Adoption coefficient as controls are added (full)

	(1)	(2)	(3)	(4)	(5)	(6)
Adoption × Post	-0.108***	-0.102***	-0.098***	-0.094***	-0.085***	-0.075***
Controls (each × Post)						
Exposure, wage, education, telework, task shares, pre-boom, trends	added cumulatively, columns (2)–(6)					
Within R^2	0.170	0.179	0.187	0.235	0.256	0.722

Notes. Companion to Table 3.

Table A6: Alternative adoption definitions

Adoption definition	Adoption × Post	p
Adoption: mean share (baseline)	-0.102***	0.000
Adoption: sum of task usage	-0.028	0.663
Adoption: rank	-0.156***	0.002

Notes. Horse race with exposure; mean and rank are significant, the sum is not.

Table A7: Alternative exposure measures

Exposure measure	Adopt.×Post	p	Expo.×Post	p
vs AIOE (baseline)	-0.102***	0.000	-0.027	0.359
vs alternative exposure	-0.103***	0.000	-0.058**	0.014

Notes. Adoption survives against each exposure measure.

Table A8: Leave-one-out estimates, all 47 sectors

Dropped sector	Est.	Dropped sector	Est.
Software Development	-0.424	Accounting	-0.101
Human Resources	-0.105	Nursing	-0.101
Marketing	-0.103	Construction	-0.101
Scientific Research & Development	-0.103	Medical Technician	-0.101
Banking & Finance	-0.103	Architecture	-0.101
Industrial Engineering	-0.103	Dental	-0.101
Medical Information	-0.103	Pharmacy	-0.101
Data & Analytics	-0.102	Production & Manufacturing	-0.101
Customer Service	-0.102	Management	-0.101
Logistic Support	-0.102	Cleaning & Sanitation	-0.101
Education & Instruction	-0.102	Aviation	-0.101
Personal Care & Home Health	-0.102	Therapy	-0.101
Installation & Maintenance	-0.102	Community & Social Service	-0.101
Project Management	-0.102	Legal	-0.100
Food Preparation & Service	-0.102	Loading & Stocking	-0.100
Sales	-0.102	Mechanical Engineering	-0.100
IT Infrastructure, Operations & Support	-0.102	Electrical Engineering	-0.100
Administrative Assistance	-0.102	Insurance	-0.100
Retail	-0.102	Social Science	-0.100
Security & Public Safety	-0.101	Media & Communications	-0.099
Driving	-0.101	IT Systems & Solutions	-0.098
Childcare	-0.101	Physicians & Surgeons	-0.098
Arts & Entertainment	-0.101	Civil Engineering	-0.096
Hospitality & Tourism	-0.101		

Notes. Baseline horse race dropping each sector.

Table A9: Reverse causality: baseline characteristics do not predict adoption

Dependent variable: standardized adoption	Coef.	p
Constant	0.000	1.000
Pre-boom	0.193	0.221
Wage	-0.074	0.762
Education	-0.101	0.717
Teleworkable	0.363	0.053
Physical-task share	-0.002	0.993
R^2	0.12	
Observations	47	

Notes. Cross-sector OLS.

Table A10: Heterogeneity, full

Dimension	Group	Adoption×Post	p	Sectors
Teleworkability	Low	-0.795	0.473	16
Teleworkability	Mid	-0.331***	0.000	15
Teleworkability	High	-0.086***	0.000	16
Language-task share	Low	-0.691	0.359	16
Language-task share	Mid	-0.742	0.112	15
Language-task share	High	-0.093***	0.000	16
Wage	Low	0.012	0.976	13
Wage	Mid	-0.302***	0.002	12
Wage	High	-0.129	0.918	13

Notes. Baseline horse race within terciles.

Table A11: Why no instrumental variable, and full Oster bounds

A. Candidate instruments: first-stage strength		
Instrument	Corr. with adoption	First-stage F
lang_share_z	0.246	2.90
phys_share_z	-0.099	0.45
B. Oster (2019) bounds, full		
R_{max}	δ	Bias-adj. β
0.333	4.06	-0.064
0.512	1.22	-0.015
1.000	0.42	0.118

Notes. Panel A: candidate task-content instruments have first-stage F far below conventional thresholds and are collinear with exposure, so no credible instrument is available. Panel B: Oster bounds across R_{max}; the effect is robust except at the extreme $R_{max} = 1$.

