

Online Appendix for “Generate” the Future of Work through AI: Empirical Evidence from Online Labor Markets

Jin Liu*

School of Economics and Management, Beijing Jiaotong University
Email: jinliu1@bjtu.edu.cn

Xingchen Xu*

Michael G. Foster School of Business, University of Washington, Seattle,
WA 98195, USA
Email: xcxu21@uw.edu

Xi Nan

Michael G. Foster School of Business, University of Washington, Seattle,
WA 98195
Email: xinan02@uw.edu

Yongjun Li

School of Management, University of Science and Technology of China
Email: lionli@ustc.edu.cn

Yong Tan

Michael G. Foster School of Business, University of Washington, Seattle,
WA 98195, USA
Email: ytan@uw.edu

Appendix A: Summary of Challenges and Robustness Checks

Table A1: Summary of Challenges and Analyses

Challenges	Solutions: Reasoning and Analyses	Related Literature	Related Section
Treatment and Control Group Definition: Our findings may be sensitive to the treatment assignment.	1. Alternative Thresholds: To avoid imposing a specific functional form for the relationship between LM-AIOE and labor-market outcomes, we partition submarkets at the median LM-AIOE value and then vary this cutoff in robustness checks to confirm that our results do not hinge on any particular threshold. We also report results under a continuous treatment specification (assuming a linear relationship) and a three-group partition of submarkets as robustness checks.	(Callaway et al. 2024, Lu and Yu 2015)	Section 3.2 and Appendices N, Z, and AA
	2. Alternative Exposure Measure: We also use an alternative LM-AIOE measure from a different paper to construct the treatment variable and rerun the estimation.	(Eloundou et al. 2024)	
	3. Seasonal Placebo Test: We conduct a placebo test using November 30, 2021, the same day in the previous year, as a pseudo-shock date. The resulting insignificant coefficients suggest that our findings are not driven by year-end effects or seasonal fluctuations, thereby reinforcing the robustness of the main results.	(Gong et al. 2023)	
Potential Seasonal Effects: Since the launch of ChatGPT occurred at the end of the year, holidays and seasonal effects might influence the observed post-shock difference.			Appendix O
Random Fluctuations: The observed effect could plausibly arise from random variation in the data rather than reflecting a genuine treatment impact.	4. Placebo Test (Treatment Reassignment): We conduct a placebo test by randomly reassigning the treatment status ($Treat * Post$) within the sample and estimate the distribution of placebo treatment effects. The resulting distribution approximates a normal distribution centered around zero, indicating that the observed treatment effects are unlikely to be driven by random chance. This further reinforces the robustness of our findings.	(Zhang et al. 2025, Ye and Shankar 2025)	Appendix P
	5. Parallel Trends: We estimate a relative time model to verify that treated and control groups exhibit similar trends prior to treatment. The results show no significant pre-treatment differences between the groups, suggesting that the treatment and control groups are comparable.	(Ananthakrishnan et al. 2025, Angrist and Pischke 2009)	Appendix Q
Treatment and Control Group Comparability: Lack of comparability between treatment and control groups may undermine the validity of the analysis.	6. Propensity Score Matching: We re-estimate the analysis using a propensity-score-matched sample of treated and control submarkets to enhance group comparability and further validate our findings.	(Ran et al. 2025, Ye and Shankar 2025, Faccio and O'Brien 2021)	Appendix R

Continued on next page

Table A1 continued

Challenges	Solutions: Reasoning and Analyses	Related Literature	Related Section
Omitted Variable Bias: Some unobservables might affect both the treatment and outcome variables.	7. Interrupted Time Series Analysis (ITSA): We employ an Interrupted Time Series Analysis to examine the impact exclusively within the treated submarkets, which yields consistent findings.	(Jiang et al. 2025, Al Balawi et al. 2023)	Appendix S
	8. Fixed Effects: If unobservable factors are either unit-invariant or time-specific, our two-way fixed effects in the model account for these influences.	(Chan et al. 2019, Angrist and Pischke 2009)	Section 4
	9. Sensitivity Analyses: We conduct sensitivity analyses by incorporating additional control variables that capture both external environmental conditions and internal market characteristics, demonstrating that our main estimates remain robust and reinforcing the validity of our findings.	(Sen et al. 2024, Petrova et al. 2021, Altonji et al. 2005)	Appendix T
Model Specification: All model specifications rest on particular assumptions, and relying on a single specification may inadvertently introduce bias arising from model selection.	10. Alternative Model for Count Variables: Several of our outcome variables are count data, taking non-negative integer values. To account for their distributional characteristics, we re-estimate the models using a negative binomial specification. The consistent results further confirm the robustness of our findings.	(Gong et al. 2023)	Appendix U
	11. Alternative Aggregation Level: We aggregate all variables at the weekly level and re-estimate the models. The results remain qualitatively consistent with the main findings, confirming the robustness and reliability of our conclusions across different levels of data granularity.	(Ran et al. 2025)	Appendix V
	12. Alternative Fixed Effect: We adopt an alternative model specification with separate fixed effects for year and month to better account for seasonal effects, enabling control over both annual fluctuations and recurring monthly patterns.	(Li and Zhan 2019)	Appendix W
	13. Alternative Outcomes: To further validate our findings, we introduce three additional outcome variables to examine market dynamics from different perspectives, thereby demonstrating the robustness of our results.	(Wang and Overby 2023)	Appendix X
	14. Removing Outliers: We exclude the top 1% of submarkets and those with an average job post count below one. The re-estimated results remain consistent with our main findings.	(Chen et al. 2025)	Appendix Y

Appendix B: Platform Information and Matching Process

The focal platform, a publicly traded entity, stands as one of the preeminent online labor marketplaces, boasting millions of visits each month. As of the close of 2023, it had amassed over 70 million registered users and hosted upwards of 20 million job postings. The platform primarily utilizes a client-driven matching mechanism, which can be elucidated through the flowchart presented in Figure B1:

1. **Project Posting:** A client (denoted as u) posts a job (project) on the platform.
2. **Bid Submission:** Freelancers can browse the available project list and potentially bid on this project.

Should no bids be submitted, the process for this project terminates.

3. **Freelancer Selection:** In cases where one or more freelancers place bids, the client u evaluates these bids and may select a freelancer (denoted as v) for the project.

4. **Project Execution:** The chosen freelancer v undertakes the project. If v does not complete the project, the procedure concludes prematurely.

5. **Payment and Conclusion:** Upon successful project completion, the client compensates the freelancer, effectively finalizing the transaction.

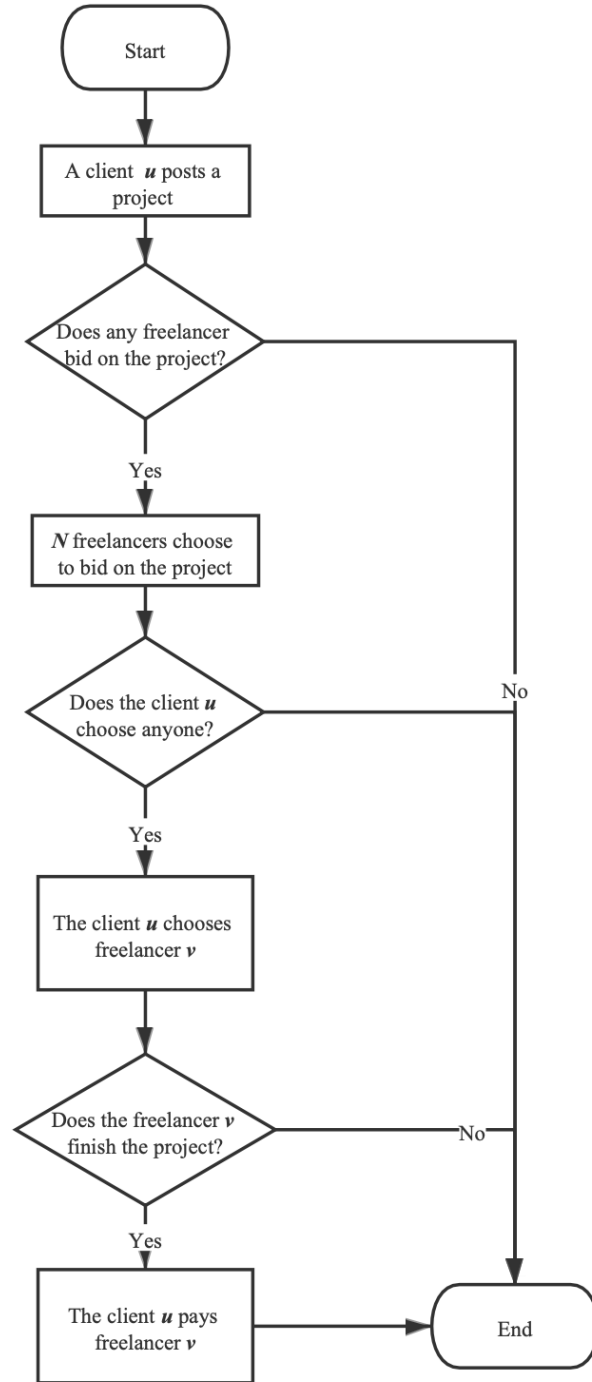


Figure B1 Matching Process

Appendix C: Detailed Job Classification, Submarket Construction Process, and LM-AIOE-Based Treatment Definition

In this appendix, we provide a detailed description of the job classification procedure and the construction of submarkets. Our raw dataset includes all job postings along with their associated skill tags (JT), with many jobs sharing identical sets of skill tags. Jobs that share the same skill set are assumed to belong to a common demand pool and therefore compete for the same group of freelancers. Furthermore, many skill sets exhibit hierarchical or subset relationships. For example, Skill Set A may include [skill tag 1, skill tag 2, skill tag 3, ..., skill tag 7], while Skill Set B consists of only [skill tag 1, skill tag 2, skill tag 3]. Since Skill Set B is a subset of Skill Set A, both are considered part of the same demand pool and are presumed to compete for a similar pool of freelance labor. Building on this logic, we also treat skill tag sets that exhibit a high degree of overlap or semantic similarity as belonging to the same demand pool, assuming they compete for comparable freelancer resources. These skill sets, which share a common demand pool and freelancer pool, can be collectively regarded as a group that we define as a single submarket on the platform.

To classify all jobs based on their associated skill tag sets and subsequently construct submarkets, we first implement a multi-step clustering procedure to aggregate similar skill sets into distinct groups. Each job is then assigned to a specific group based on its associated skill set. Each resulting group represents a collection of jobs that require similar skill sets, thereby reflecting a common type of labor demand and forming the basis of a submarket. The relationships among jobs, skill sets, skill tag clusters, and submarkets are illustrated in Figure C1.

After clustering, we identify 1,082 distinct submarkets from 409,119 unique skill sets, which serve as the units of our market-level analysis. Since each job is associated with a specific skill set, and each skill set is mapped to a single cluster, we are able to assign every job to a corresponding submarket. This mapping provides us with a complete classification of jobs into submarkets. Using these submarkets, we proceed to define treatment and control groups at both the market and freelancer levels. Specifically, we leverage the LM-AIOE, a variant of the AI Occupational Exposure (AIOE) Index developed by Felten et al. (2023, 2021), which quantifies the degree to which various occupations are exposed to AI tools based on large language models. We begin by calculating the semantic similarity between each of the platform's 2,719 skill tags and the 774 occupations for which LM-AIOE scores are available. Each skill tag is then assigned an exposure score (LM-AIOE value) based on the most semantically aligned occupation. Next, we compute a weighted

average LM-AIOE score for each submarket, where the weights reflect the distribution of skill tags within the submarket. We then adopt a conservative method by setting a threshold to define a binary treatment variable, classifying submarkets into treated and control groups. The details of this process are summarized in the pseudocode presented in Algorithm 1, and the complete procedure for the multi-step clustering and the construction of treatment and control submarkets is outlined below:

1. **Skill Set Embedding:** We use the Sentence Transformer model (SBERT)⁷ to generate vector representations for each skill tag set ST .

2. **Clustering of Skill Tag Sets Based on Embedding Representations:** We cluster the skill tag sets ST based on their embedding representations using the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm.

3. **LM-AIOE Value Assignment for Skill Tags:** We leverage the 774 occupations OT in the LM-AIOE data, each associated with a specific LM-AIOE score, to assign exposure values to skill tags. Specifically, we compute the cosine similarity between each of the platform's 2,719 skill tags ST and each occupation. Then each skill tag is matched to the most semantically similar occupation, and its corresponding LM-AIOE value is assigned as the LM-AIOE value of this skill tag.

4. **Submarket LM-AIOE Values Calculation:** We use the LM-AIOE values of skill tags to compute the weighted LM-AIOE score for each submarket. Specifically, each submarket consists of multiple skill sets, and each skill set is composed of multiple skill tags. Within each submarket, skill tags occur with different frequencies. We define the frequency of a skill tag as $\frac{n_{\text{tag } ST}}{\sum n_{\text{all tags}}}$ and use this value as its weight. The submarket's LM-AIOE score is then calculated as the weighted average of the LM-AIOE values of its skill tags.

5. **Submarket Classification Based on a Defined LM-AIOE Threshold:** We define a binary treatment variable $Treat$ for each submarket based on a predefined LM-AIOE threshold TH . A submarket is assigned $Treat = 1$ if its LM-AIOE value exceeds the threshold TH , and $Treat = 0$ otherwise.

⁷ The Sentence Transformer model (SBERT) is a state-of-the-art neural architecture that generates semantically meaningful sentence or text embeddings (Reimers and Gurevych 2019).

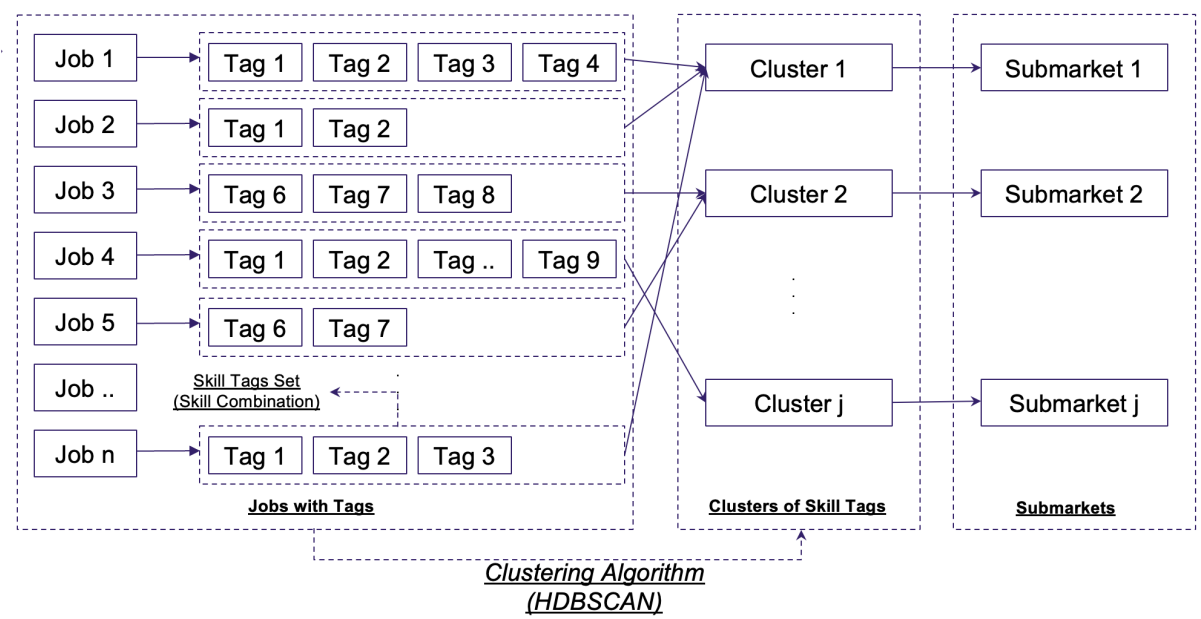


Figure C1 The Relationship between Jobs, Skill Sets, Skill Tag Clusters, and Submarkets

Algorithm 1 Submarket Construction and LM-AIOE-Based Treatment Definition

Require: Skill Tag Sets ST , LM-AIOE Dataset Occupations OT , Skill Tags List**1. Skill Set Embedding Generation****for** each skill tag set $st \in ST$ **do** Compute embedding E_{st} using Sentence Transformer (SBERT)⁸.**end for****2. Skill Tag Set Clustering**Cluster all embedding vectors E_{st} using HDBSCAN algorithm.

Obtain clusters as distinct submarkets.

3. Assign LM-AIOE Values to Skill Tags**for** each skill tag $t \in$ Skill Tags List **do** **for** each occupation $ot \in OT$ **do** Compute cosine similarity $sim(t, ot)$ **end for** Assign the LM-AIOE value of the most similar occupation to tag t **end for****4. Calculate Submarket LM-AIOE Values****for** each submarket s **do** Initialize submarket LM-AIOE score $LM-AIOE_s$ Count frequency of each skill tag ST_s within submarket s **for** each skill tag $t \in ST_s$ **do** Calculate tag frequency weight $w_t = \frac{n_t}{\sum_{t' \in ST_s} n_{t'}}$ Update submarket LM-AIOE: $LM-AIOE_s += w_t \times LM-AIOE_t$ **end for****end for**

Algorithm 1 Submarket Construction and LM-AIOE-Based Treatment Definition (Continued)

5. Define Treatment Variable for Submarkets**for** each submarket s **do** **if** $LM-AIOE_s > TH$ **then** Set $Treat_s = 1$ **else** Set $Treat_s = 0$ **end if****end for**

Appendix D: Variable Summary Statistics

This section reports summary statistics for the primary variables of interest in Table D1. Because each submarket appears on the platform at different times, the number of observations varies by submarket. We define a submarket’s inception as the month of its first job posting; for example, if the first posting occurs in January 2022, we set January 2022 as the submarket’s start date. In each subsequent month, we record the count of job postings, assigning zero to months with no activity. Prior to January 2022, there are no observations for that submarket. As for freelancers, we treat their registration time as the moment they first enter the market.

It is notable that the variables exhibit substantial positive skewness. After applying the natural logarithm, the means and standard deviations become comparable in scale, rendering conventional count models unsuitable. The same pattern arises in the freelancer-level data: while some active freelancers submit bids to hundreds of jobs, many place only a few bids due to the cold-start problem. Accordingly, we apply log transformations to all variables except *Programming%*, *Rating*, and *Developed*, whose limited ranges obviate the need for transformation. It is noteworthy that 301 occupations in our dataset are associated with at least one skill tag, and that the median LM-AIOE score for these occupations is 0.4. Using this value as a threshold, approximately 85% of submarkets qualify as “treated”. This result accords with our theoretical framework and empirical observations: a substantial share of online labor-market tasks produce text or code as their primary outputs and are therefore highly related to the capabilities of LLMs.

Additionally, several data imbalances warrant attention. For the freelancer-month dataset, both *Programming%* and $\ln(\text{Avg_Bids_job} + 1)$ are defined as ratios with *Bids* in the denominator, and *Bids* may be zero. Moreover, *Rating* values are missing for freelancers who have not yet received client reviews; in our DDD analyses, we impute these missing ratings as zero. Finally, *Tenure* has a small number of missing values, and these observations are omitted from the DDD analyses. For those with a missing value in *Tenure*, we regard their first bidding time as the beginning of the panel data.

Table D1 Variable Summary Statistics

Variable	Observation	Mean	Std	Min	Max
Submarket-Month-Level					
$\ln(\text{Jobs} + 1)$	25,833	2.949	1.244	0	6.926
$\ln(\text{Bids} + 1)$	25,833	5.353	1.858	0	10.636
$\ln(\text{Transaction} + 1)$	25,833	5.524	3.019	0	12.386
$\ln(\text{Avg_Bids_job} + 1)$	25,833	2.132	0.999	0	5.118
Freelancer-Month-Level					
<i>Programming%</i>	711,738	0.320	0.341	0	1
$\ln(\text{Bids} + 1)$	1,705,081	0.663	1.107	0	7.840
$\ln(\text{Transaction} + 1)$	1,705,081	0.192	0.996	0	11.707
$\ln(\text{Avg_Earn_Bid} + 1)$	711,738	0.213	0.828	0	11.707
Freelancer-Level					
<i>Treat</i>	132,260	0.879	0.275	0	1
<i>Developed</i>	132,260	0.085	0.279	0	1
$\ln(\text{Avg_Earn_Bid_Pre} + 1)$	132,260	0.222	0.773	0	11.014
$\ln(\text{Tenure} + 1)$	130,686	5.954	1.516	0.003	8.499
<i>Rating</i>	34,325	4.816	0.695	0	5
Submarket-Level					
<i>Treat</i>	1,082	0.848	0.360	0	1
<i>Programming</i>	1,082	0.390	0.488	0	1

Appendix E: Notes on the Difference-in-Differences-in-Differences Specification

In Section 5.2, we employ a triple-difference (DDD) specification that includes interactions among *Treat*, *Post*, and *Programming*. However, upon incorporating two-way fixed effects, we are left with only two coefficients to examine, as shown in Equation 2. This phenomenon arises because, in our analysis, any submarket categorized under *Programming* = 1 is simultaneously considered a treated submarket (*Treat* = 1), due to its high relatedness to large language models and correspondingly elevated LM-AIOE. In other words, programming-intensive submarkets are a subset of the treated submarkets. Consequently, the interaction term *Treat* * *Programming* is equivalent to *Programming* and is also absorbed by the unit fixed effect. Similarly, *Programming* * *Post* is indistinguishable from *Treat* * *Programming* * *Post*, leading to the omission of one variable. After omitting these two interaction terms, Equation 2 is formulated, wherein *Treat* * *Post* captures the post-ChatGPT changes in non-programming-intensive treated submarkets relative to the control submarkets. Meanwhile, *Treat* * *Programming* * *Post* captures additional differences in programming-intensive submarkets compared to non-programming-intensive submarkets, highlighting the heterogeneity of interest.

Appendix F: Interrupted Time Series Analysis of Freelancer-Level Skill Transition

As a supplement to Section 5.2, we further examine skill-transition effects at the freelancer level using an Interrupted Time Series Analysis (ITSA). We estimate the ITSA model using the specification described below, with the results presented in Table F1.

$$Programming\%_{it} = \beta_0 + \beta_1 * Time_t + \beta_2 * Post_t + \beta_3 * PostTime_t + u_i + \epsilon_{it} \quad (F1)$$

Here, $Programming\%_{it}$ denotes the average programming intensity of bids submitted by freelancer i in month t . $Time_t$ represents the number of months elapsed since the start of the observation period. $Post_t$ is a binary indicator equal to 0 during the pre-treatment period and 1 during the post-treatment period. $PostTime_t$ equals 0 before the treatment and increases by one in each subsequent post-treatment month. We also include u_i to control for freelancer-level fixed effects, while the error term ϵ_{it} captures unobserved factors that may influence the outcome. To ensure that the skill-transition process is observable, we restrict the sample to freelancers who submitted at least one bid in both the pre- and post-treatment periods. We find a statistically significant post-treatment increase in $Programming\%$ ($\beta_2 = 0.0500$, $p < 0.01$), indicating that freelancers indeed transitioned toward programming jobs following the introduction of ChatGPT.

Table F1 Results for Interrupted Time Series Analysis

	<i>Programming%</i>
<i>Time</i>	0.0010*** (0.0001)
<i>Post</i>	0.0500*** (0.0009)
<i>PostTime</i>	-0.0001 (0.0003)
Freelancer FE	Yes
Observations	432,558

Note: Robust standard errors are reported at the freelancer level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix G: Freelancer-Level Heterogeneity in Skill Transition: DDD and Event-Study Evidence

To complement the analysis in Section 5.3, we investigate heterogeneity in freelancers' skill-transition behavior by estimating a difference-in-differences-in-differences (DDD) model. The specification is as follows, with results reported in Table G1:

$$\begin{aligned} \text{Programming}\%_{it} = & \beta_0 + \beta_1 * \text{Treat}_i * \text{Post}_t * \text{Moderator}_i + \beta_2 * \text{Treat}_i * \text{Post}_t \\ & + \beta_3 * \text{Post}_t * \text{Moderator}_i + u_i + T_t + \epsilon_{it} \end{aligned} \quad (\text{G1})$$

In this specification, Treat_i is a continuous variable ranging from 0 to 1, representing the proportion of each freelancer's pre-treatment bids submitted in treated submarkets. This variable captures the degree of initial skill relatedness to ChatGPT's capabilities. Moderator_i reflects freelancer characteristics, operationalized in four ways across the columns: freelancer rating (Column (1)), the logarithm of pre-treatment earning efficiency (Column (2)), a binary indicator for whether the freelancer is located in a developed country (Column (3)), and the logarithm of platform tenure (Column (4)).