Table A12: Treatment concentration and inference

	Adoption, level	Adoption, rank
Horse-race coefficient	-0.102	-0.156
Clustered standard error	(0.018)	(0.049)
Leverage share of top sector	0.921	0.061
Effective number of clusters (of 47)	1.2	26.1
p-value, clustered t	0.000	0.003
p-value, wild cluster bootstrap	0.635	0.007
p-value, randomization inference	0.002	0.006

Adoption in levels is dominated by a single high-leverage sector (software development, standardized adoption +6.5), so the effective number of clusters carrying treatment variation is near one and the wild cluster bootstrap under-rejects, a known pathology under high leverage (MacKinnon and Webb, 2018). The rank measure spreads leverage across the sectors (effective clusters 26 of 47) and is significant under the clustered t, the wild cluster bootstrap, and randomization inference. Randomization inference, which does not rely on the number of clusters, is the appropriate benchmark for both. Wild cluster bootstrap uses 9,999 Rademacher draws; randomization inference uses 999 joint permutations of the exposure and adoption values across sectors.

Table A13: Metropolitan geographic design

	Metro AI exposure × Post
Coefficient	−0.027
Clustered standard error	(0.005)
p-value	< 0.001
Metropolitan areas	489
Metro months	29,340
Effective number of clusters	166
Average post effect, raw	−0.029
Average post effect, prior trend removed	−0.100

A metro month panel of 489 metropolitan areas, January 2021 through December 2025, with metro and month fixed effects and standard errors clustered by metro. Metro AI exposure is the Felten-Raj-Seamans geographic exposure of the metro's central county, standardized across metros; it is matched for 489 of the 593 metros in the Indeed data (83 percent), and exposure is measured at the central county rather than the whole area. The effective number of clusters, the inverse Herfindahl index of the metros' leverage, is 166 of 489, so the geographic identifying variation is spread across many units rather than concentrated as in the sector design. The prior trend removed line subtracts a linear extrapolation of the eight prerelease quarters, as in the sector event study; here it strengthens the effect rather than halving it, because the prerelease trend is upward. Figure A8 plots the event study.

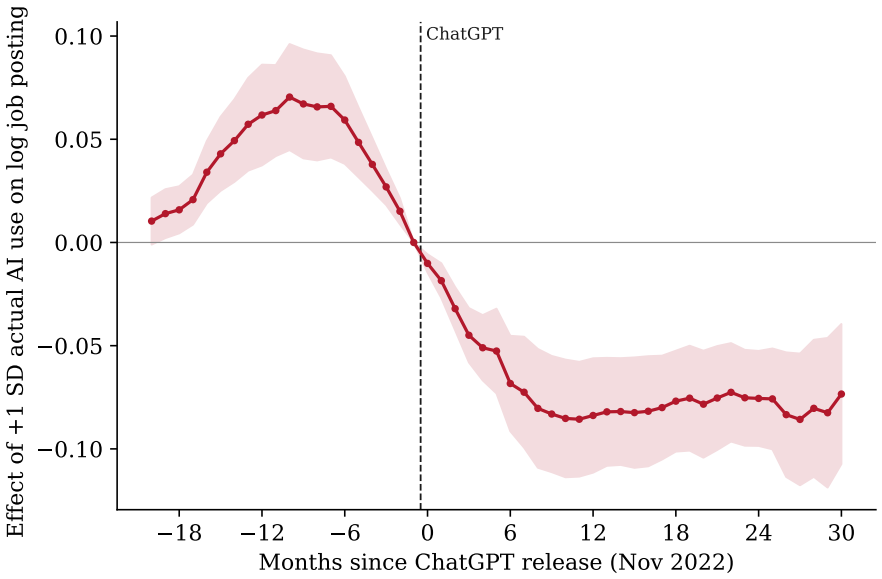


Figure A1: Adoption event study. Coefficients θ_k from equation (2), the interaction of standardized adoption with indicators for each month relative to the ChatGPT release, with the month before the release omitted, on the weekly panel of 47 sectors with sector and week fixed effects. Shaded bands are 95 percent confidence intervals from standard errors clustered by sector. The path rises before the release, the cognitive hiring boom, peaks in the middle of 2022, and falls after the release to about 9 log points below baseline within a year.