The estimates in Columns (1) and (2) are consistent with those from the main analysis: higher-skilled freelancers (proxied by higher ratings or by greater earning efficiency) are significantly more likely to shift into programming work ($\beta_1 = 0.0256$ and 0.0328 , respectively; $p < 0.01$). Column (3) reveals no statistically significant difference in transition behavior between freelancers in developed and developing countries ($\beta_1 = 0.0040$; $p > 0.10$). Finally, Column (4) indicates that freelancers with longer platform tenure are more inclined to move into programming-intensive jobs ($\beta_1 = 0.027$; $p < 0.01$).

Furthermore, we estimate an event-study (relative-time) specification to test the parallel trends assumption. As shown in Figure G1, the pre-treatment coefficients cluster around zero and are statistically insignificant, indicating no violation of the parallel trends assumption.

$$\begin{aligned} \text{Outcome}_{it} &= \sum_{k \neq -1} \beta_k D_{i,t=k} + u_i + T_t + \epsilon_{it} \\ &= \sum_{k \neq -1} \beta_k (D_i \cdot \mathbf{1}\{t - \tau_i = k\}) + u_i + T_t + \epsilon_{it}, \end{aligned} \quad (\text{G2})$$

Table G1 DDD Analyses about Skill-Transition Effects

	<i>Programming%</i>			
	(1)	(2)	(3)	(4)
<i>Treat * Post * Moderator</i>	0.0256*** (0.0006)	0.0328*** (0.0013)	0.0040 (0.0045)	0.0271*** (0.0010)
<i>Treat * Post</i>	-0.0366*** (0.0022)	0.0084*** (0.0017)	0.0226*** (0.0016)	-0.153*** (0.0068)
<i>Post * Moderator</i>	-0.0165*** (0.0005)	-0.0212*** (0.0012)	-0.0193*** (0.0043)	-0.0155*** (0.0009)
Constant	0.328*** (0.0008)	0.316*** (0.0006)	0.313*** (0.0006)	0.348*** (0.0024)
Moderator	<i>Rating</i>	<i>ln(Avg_Earn_Bid_Pre + 1)</i>	<i>Developed</i>	<i>ln(Tenure + 1)</i>
Freelancer FE	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes
Observations	711,738	711,738	711,738	697,908
R^2	0.858	0.858	0.858	0.857

Note: Cluster-robust standard errors are reported at the freelancer level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

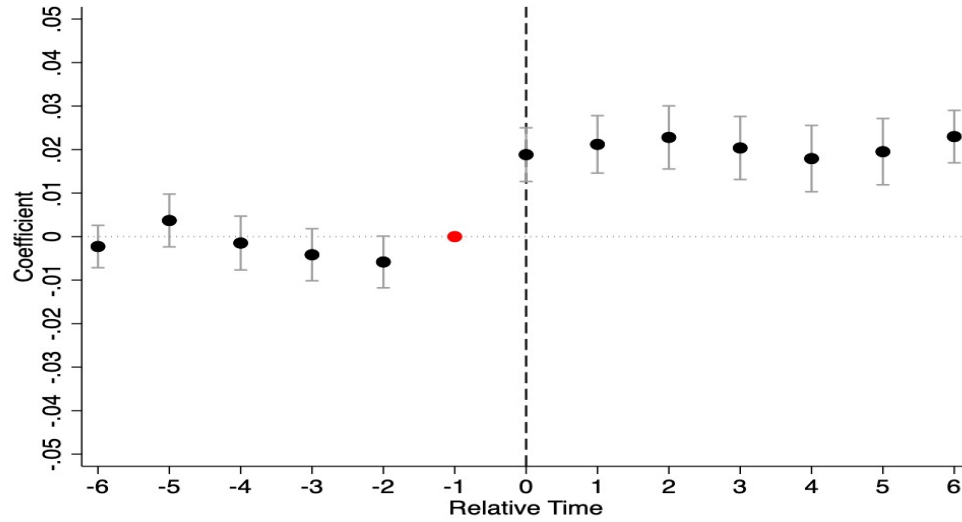


Figure G1 Pre-Trends and Temporal Dynamics of Freelancers' Programming Proportion

Appendix H: Additional Details on Freelancer-Level Earnings

Similar to our market-level analysis, we examine the effect of ChatGPT on freelancer-level earnings. Using the specification in Equation 1, we estimate the impact on freelancers' total transaction value and the mechanisms driving the changes, including changes in bidding behavior and earning efficiency.

Since freelancers who did not receive any payment during their pre-treatment period had no potential for income loss, we first focus on freelancers who were awarded at least one job and had prior earning activity, as shown in Table H1. We examine how the magnitude of the change in transactions, which varies across freelancers, is affected by ChatGPT. Here, $Treat$ is a continuous variable ranging from 0 to 1, representing the proportion of bids submitted to treated submarkets. A higher value of $Treat$ indicates that the freelancer is more likely to be affected by ChatGPT's demand displacement effect. We find a significant decline in transaction value ($\beta_1 = -0.116, p < 0.05$), which is primarily driven by a substantial reduction in the number of bids submitted ($\beta_1 = -0.0870, p < 0.01$) and a small, statistically insignificant decrease in earning efficiency ($\beta_1 = -0.0041, p > 0.1$).

In Tables H2 and H3, we separately analyze freelancers who made a skill transition and those who did not. We define skill-transitioning freelancers as those whose *Programming%* increases after the introduction of ChatGPT. Freelancers who transitioned exhibit no significant decline in transaction value; their bidding activity remains stable, and their earning efficiency significantly improves. In contrast, those who did not transition experience substantial reductions in both transaction value and bid volume, along with a significant decline in earning efficiency.

For completeness, Tables H4, H5, and H6 replicate these analyses including all freelancers (even those with zero pre-treatment earnings). Unsurprisingly, the estimated effects are attenuated because most freelancers report zero earnings before treatment, which precludes substantial further reduction, but the qualitative patterns are consistent.

Table H1 Impact on Freelancers' Earnings

	$\ln(y + 1)$		
	<i>Transaction</i>	<i>Bids</i>	<i>Avg_Earn_Bid</i>
	(1)	(2)	(3)
<i>Treat * Post</i>	-0.116** (0.0467)	-0.0870*** (0.0265)	-0.0041 (0.0351)
<i>Constant</i>	1.005*** (0.0152)	1.727*** (0.0086)	0.629*** (0.011)
Freelancer FE	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes
Observations	310,771	310,771	218,716
R^2	0.279	0.670	0.244

Note: Cluster-robust standard errors are reported at the freelancer level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table H2 Impact on Freelancers' Earnings (Transition)

	$\ln(y + 1)$		
	<i>Transaction</i>	<i>Bids</i>	<i>Avg_Earn_Bid</i>
	(1)	(2)	(3)
<i>Treat * Post</i>	0.0695 (0.0523)	0.0483 (0.0322)	0.118*** (0.0452)
<i>Constant</i>	1.055*** (0.0175)	1.888*** (0.0108)	0.593*** (0.0149)
Freelancer FE	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes
Observations	217,637	217,637	160,571
R^2	0.290	0.670	0.246

Note: Cluster-robust standard errors are reported at the freelancer level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table H3 Impact on Freelancers' Earnings (Non-Transition)

	$\ln(y + 1)$		
	<i>Transaction</i>	<i>Bids</i>	<i>Avg_Earn_Bid</i>
	(1)	(2)	(3)
<i>Treat * Post</i>	-0.208*** (0.0540)	-0.146*** (0.0340)	-0.156*** (0.0496)
<i>Constant</i>	0.762*** (0.0157)	1.212*** (0.0099)	0.683*** (0.0131)
Freelancer FE	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes
Observations	98,504	98,504	60,573
R^2	0.230	0.609	0.265

Note: Cluster-robust standard errors are reported at the freelancer level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table H4 Impact on Freelancers' Earnings, Including Those with Zero Pre-Treatment Earnings

	$\ln(y + 1)$		
	<i>Transaction</i>	<i>Bids</i>	<i>Avg_Earn_Bid</i>
	(1)	(2)	(3)
<i>Treat * Post</i>	-0.0228*** (0.0072)	-0.0425*** (0.0057)	-0.0140 (0.0094)
Constant	0.200*** (0.0025)	0.678*** (0.0020)	0.218*** (0.0033)
Freelancer FE	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes
Observations	1,705,081	1,705,081	711,738
R^2	0.361	0.671	0.322

Note: Cluster-robust standard errors are reported at the freelancer level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table H5 Impact on Freelancers' Earnings, Including Those with Zero Pre-Treatment Earnings (Transition)

	$\ln(y + 1)$		
	<i>Transaction</i>	<i>Bids</i>	<i>Avg_Earn_Bid</i>
	(1)	(2)	(3)
<i>Treat * Post</i>	0.0109 (0.0087)	0.0229*** (0.0073)	0.0084 (0.0121)
<i>Constant</i>	0.268*** (0.0030)	0.823*** (0.0025)	0.254*** (0.0043)
Freelancer FE	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes
Observations	932,942	932,942	430,712
R^2	0.376	0.699	0.318

Note: Cluster-robust standard errors are reported at the freelancer level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table H6 Impact on Freelancers' Earnings, Including Those with Zero Pre-Treatment Earnings
(Non-Transition)

	$\ln(y + 1)$		
	<i>Transaction</i>	<i>Bids</i>	<i>Avg_Earn_Bid</i>
	(1)	(2)	(3)
<i>Treat * Post</i>	-0.0117 (0.0077)	0.0033 (0.0067)	-0.0179 (0.0118)
<i>Constant</i>	0.098*** (0.0026)	0.449*** (0.0023)	0.154*** (0.0039)
Freelancer FE	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes
Observations	814,668	814,668	294,952
R^2	0.291	0.539	0.341

Note: Cluster-robust standard errors are reported at the freelancer level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix I: The Heterogeneous Effects of Generative AI Across Programming Domains

In our market-level analysis, we find that although programming-intensive and non-programming-intensive submarkets exhibit no consistent differences in labor demand, the contraction in labor supply is significantly smaller in programming-intensive submarkets. We aim to take one step further in exploring whether any heterogeneity exists among programming jobs. Large language models support numerous programming languages since they have been trained on multi-language code corpora (Nguyen et al. 2023, Kocetkov et al. 2023) and evaluated on benchmarks spanning a wide range of programming languages (Cassano et al. 2023). Broadly, programming languages can be categorized into two distinct types: scripting languages and compiled languages (Prechelt 2000, Ousterhout 1998). These categories differ in syntactic complexity and learning cost: scripting languages typically feature simpler syntax and lower learning barriers, while compiled languages are generally more complex and demanding (Lokkila et al. 2023), potentially influencing how different languages mediate the impact of generative AI.

In light of this, we further divide the programming-intensive submarkets into two segments based on the predominant type of programming language, in order to examine the heterogeneous impact on scripting-language-dominated versus compiled-language-dominated submarkets. Specifically, we classify each submarket according to the relative proportions of scripting and compiled languages used. If the proportion of scripting languages exceeds that of compiled languages within a submarket, it is categorized as scripting-language-dominated; otherwise, it is considered compiled-language-dominated. A small number of submarkets, however, do not predominantly feature either language type and instead rely on alternative paradigms, such as the declarative language SQL. As there are only four such submarkets, we exclude them from the analysis.

Subsequently, following the model specifications in Section 4, we separately compare the demand, supply, transaction, and competition in scripting-language-dominated and compiled-language-dominated submarkets with those in the control submarkets. The estimation results are reported in Table I1. Panel A compares scripting-language-dominated submarkets with the control group, while Panel B compares compiled-language-dominated submarkets with the control group. Consistent with the main results, both types of submarkets show a notable decline in labor demand ($\beta_1 = -0.188$ or -0.234 , $p < 0.01$) and total transaction volume ($\beta_1 = -0.283$, $p < 0.01$ or $\beta_1 = -0.268$, $p < 0.05$) relative to the control group. However, it is

worth noting that on the supply side, measured by the total number of bids, scripting-language-dominated submarkets do not exhibit a significant decline relative to the control group ($\beta_1 = -0.064$, $p > 0.1$). In contrast, compiled-language-dominated submarkets still experience a significant reduction in labor supply ($\beta_1 = -0.131$, $p < 0.05$). This result may reflect the lower barrier to entry associated with scripting languages compared to compiled languages, which generally require more formal training. Such suggestive evidence indicates that the skill-transition effect is inherently constrained: freelancers cannot move freely into every role without encountering substantive limitations.

Table I1 The Heterogeneous Effects of Generative AI Across Different Types of Programming Submarkets

Panel A				
Programming Type: Scripting Language vs. Control				
$\ln(y + 1)$				
	<i>Jobs</i>	<i>Bids</i>	<i>Transaction</i>	<i>Avg_Bids_job</i>
	(1)	(2)	(3)	(4)
<i>Treat * Post</i>	-0.188*** (0.0348)	-0.0640 (0.0524)	-0.283*** (0.105)	0.213*** (0.0408)
Constant	2.863*** (0.0082)	5.199*** (0.0123)	5.467*** (0.0246)	2.095*** (0.0096)
Submarket FE	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes
Observations	10,268	10,268	10,268	10,268
R^2	0.895	0.848	0.636	0.544
Panel B				
Programming Type: Compiled Language vs. Control				
$\ln(y + 1)$				
	<i>Jobs</i>	<i>Bids</i>	<i>Transaction</i>	<i>Avg_Bids_job</i>
	(1)	(2)	(3)	(4)
<i>Treat * Post</i>	-0.234*** (0.0371)	-0.131** (0.0560)	-0.268** (0.103)	0.181*** (0.0421)
Constant	2.759*** (0.0067)	4.887*** (0.0102)	5.151*** (0.0187)	1.933*** (0.0076)
Submarket FE	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes
Observations	7,496	7,496	7,496	7,496
R^2	0.898	0.847	0.645	0.524

Note: Cluster-robust standard errors are reported at the submarket level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix J: The Impact of Generative AI on Market Entry Decisions of New Freelancers

Given that some freelancers can be active in online labor markets for an extended period and work on diverse jobs, several potential mechanisms may explain the average and heterogeneous impacts on the labor supply side, as documented in Sections 5.1 and 5.2. Utilizing the market-level data, we explore the influx of new freelancers as one such mechanism. To this end, we introduce a new outcome variable, *New_Freelancer*, defined as the monthly count of freelancers who newly register and submit at least one bid within specific submarkets. We employ the same econometric specifications to perform DiD and DDD analyses, with the results presented in Table J1. The estimates suggest that the decline in new participants is one contributing mechanism to the observed decrease in average labor supply across treated submarkets ($\beta_1 = -0.145$, $p < 0.01$). However, there is no significant heterogeneity in the number of new freelancers between programming-intensive and non-programming-intensive submarkets ($\beta_2 = -0.0430$, $p > 0.1$), effectively ruling out differences in new participation as the driving mechanism for heterogeneous effects on labor supply. We also test the parallel trends assumption and examine whether any differential trends existed prior to the introduction of ChatGPT. As shown in Figure J1, the confidence intervals for all pre-treatment periods encompass zero, suggesting an absence of systematic differences before the intervention.

Table J1 The Impact of Generative AI on Market Entry Decisions of New Freelancers

	$\ln(New_Freelancer + 1)$	
	(1)	(2)
$Treat * Post$	-0.145*** (0.0371)	-0.125*** (0.0411)
$Treat * Post * Programming$		-0.0430 (0.0283)
Constant	3.591*** (0.0119)	3.591*** (0.0119)
Submarket FE	Yes	Yes
Year-Month FE	Yes	Yes
Observations	25,833	25,833
R^2	0.900	0.901

Note: Cluster-robust standard errors are reported at the submarket level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

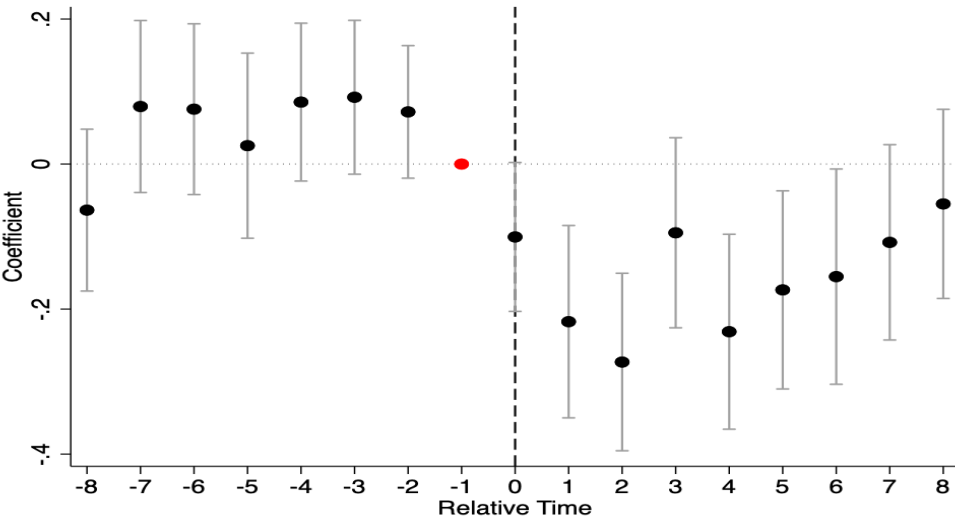


Figure J1 Pre-Trends and Temporal Dynamics of New Freelancers

Appendix K: The Impacts on Project Budgets

To better understand the underlying motivations behind freelancers' transitions, we further examine how the introduction of ChatGPT affects job budgets. As demonstrated in Section 5.2, freelancers shift toward bidding on programming-intensive jobs. While these jobs have traditionally commanded higher wages than non-programming-intensive ones, ChatGPT may have further widened this wage gap. This widening gap likely strengthens freelancers' motivation to transition into programming-intensive submarkets as a way to mitigate ChatGPT's negative impact. Therefore, we examine the wages clients are willing to offer for completing a given task, as reflected in job budgets, and analyze how these budgets have changed differentially between programming-intensive and non-programming-intensive submarkets following the introduction of ChatGPT.

Specifically, we define a new dependent variable, *Budget*, as the average project budget of all jobs posted within a submarket during a given month. We then follow the DDD model specification outlined in Section 5.2 to investigate the heterogeneous treatment effects (HTE) on job budgets. The corresponding estimation results are presented in Table K1. The negative and statistically insignificant coefficient of the interaction term $Treat * Post * Programming$ suggests that the wage gap was not widened by the introduction of ChatGPT, which allows us to rule out this explanation.

Table K1 The Impact of Generative AI on Project Budgets		
	$\ln(Budget + 1)$	
	(1)	(2)
$Treat * Post$	0.0575* (0.0327)	0.0727** (0.0349)
$Treat * Post * Programming$		-0.0327 (0.0223)
Constant	5.669*** (0.0105)	5.669*** (0.0105)
Submarket FE	Yes	Yes
Year-Month FE	Yes	Yes
Observations	25,086	25,086
R^2	0.381	0.381

Note: Cluster-robust standard errors are reported at the submarket level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix L: Additional Evidence on Demand-Side Market Dynamics

To provide additional insight into market demand dynamics, we visualize the temporal evolution of three key indicators within the treated submarkets: the proportion of programming-intensive jobs, the average number of required skills per job, and the average job budget. Specifically, we calculate the monthly proportion of programming-intensive jobs in the overall market and plot its trend over time. The results are presented in Figure L1, where the horizontal axis represents the number of months relative to the treatment point, and the vertical axis indicates the proportion of programming-intensive jobs. The blue dots show the observed monthly values, while the red line illustrates the fitted trend estimated through linear regression. Similarly, we calculate the monthly average number of required skills per job and average job budget across the entire market and plot their changes over time in Figures L2 and L3.

Based on the results presented in Figure L1, the proportion of programming-intensive jobs remains largely stable over the observation period, exhibiting only a slight decline. The overall variation is minimal, and the fitted trend line is nearly flat. This suggests that, on the demand side, clients have not demonstrated any notable shift in preference or a growing tendency toward programming-intensive tasks. Figure L2 presents the temporal trend in the number of required skills per job. The fitted trend line appears relatively flat with a slight upward slope, indicating a gradual increase in the number of skills requested by clients. Figure L3 illustrates the monthly evolution of the average job budget, revealing a modest upward trend over time, particularly following the introduction of ChatGPT. This pattern suggests a gradual rise in job budgets within the treated submarkets and aligns with the observed increase in skill requirements. We re-estimate the DiD model separately to compare high- and low-budget submarkets with the control group. As reported in Table L1, the results show a higher demand displacement effect in low-budget submarkets.

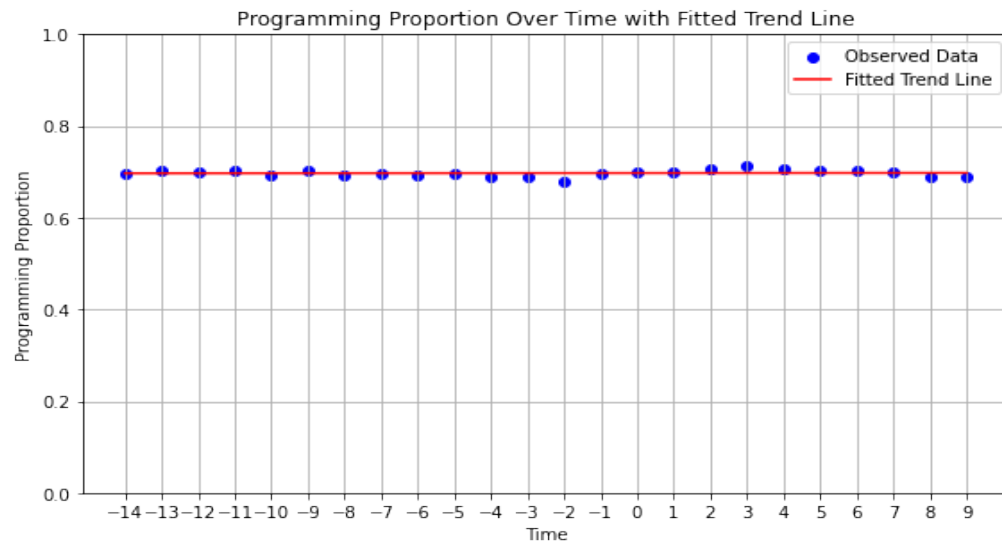


Figure L1 Temporal Trends in Proportion of Programming Jobs

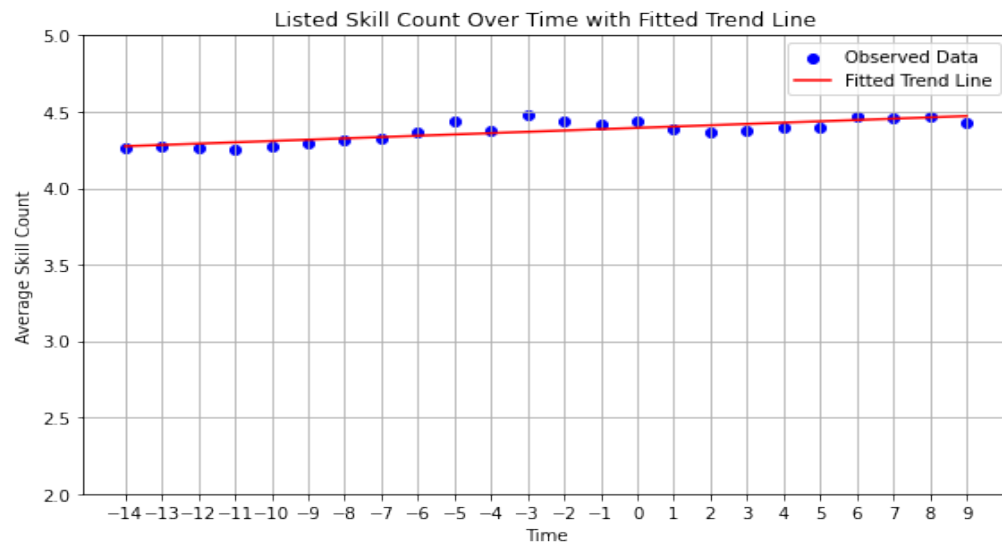


Figure L2 Temporal Trends in Average Skill Count

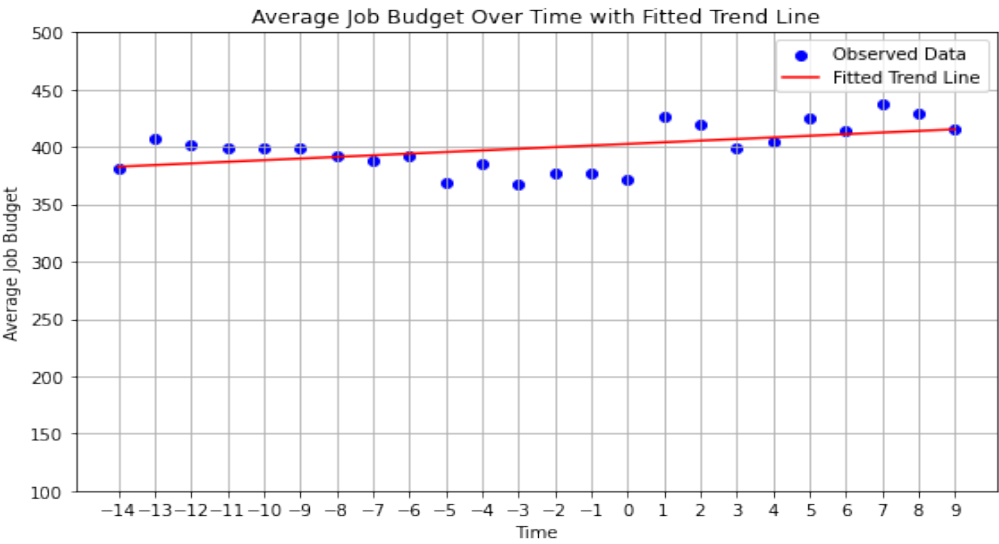


Figure L3 Temporal Trends in Average Budget

Table L1 The Heterogeneous Effects of Generative AI Across Submarkets with Different Budget Levels