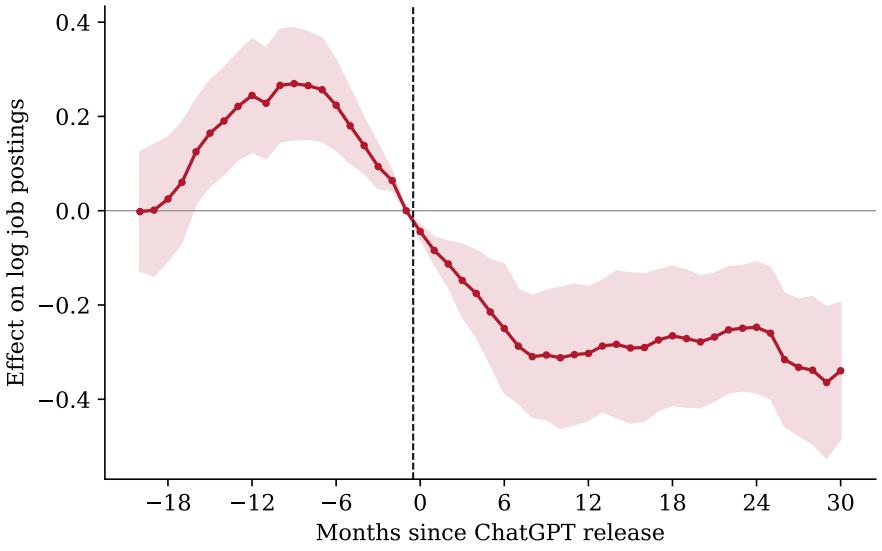


Figure A2: Adoption event study excluding software development. The specification of Figure A1 with the most adopted sector dropped, so the identifying variation comes from the remaining 46 sectors. The decline after the release remains, which shows the reversal is not carried by the single highest leverage sector.

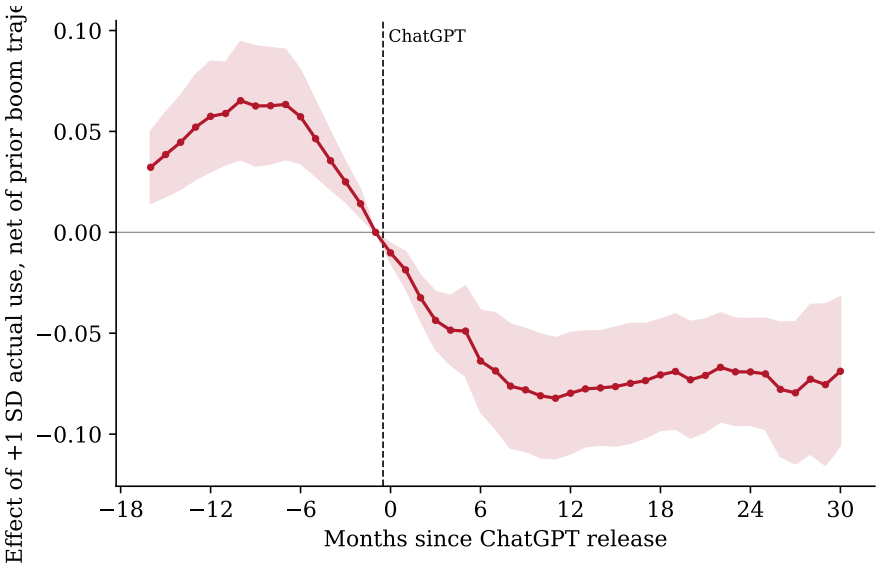


Figure A3: Adoption event study net of the prior boom. The event study of Figure A1 after also interacting each sector's prerelease hiring boom with the event month indicators, so the estimated adoption path is purged of the trajectory the prior boom alone would predict. The decline after the release survives this adjustment, which separates it from a continuation of the cognitive hiring boom.