Panel A				
High-Budget Submarkets				
	$\ln(y + 1)$			
	<i>Jobs</i>	<i>Bids</i>	<i>Transaction</i>	<i>Avg.Bids.job</i>
	(1)	(2)	(3)	(4)
<i>Treat * Post</i>	-0.185*** (0.0381)	-0.156*** (0.0566)	-0.233** (0.110)	0.0886** (0.0400)
Constant	3.078*** (0.0121)	5.615*** (0.0181)	5.713*** (0.0351)	2.272*** (0.0128)
Submarket FE	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes
Observations	17,421	17,421	17,421	17,421
R ²	0.889	0.836	0.597	0.501
Panel B				
Low-Budget Submarkets				
	$\ln(y + 1)$			
	<i>Jobs</i>	<i>Bids</i>	<i>Transaction</i>	<i>Avg.Bids.job</i>
	(1)	(2)	(3)	(4)
<i>Treat * Post</i>	-0.295*** (0.0574)	-0.192** (0.0873)	-0.489*** (0.154)	0.243*** (0.0713)
Constant	2.897*** (0.0183)	4.972*** (0.0279)	5.443*** (0.0493)	1.707*** (0.0228)
Submarket FE	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes
Observations	8,412	8,412	8,412	8,412
R ²	0.895	0.838	0.643	0.532

Note: Cluster-robust standard errors are reported at the submarket level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix M: The Impacts on Image-Related Jobs

We further investigate how image-related work is affected by the emergence of generative AI. Specifically, we consider two key events: the release of the first publicly available text-to-image AI model, and the subsequent launch of ChatGPT. To analyze their respective impacts on the image-related submarket, we estimate two regression specifications. Consistent with our earlier submarket construction procedure, we apply the same clustering approach to the skill sets of image-related jobs to define the corresponding submarkets. In this setup, control submarkets serve as the baseline, while the treatment group comprises the image-related submarkets.

For the first specification, we use $Post_{image,t}$ to denote the first publicly available version as the treatment time, set at July 2022 on a monthly level. For the second specification, we adopt the same $Post$ variable from our main analyses, using the ChatGPT launch date as the start of the treatment period. We define $Treat_{image,i}$ as an indicator variable equal to 1 if unit i corresponds to an image-related submarket and 0 otherwise.

$$Outcome_{it} = \beta_0 + \beta_1 * Treat_{image,i} * Post_{image,t} + u_i + T_t + \epsilon_{it} \quad (M1)$$

$$Outcome_{it} = \beta_0 + \beta_1 * Treat_{image,i} * Post_t + u_i + T_t + \epsilon_{it} \quad (M2)$$

Equation M1 examines changes in demand, supply, and matching outcomes in image-related submarkets relative to control submarkets, in response to the introduction of text-to-image generative AI. This analysis includes data only up to November 2022, thereby excluding the post-ChatGPT period, with the corresponding results reported in Table M1. In the second analysis, we explore whether the widespread attention generated by ChatGPT led to increased interest in text-to-image AI applications, potentially encouraging broader adoption and influencing labor market outcomes. For this purpose, we extend the observation window to include the post-ChatGPT period and estimate relative changes in image-related submarkets. The results of this spillover analysis are presented in Table M2.

Both sets of analyses show insignificant results across all measurements. This suggests that current text-to-image generative AI may not be general-purpose or powerful enough to influence a wide range of jobs in online labor markets.

Table M1 The Effect of Generative AI on the Image-Related Labor Market

	$\ln(y+1)$			
	<i>Jobs</i>	<i>Bids</i>	<i>Transaction</i>	<i>Avg_Bids_job</i>
	(1)	(2)	(3)	(4)
$Treat_{image} * Post_{image}$	-0.0021 (0.0402)	0.0854 (0.0571)	-0.150 (0.139)	0.0233 (0.0600)
Constant	3.047*** (0.0075)	5.460*** (0.0107)	5.168*** (0.0261)	2.192*** (0.0113)
Submarket FE	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes
Observations	5,461	5,461	5,461	5,461
R^2	0.925	0.913	0.660	0.574

Note: Cluster-robust standard errors are reported at the submarket level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table M2 The Spillover Effect of ChatGPT's Success on the Image-Related Labor Market

	$\ln(y+1)$			
	<i>Jobs</i>	<i>Bids</i>	<i>Transaction</i>	<i>Avg_Bids_job</i>
	(1)	(2)	(3)	(4)
$Treat_{image} * Post$	-0.0536 (0.0337)	-0.0619 (0.0470)	-0.0775 (0.0959)	0.0683 (0.0421)
Constant	2.983*** (0.0071)	5.510*** (0.0099)	5.015*** (0.0201)	2.303*** (0.0088)
Submarket FE	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes
Observations	8,792	8,792	8,792	8,792
R^2	0.919	0.904	0.660	0.573

Note: Cluster-robust standard errors are reported at the submarket level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix N: Robustness Check - Alternative Treatment/Control Group Definition Thresholds

Given that our submarket classification is based on a conservative approach using the median value of LM-AIOE to distinguish between submarkets with relatively high and low LM-AIOE levels, we acknowledge that observations near the threshold may introduce sensitivity into the estimation. Varying the threshold can alter submarket classifications, particularly for those with LM-AIOE values close to the cutoff. For example, lowering the threshold may lead to some submarkets that were previously assigned to the control group being reclassified into the treatment group.

To address this concern and enhance the robustness of our findings, we implement an alternative thresholding strategy to ensure that the results are not driven by the specific choice of cutoff. Specifically, we assess the sensitivity of our results by adjusting the original median-based threshold by ± 0.05 units. The lower threshold is defined as the median minus 0.05, while the higher threshold is defined as the median plus 0.05. This approach enables us to assess whether submarkets classified as high or low LM-AIOE continue to exhibit consistent patterns that are not influenced by the classification criteria. The estimation results based on these alternative thresholds are presented in Table N1 and Table N2. Table N1 reports the results using the lower threshold, while Table N2 presents the results using the higher threshold.

Interestingly, we find that although the estimated coefficients exhibit slight fluctuations compared to those reported in Table 2, they remain statistically significant. Specifically, the outcomes for demand, supply, and transaction value consistently show a moderate decline, indicating stable and coherent patterns of heterogeneous treatment effects. These results confirm the robustness and consistency of our findings under alternative threshold specifications.

Table N1 Main Effects Replication with Low Threshold

	$\ln(y+1)$			
	<i>Jobs</i>	<i>Bids</i>	<i>Transaction</i>	<i>Avg_Bids_job</i>
	(1)	(2)	(3)	(4)
<i>Treat * Post</i>	-0.220*** (0.0359)	-0.183*** (0.0535)	-0.332*** (0.101)	0.130*** (0.0397)
<i>Treat * Post * Programming</i>	0.0207 (0.0224)	0.133*** (0.0346)	0.0523 (0.0667)	0.109*** (0.0267)
Constant	3.018*** (0.0109)	5.393*** (0.0162)	5.625*** (0.0311)	2.074*** (0.0121)
Submarket FE	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes
Observations	25,833	25,833	25,833	25,833
R^2	0.890	0.841	0.612	0.541

Note: Cluster-robust standard errors are reported at the submarket level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table N2 Main Effects Replication with High Threshold

	$\ln(y+1)$			
	<i>Jobs</i>	<i>Bids</i>	<i>Transaction</i>	<i>Avg_Bids_job</i>
	(1)	(2)	(3)	(4)
<i>Treat * Post</i>	-0.198*** (0.0339)	-0.195*** (0.0513)	-0.299*** (0.0946)	0.0862** (0.0372)
<i>Treat * Post * Programming</i>	0.0294 (0.0229)	0.149*** (0.0352)	0.0657 (0.0680)	0.112*** (0.0273)
Constant	3.006*** (0.0096)	5.391*** (0.0145)	5.608*** (0.0272)	2.089*** (0.0106)
Submarket FE	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes
Observations	25,833	25,833	25,833	25,833
R^2	0.890	0.841	0.612	0.541

Note: Cluster-robust standard errors are reported at the submarket level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix O: Robustness Check - Pseudo Treatment Timing (Placebo Test)

Since the launch of ChatGPT occurred at the end of the year, holidays and seasonal effects might influence the observed post-shock differences. For instance, these differences could be attributed to an alternative mechanism where treated submarkets naturally experience a more pronounced downward trend around the end of the year compared to control submarkets. To address this concern, we conduct another placebo test using November 30, 2021, the same day in the previous year, as a pseudo-shock date. We maintain the time period from September 2021 to August 2022 and rerun all regressions. As shown in Table O1, our main findings remain robust, suggesting that they are not driven by end-of-year effects or seasonal factors.

Table O1 Placebo Test with a Pseudo Shock Date

	$\ln(y + 1)$			
	<i>Jobs</i>	<i>Bids</i>	<i>Transaction</i>	<i>Avg_Bids_job</i>
	(1)	(2)	(3)	(4)
<i>Treat * Post</i>	-0.0203	0.0229	0.202	0.0706
	(0.0315)	(0.0540)	(0.128)	(0.0476)
Constant	3.120***	5.345***	5.614***	1.906***
	(0.0196)	(0.0336)	(0.0796)	(0.0296)
Submarket FE	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes
Observations	11,769	11,769	11,769	11,769
R^2	0.928	0.873	0.666	0.606

Note: Cluster-robust standard errors are reported at the submarket level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix P: Robustness Check - Treatment Reassignment (Placebo Tests)

Statistically, there is a possibility that the observed effects are driven by random variations across observations. To address this, we conduct a placebo test by randomly reassigning the treatment ($Treat * Post$) within our sample. We simulate this permutation procedure 1,000 times and capture the distribution of the placebo treatment effects for all four main outcome variables based on random treatment assignment. The distribution of placebo treatment effects resembles a normal distribution, as shown in Figure P1, and our estimated coefficient lies far in the tail of this distribution. This placebo test indicates that the treatment effects are not driven by chance, further reinforcing the robustness of our findings.

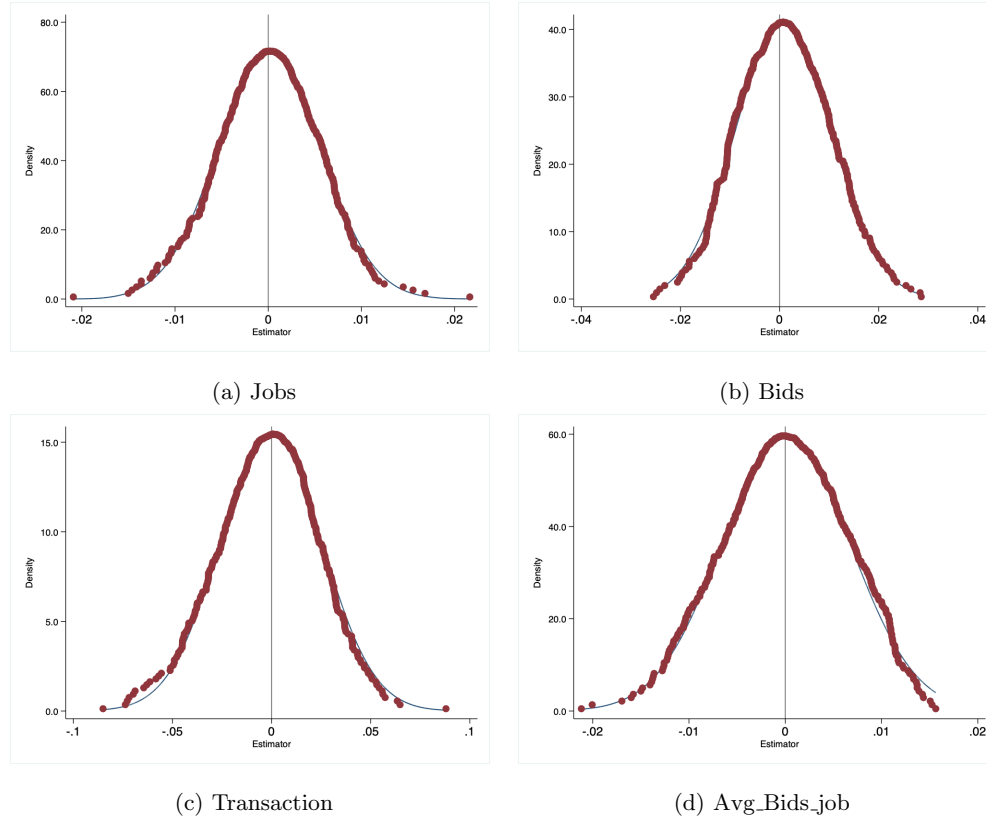


Figure P1 Placebo Tests

Appendix Q: Robustness Check - Pretrends and Temporal Dynamics

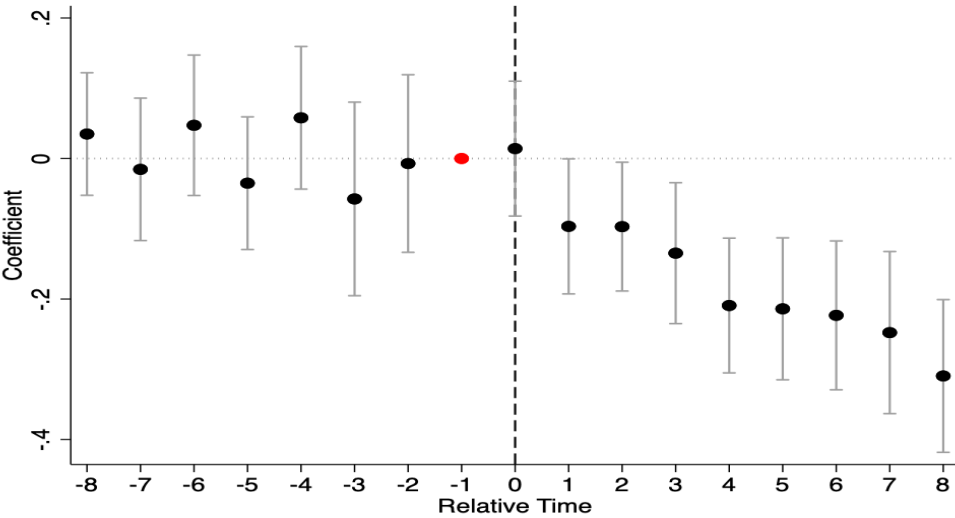
We use a DiD specification for most analyses to uncover how generative AI introduces different trends for treated submarkets versus control submarkets. To attribute post-shock differences to the shock itself, we first need to examine whether these two sets of submarkets follow similar trends before the shock (parallel trends assumption). Additionally, we are interested in whether the observed effects appear primarily at the beginning of the shock (due to novelty effects, etc.) and attenuate over time, or whether these differences strengthen over time. With these objectives in mind, we employ the relative time model to obtain per-period effect estimates, following previous scholarly investigations (Braghieri et al. 2022, Perez-Truglia 2020).

$$\begin{aligned} Outcome_{it} &= \sum_{k \neq -1} \beta_k D_{i,t=k} + u_i + T_t + \epsilon_{it} \\ &= \sum_{k \neq -1} \beta_k (D_i \cdot \mathbf{1}\{t - \tau_i = k\}) + u_i + T_t + \epsilon_{it}, \end{aligned} \quad (Q1)$$

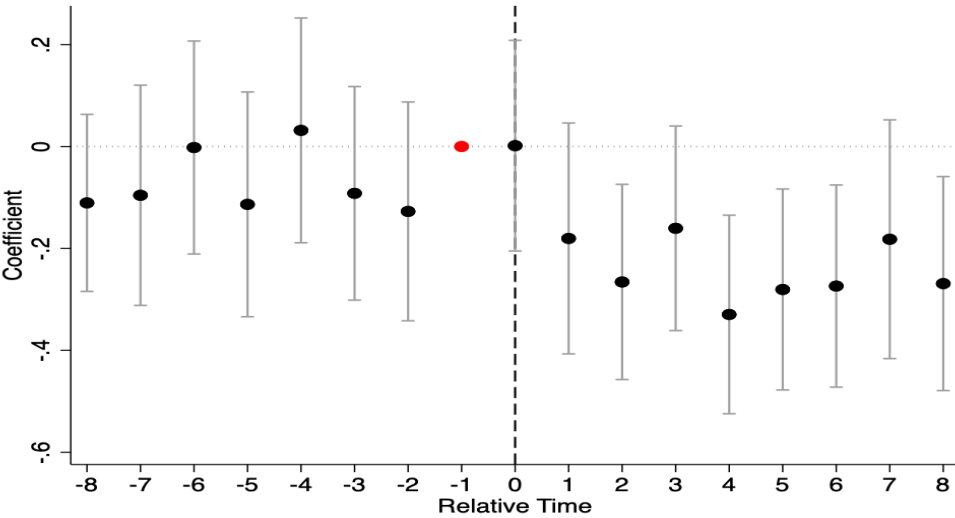
where $D_{i,t=k}$ equals 1 if unit i is treated and period t is k periods relative to unit i 's event (adoption) time τ_i , and 0 otherwise; the reference period $k = -1$ is omitted. Equivalently, $D_{i,t=k}$ can be written as $D_i \cdot \mathbf{1}\{t - \tau_i = k\}$, where D_i indicates whether unit i is from the treatment group and $\mathbf{1}\{\cdot\}$ is the indicator function. u_i and T_t denote unit and time fixed effects, respectively, and ϵ_{it} is the idiosyncratic error term.

To estimate the relative time model, we utilize the full panel of data. Following common practice in event-study analyses (Schmidheiny and Siegloch 2023), we merge all periods prior to event time $k = -8$ into a single bin at $k = -8$, and $k = +8$ is the final time period. This binning approach reduces noise and ensures a more balanced panel. Importantly, our main findings remain robust without this binning: even when all relative time periods are included without aggregation, we continue to find no evidence of differential pre-trends.

The results are shown in Figures Q1 and Q2. First, the confidence intervals for the periods leading up to the launch of ChatGPT all include zero, indicating no significant pre-existing trend differences. This finding strengthens the credibility of attributing the observed post-shock differences to the launch of LLM-based generative AI. Second, the figures reveal an increasing impact over time in the post-treatment periods, suggesting that the influence of these technologies is not a one-time event but deepens with increasing adoption.

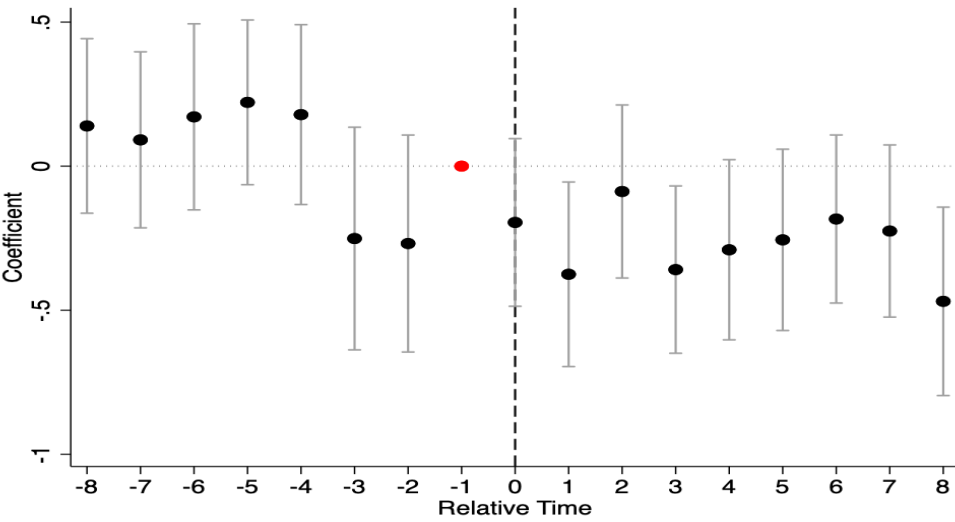


(a) Jobs

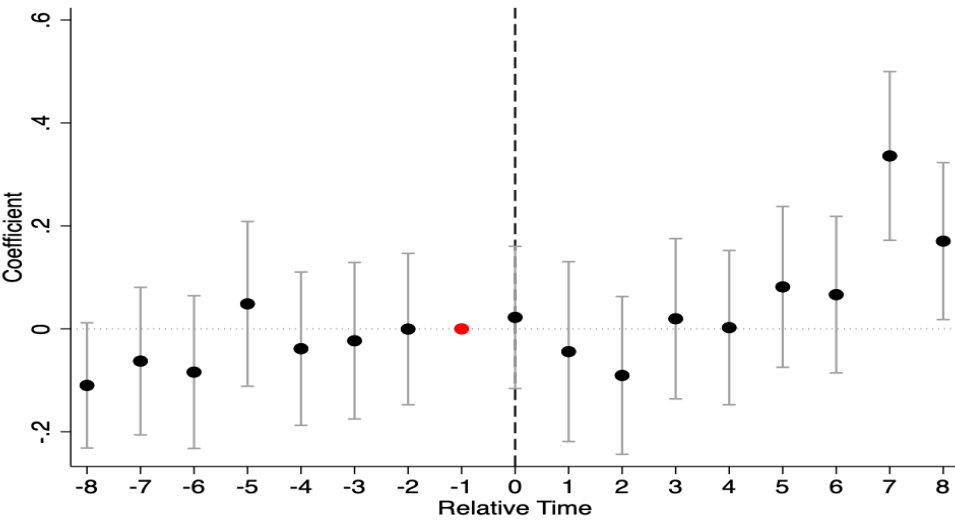


(b) Bids

Figure Q1 Pretrends and Temporal Dynamics in Demand and Supply



(a) Transaction



(b) Avg_Bids_job

Figure Q2 Pretrends and Temporal Dynamics in Matching Outcomes

Appendix R: Robustness Check - Propensity Score Matching

Although the treated and control submarkets follow similar pre-shock trends, as demonstrated in Appendix Q, they may still differ across various dimensions such as market size prior to the treatment. To address this concern and enhance the comparability between the two sets of submarkets, we re-estimate our analysis using a propensity-score-matched sample of treated and control submarkets. Specifically, we estimate a probit selection model across submarkets to predict each submarket's probability of receiving treatment. To improve the comparability of the matched submarkets, we include all relevant pre-treatment variables (*Jobs*, *Clients*, *Bids*, *Bidders*, *Matches*, *Transaction*, and *Avg_Bids_job*) as covariates in the matching process. Based on the predicted propensity scores, we apply a nearest-neighbor matching procedure following standard practices in the literature (Ran et al. 2025, Ye and Shankar 2025, Faccio and O'Brien 2021) to identify the closest control submarket for each treated submarket. The resulting matched sample consists of 952 submarkets.

Next, we evaluate the quality of the matches by comparing the treated and control submarkets in terms of the covariates used in the matching procedure. Table R1 presents the mean values of the matching variables. In contrast to the full sample results shown in the Before-Matching columns, the After-Matching columns reveal no statistically significant differences between the treated and control submarkets. This indicates that the propensity score matching procedure effectively mitigates key pre-treatment differences between the two groups. Additionally, Figures R1 and R2 illustrate the distribution of propensity scores before and after matching, demonstrating that the matched sample is substantially more balanced.

In Table R2, we re-estimate our main regression model using the matched submarket sample. Following the specification outlined in Section 4, the coefficient on the interaction term $Treat * Post$ (β_1) remains negative and statistically significant. These results are consistent with our baseline findings. Overall, the robustness checks suggest that our main results are not driven by insufficient matching quality.