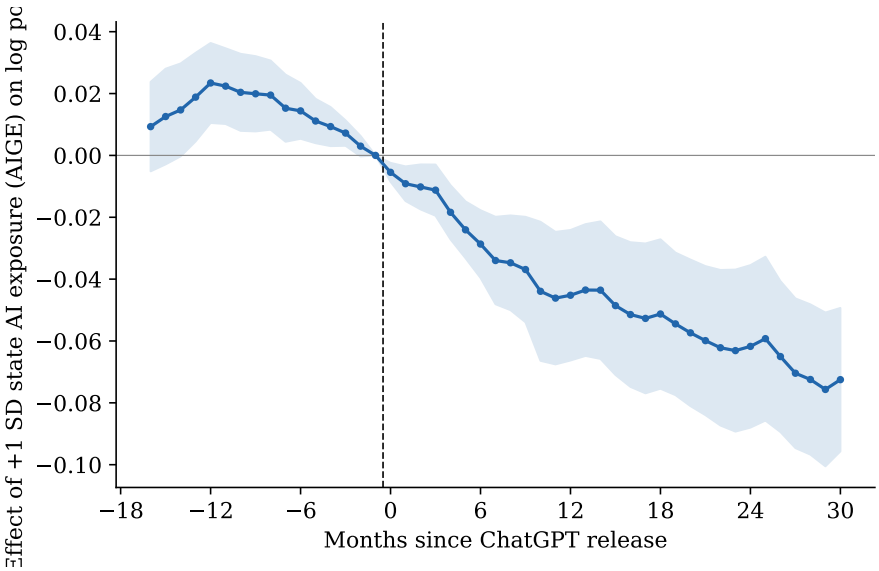


Figure A4: State geographic design event study. Coefficients from the state analogue of equation (2), the interaction of standardized state AI exposure, the AI Geographic Exposure index formed as the state's occupational composition weighted by the national exposure of each occupation, with month indicators, on a panel of 51 states by month with state and month fixed effects and standard errors clustered by state. Bands are 95 percent confidence intervals. The decline after the release mirrors the sector design in an independent source of variation.

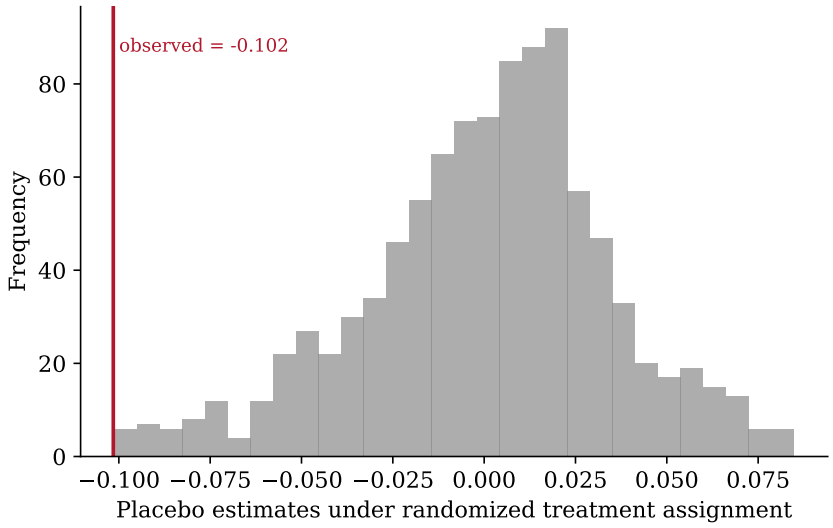


Figure A5: Randomization inference. The distribution of the adoption coefficient across 999 draws that permute the exposure and adoption values jointly across the 47 sectors and estimate the horse race again, preserving the cross sectional correlation of the two measures. The vertical line is the observed estimate, which lies in the extreme tail, for a two sided p value of 0.002. Randomization inference makes no appeal to the number of clusters and is the appropriate benchmark when treatment is assigned across a small number of units.

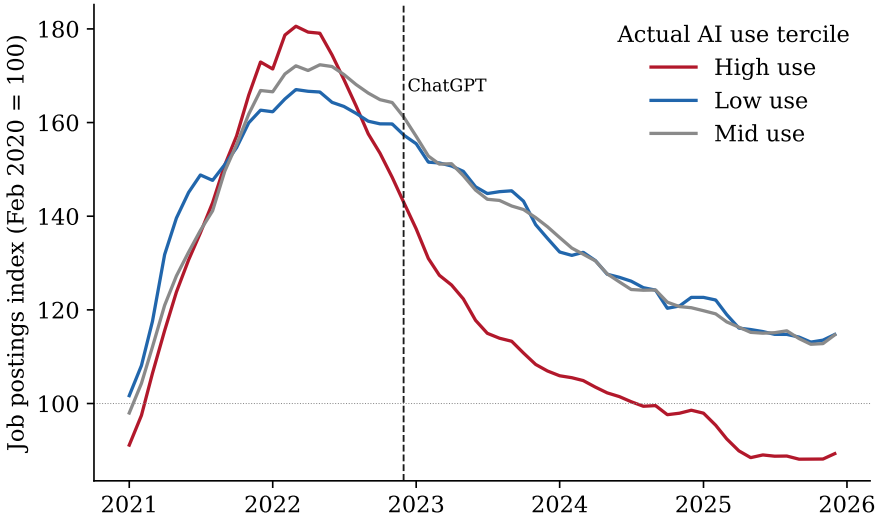


Figure A6: Job postings by adoption tercile, unadjusted. The raw Indeed postings index relative to February 2020, averaged within the most, middle, and least adopted thirds of the 47 sectors, with no regression adjustment. Through 2021 and 2022 the most adopted third posted more vacancies than the least adopted third; the two groups cross in the months around the release; and by 2024 the most adopted third sits below the least adopted third. The crossing that the regression dates and signs is already visible in the unadjusted series.