Table R1 Statistical Validation of Pre- and Post-Matching Balance

Variables	Before-Matching			After-Matching		
	Control	Treat	<i>p</i> -value	Control	Treat	<i>p</i> -value
<i>Jobs</i>	24.38	47.49	<0.001	28.65	35.68	0.112
<i>Clients</i>	21.30	41.17	<0.001	25.03	31.04	0.110
<i>Bids</i>	406.14	774.44	0.002	490.80	591.41	0.349
<i>Bidders</i>	234.65	443.35	0.001	281.71	363.47	0.189
<i>Matches</i>	3.28	6.35	<0.001	3.85	4.71	0.199
<i>Transaction</i>	1677.13	3175.91	<0.001	1995.10	2366.75	0.286
<i>Avg_Bids_job</i>	8.30	10.00	0.004	9.42	10.00	0.375

Table R2 Effects of Generative AI on Labor Market Outcomes (After PSM)

	$\ln(y + 1)$			
	<i>Jobs</i>	<i>Bids</i>	<i>Transaction</i>	<i>Avg_Bids_job</i>
	(1)	(2)	(3)	(4)
<i>Treat * Post</i>	-0.213*** (0.0471)	-0.159** (0.0694)	-0.362*** (0.117)	0.112** (0.0443)
Constant	2.807*** (0.0089)	5.162*** (0.0131)	5.275*** (0.0221)	2.129*** (0.0084)
Submarket FE	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes
Observations	38,962	38,962	38,962	38,962
R^2	0.871	0.799	0.584	0.474

Note: Cluster-robust standard errors are reported at the submarket level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

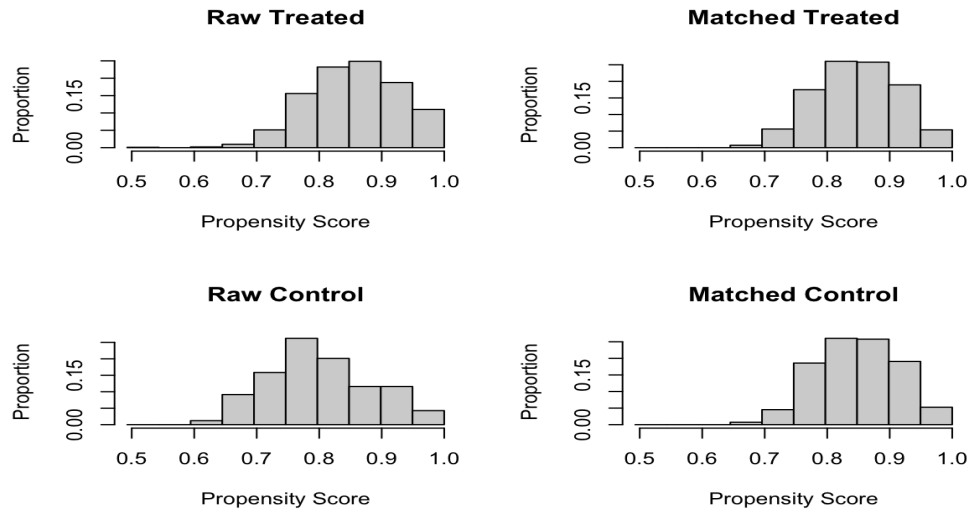


Figure R1 Propensity Score Distribution Pre- and Post-Matching

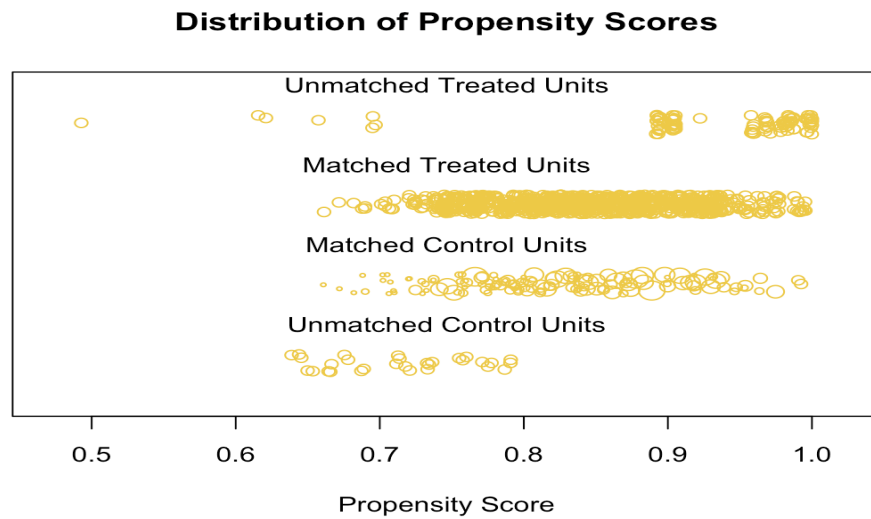


Figure R2 Comparison Between Matched and Unmatched Units in Propensity Score Matching

Appendix S: Robustness Check - Interrupted Time Series Analysis (ITSA)

ITSA is a quasi-experimental approach that projects the pre-intervention trend forward to form a counterfactual baseline, and then measures the intervention's effect as the deviation of actual post-intervention outcomes from that projected path (Kontopantelis et al. 2015, Campbell and Stanley 2015). We estimate the regression model as follows:

$$Outcome_{it} = \beta_0 + \beta_1 * Time_t + \beta_2 * Post_t + \beta_3 * PostTime_t + u_i + \epsilon_{it}. \quad (S1)$$

Here, $Outcome_{it}$ denotes the dependent variable (*Jobs*, *Bids*, *Transaction*, or *Avg_Bids_job*) for submarket i in month t . $Time_t$ represents the time elapsed since the beginning of the observation period. $Post_t$ is a binary indicator equal to 0 during the pre-treatment period and 1 during the post-treatment period. $PostTime_t$ equals 0 prior to the treatment (i.e., the launch of ChatGPT) and increases by one for each subsequent time period following the treatment. Furthermore, we incorporate u_i for the unit-level fixed effect. The error term ϵ_{it} encapsulates unobserved factors that influence the outcome. Our analysis centers on two coefficients of interest, β_2 and β_3 . The coefficient on $Post_t$ (β_2) captures the immediate effect of the treatment: a statistically significant value indicates a discrete level change in the outcome attributable to the emergence of ChatGPT. The coefficient on $PostTime_t$ (β_3) captures the trend effect, reflecting the change in the slope of the outcome trajectory following the intervention. Distinguishing the two matters because the influence of LLM-based generative AI may manifest both as an abrupt shift at the shock and as a gradual change that accumulates over the post-treatment period.

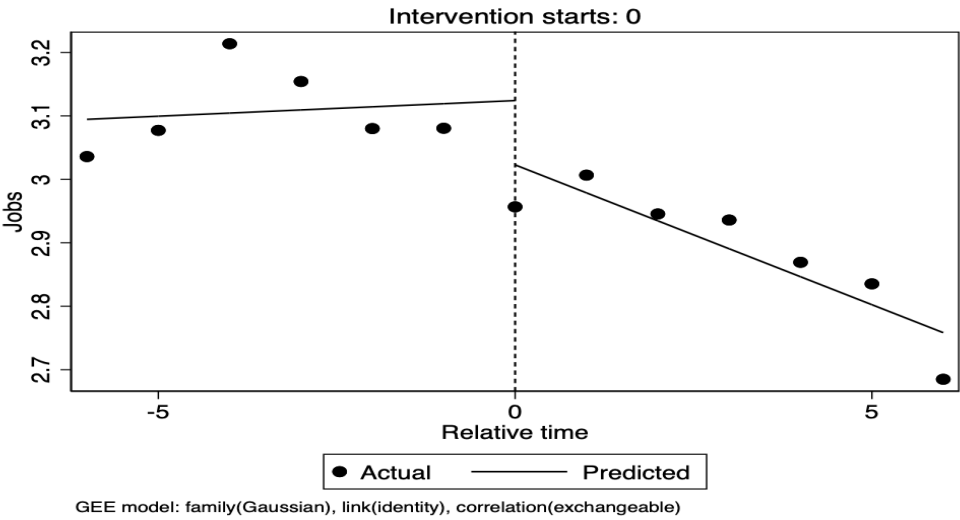
We estimate the model specified in Equation S1 and report the results in Table S1. The findings align closely with those from the DiD estimation. Specifically, labor demand (*Jobs*) and market transaction volume (*Transaction*) exhibit a significant immediate decline following the treatment ($\beta_2 = -0.101$ and -0.287 , respectively; $p < 0.01$), and labor demand also continues along a significant downward trend thereafter ($\beta_3 = -0.0490$ for *Jobs*; $p < 0.01$). On the supply side, by contrast, labor supply (*Bids*) shows no significant immediate response ($\beta_2 = -0.0055$, $p > 0.1$) and instead declines only gradually, as captured by a significant negative trend ($\beta_3 = -0.0361$, $p < 0.01$) whose absolute magnitude is smaller than its demand-side counterpart. As a result, competition within the submarket intensifies, reflected in a significant rise in the average number of bids per job (*Avg_Bids_job*; $\beta_2 = 0.0726$, $p < 0.01$). Consistent with our main analysis, these dynamics point to an asymmetric adjustment across the two sides of the market: demand contracts almost

at once as some clients substitute the technology for freelance labor, whereas the supply side responds with a lag, with freelancers recalibrating their participation only gradually. Figures S1 and S2 provide a more detailed and visually intuitive representation of these trends.

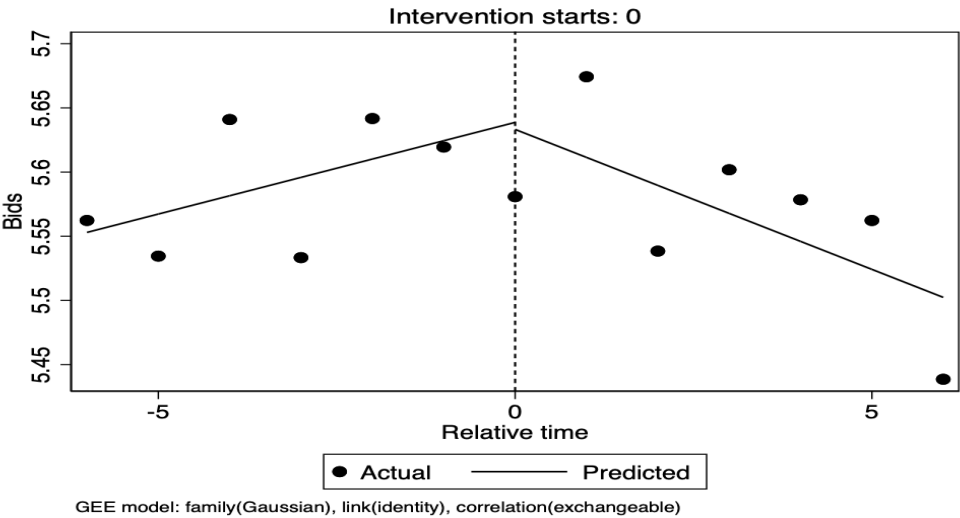
Table S1 Results for Interrupted Time Series Analysis

	$\ln(y+1)$			
	<i>Jobs</i>	<i>Bids</i>	<i>Transaction</i>	<i>Avg_Bids_job</i>
	(1)	(2)	(3)	(4)
<i>Time</i>	0.0049* (0.0028)	0.0143*** (0.0055)	-0.0193 (0.0144)	0.0132** (0.0053)
<i>Post</i>	-0.101*** (0.0128)	-0.0055 (0.0248)	-0.287*** (0.0715)	0.0726*** (0.0240)
<i>PostTime</i>	-0.0490*** (0.0040)	-0.0361*** (0.0075)	-0.0177 (0.0184)	0.0456*** (0.0076)
Constant	3.095*** (0.0356)	5.553*** (0.0496)	5.997*** (0.0795)	2.084*** (0.0282)
Submarket FE	Yes	Yes	Yes	Yes
Observations	11,921	11,921	11,921	11,921

Note: Cluster-robust standard errors are reported at the submarket level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