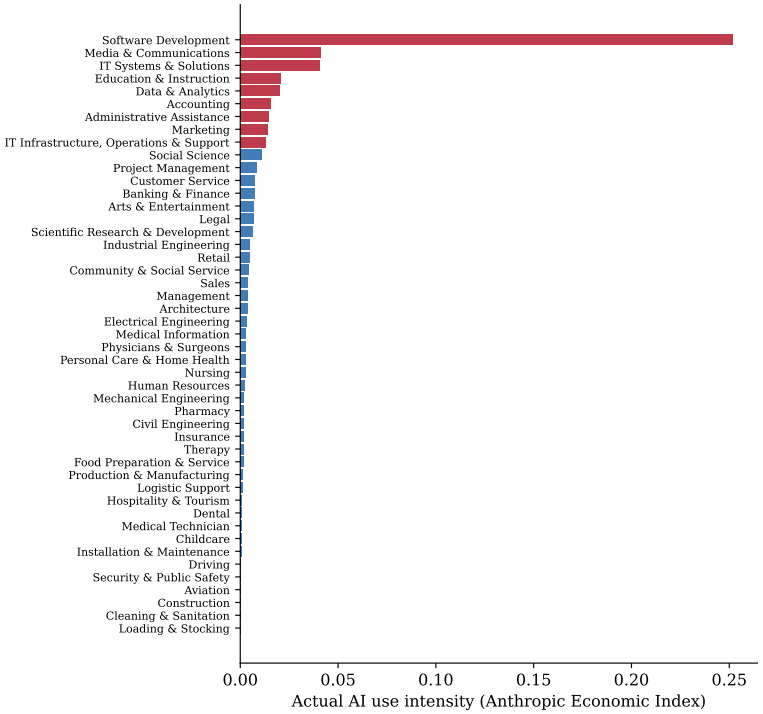


Figure A7: The 47 occupational sectors ranked by realized AI adoption. Adoption is the Anthropic Economic Index mean task usage share. Software development leads by a wide margin, followed by information technology, data and analytics, media and communications, and education, while manual sectors register almost no usage. This is the concentration summarized by the adoption Gini coefficient of 0.76 in Table 1 and by the Lorenz curve in Figure 2.

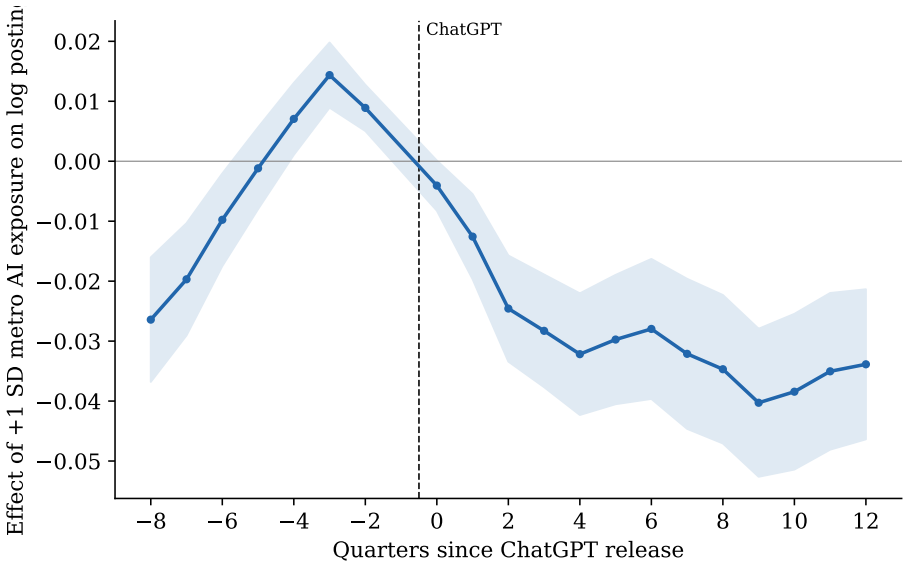


Figure A8: Metropolitan geographic design event study. Coefficients on standardized central county AI exposure interacted with quarter indicators across 489 metropolitan areas, with metro and month fixed effects and standard errors clustered by metro; bands are 95 percent confidence intervals, quarters relative to the release. The decline after the release is monotonic and the prerelease path is rising rather than flat, so removing a linear extrapolation of the prior trend strengthens the estimate rather than halving it. Table A13 reports the corresponding difference in differences coefficient.