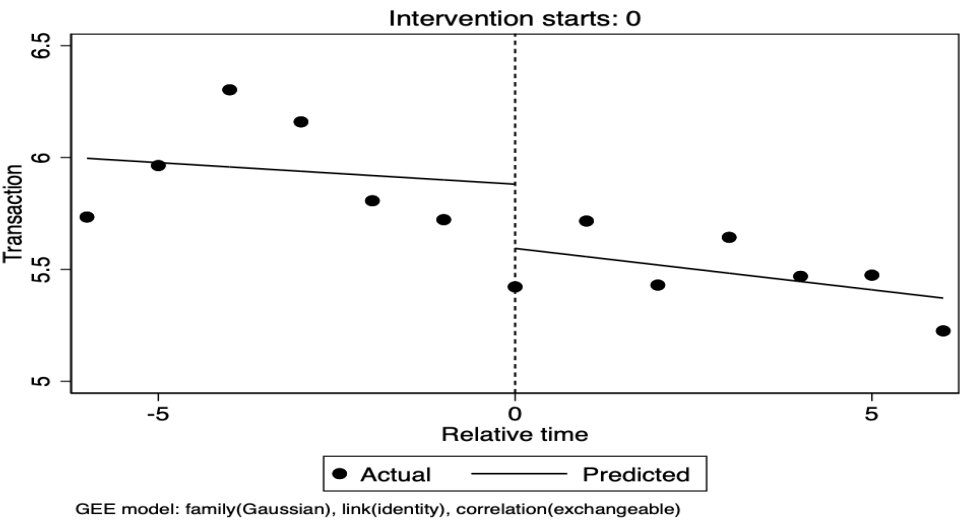


(a) Jobs

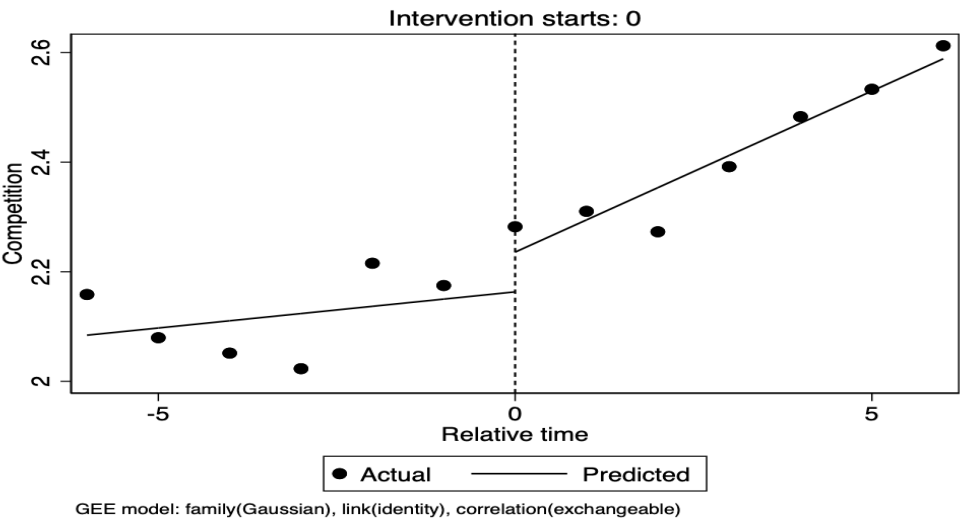


(b) Bids

Figure S1 Interrupted Time Series Analysis (ITSA)



(a) Transaction



(b) Avg_Bids_job

Figure S2 Interrupted Time Series Analysis (ITSA)

Appendix T: Robustness Check - More Details on Control Variable Inclusion and Omitted Variable Sensitivity Analyses

In this section, we incorporate control variables that capture both external conditions and internal market characteristics to obtain a conservative estimate of ChatGPT's impact and to conduct a sensitivity analysis.

Since online labor markets may be affected by external factors such as trends in the broader employment landscape, we account for these influences by collecting monthly Google Trends data for each skill, which reflects the relative popularity of each skill in a given month. Using the skill composition of each submarket, we construct a weighted monthly Google Trends index for each submarket to capture and control for fluctuations in the external environment. In addition, within the labor market, demand and supply interact dynamically. Demand in one period may influence supply in the subsequent period, while supply may, in turn, affect future demand. These two forces jointly determine the market transaction value and competition in the periods that follow. Therefore, when estimating the effect on demand (*Jobs*), we include lagged supply (*Lag_Bids*) as a control variable. Similarly, when estimating the effect on supply (*Bids*), we control for lagged demand (*Lag_Jobs*). In the estimation of transaction value and competition, both lagged demand and lagged supply (*Lag_Jobs* and *Lag_Bids*) are included as control variables to account for their joint influence. The estimation results are presented in Table T1 and remain consistent with our main findings.

After incorporating these observable controls, the estimated coefficients β_1 change from -0.221 to -0.222 for *Jobs*, from -0.168 to -0.141 for *Bids*, from -0.315 to -0.267 for *Transaction*, and from 0.138 to 0.111 for *Avg_Bids-job*.

Including post-treatment covariates is likely to attenuate the estimated treatment effect. A consistent estimator should capture the total influence of ChatGPT on the outcome, that is, it should capture the cumulative effect of all underlying mechanisms. In practice, ChatGPT may first alter demand and subsequently affect labor supply and transaction volume. Competition, in particular, can rise both because the number of available jobs falls and because changes in job mix propagate through bidding behavior before they are reflected in competitive metrics. Variables such as *Jobs* or *Bids* therefore lie on the causal pathway from the treatment to the final outcome, making them endogenous to ChatGPT adoption. Conditioning on such mediators severs part of the causal chain and introduces the classic "bad-control" problem, which mechanically biases the treatment coefficient toward zero. Hence, our reported estimates should be interpreted as conservative lower bounds; even so, they remain both economically meaningful and statistically significant.

Furthermore, to evaluate the risk of omitted variable bias, we conduct a sensitivity analysis following the approach proposed by Altonji et al. (2005), which has been widely applied in prior empirical studies (Sen et al. 2024, Petrova et al. 2021, Galor and Özak 2016). Specifically, we assess the relative sensitivity of unobserved factors by incorporating these additional control variables, thereby generating an alternative set of estimates. We then compare the new coefficient to that obtained from the regression without the additional controls, using the ratio defined as $\frac{\beta_{original}}{\beta_{original} - \beta_{control}}$. The rationale behind this ratio is straightforward: a smaller difference between the uncontrolled and controlled estimates suggests that the parameter of interest is less affected by selection on observables, thus requiring a stronger degree of selection on unobservables relative to observables to fully explain the observed effect. In our estimated results, the similarity of the coefficients suggests that selection on unobservables would need to be at least five times stronger than selection on observables to fully account for the observed effect (Oster 2019). This substantially alleviates concerns about omitted variable bias.

Table T1 Effects of Generative AI on Demand, Supply, and Matching Outcomes

	$\ln(y + 1)$			
	<i>Jobs</i>	<i>Bids</i>	<i>Transaction</i>	<i>Avg_Bids_job</i>
	(1)	(2)	(3)	(4)
<i>Treat * Post</i>	-0.222*** (0.0316)	-0.141*** (0.0469)	-0.267*** (0.0886)	0.111*** (0.0355)
Constant	2.234*** (0.181)	3.958*** (0.309)	2.972*** (0.651)	1.904*** (0.225)
External Confounder Controls	Yes	Yes	Yes	Yes
Internal Confounder Controls	Yes	Yes	Yes	Yes
Submarket FE	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes
Observations	25,833	25,833	25,833	25,833
R^2	0.892	0.842	0.614	0.545

Note: Cluster-robust standard errors are reported at the submarket level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix U: Robustness Check - Alternative Model for Count Variables

It is noteworthy that several of our variables, including *Jobs* and *Bids*, are count variables, taking on non-negative integer values. As shown in Appendix D, the variables exhibit substantial skewness, rendering count models not perfectly appropriate. Since the negative binomial model is relatively better suited for fitting such data distributions compared to the Poisson model (Stone 2013), we employ it to re-estimate the impact on these count variables. Table U1 presents the results of this analysis. The estimation results are qualitatively aligned with our main findings, corroborating the robustness of our conclusions. Specifically, we observe a significant decrease in labor demand in programming-intensive or non-programming-intensive submarkets, with no heterogeneity observed between these two types of treated submarkets. On the labor supply side, we still observe a significant decrease in non-programming-intensive submarkets, while the programming-intensive submarkets incur a significantly smaller supply loss, particularly in terms of the number of bids.

Table U1 Alternative Model for Count Variables		
	Negative Binomial Regressions	
	<i>Jobs</i>	<i>Bids</i>
	(1)	(2)
<i>Treat * Post</i>	-0.189*** (0.0437)	-0.207*** (0.0500)
<i>Treat * Post * Programming</i>	0.0070 (0.0332)	0.143*** (0.0391)
Constant	3.752*** (0.0106)	5.937*** (0.0173)
Submarket FE	Yes	Yes
Year-Month FE	Yes	Yes
Observations	25,833	25,833

Note: Cluster-robust standard errors are reported at the submarket level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix V: Robustness Check - Alternative Aggregation Level (Week)

Leveraging the two-year dataset, our primary analyses use variables aggregated at the monthly level to smooth trends, particularly for smaller submarkets. In this section, we aggregate all variables to the weekly level and rerun the estimations. The outcomes of this analysis are presented in Table V1. The estimation results remain qualitatively consistent with the main findings, confirming the robustness and reliability of our conclusions across different levels of data granularity. However, it is important to note that at the weekly level, there are a substantial number of zero observations, which makes it less appropriate to directly interpret the coefficients as proportional changes.

Table V1 Effects of Generative AI on Labor Market Outcomes (Week-level)

	$\ln(y + 1)$			
	<i>Jobs</i>	<i>Bids</i>	<i>Transaction</i>	<i>Avg_Bids_job</i>
	(1)	(2)	(3)	(4)
<i>Treat * Post</i>	-0.181*** (0.0225)	-0.127*** (0.0388)	-0.294*** (0.0520)	0.106*** (0.0281)
Constant	1.730*** (0.0073)	3.706*** (0.0126)	3.177*** (0.0169)	1.895*** (0.0091)
Submarket FE	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes
Observations	112,744	112,744	112,744	112,744
R^2	0.799	0.725	0.485	0.374

Note: Cluster-robust standard errors are reported at the submarket level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix W: Robustness Check - Alternative Fixed Effect

To more effectively account for seasonal effects, we also adopt an alternative specification that includes separate fixed effects for year (Y_t) and month (M_t), allowing us to control for both annual variations and recurring monthly patterns. This approach enhances interpretability by decomposing temporal variation into distinct components. The definitions and constructions of the remaining variables remain consistent with those outlined in Section 4. The model specification is presented as follows:

$$Outcome_{it} = \beta_0 + \beta_1 * Treat_i * Post_t + \beta_2 * Post_t + u_i + Y_t + M_t + \epsilon_{it} \quad (W1)$$

Using this model specification, we re-estimate the market-level outcomes, with the results reported in Table W1. The findings suggest that the alternative fixed effects structure yields results that are broadly consistent with those of the main specification.

Table W1 Effects of Generative AI on Demand, Supply, and Matching Outcomes (Alternative Fixed Effects)

	$ln(y + 1)$			
	<i>Jobs</i>	<i>Bids</i>	<i>Transaction</i>	<i>Avg_Bids_job</i>
	(1)	(2)	(3)	(4)
<i>Treat * Post</i>	-0.220*** (0.0318)	-0.168*** (0.0476)	-0.314*** (0.0896)	0.137*** (0.0356)
<i>Post</i>	0.147*** (0.0327)	0.211*** (0.0557)	-0.108 (0.121)	0.0343 (0.0467)
Constant	2.964*** (0.0065)	5.327*** (0.0129)	5.665*** (0.0358)	2.075*** (0.0120)
Submarket FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	25,833	25,833	25,833	25,833
R^2	0.890	0.840	0.610	0.538

Note: Cluster-robust standard errors are reported at the submarket level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix X: Robustness Check - Alternative Outcomes

In Section 5, we have analyzed the average post-ChatGPT changes in labor demand, supply, and matching outcomes for treated submarkets relative to control submarkets, with a primary focus on changes in the overall volume of demand, supply, and transactions. To further substantiate these findings, we introduce three additional count variables corresponding to the demand side, supply side, and matching outcomes. These variables serve to both complement and validate the aggregate analysis by capturing changes in the number of platform participants and successfully matched jobs. Similar to other market-level variables used in our analysis, these are constructed at the submarket-month level to ensure consistency in temporal and unit resolution across all outcome measures. Specifically, on the demand side, we introduce the variable *Clients*, denoting the number of unique clients who post jobs within the submarket in a given month. On the supply side, we adopt the variable *Freelancers*, representing the number of distinct freelancers who have placed at least one bid within a particular submarket during a month. To complement the analysis of market matching outcomes, we construct the variable *Matches*, denoting the number of successful matches made (i.e., successfully completed jobs) within a submarket in a given month.

Following the same procedure as in the main analysis presented in Section 5, we estimate the results using Equation 1, as specified in Section 4. The findings further confirm the observed contractions in demand, supply, and matching outcomes in the online labor market. Specifically, the results indicate a significant decrease in the number of participating clients ($\beta_1 = -0.199$, $p < 0.01$), active freelancers ($\beta_1 = -0.141$, $p < 0.01$), and successful matches ($\beta_1 = -0.102$, $p < 0.01$) in the treated submarkets relative to the control submarkets following the launch of ChatGPT. These additional analyses provide further support for the validity of our findings.

Summary statistics for these additional variables are reported in Table X1. The corresponding estimation results, based on our empirical specification, are presented in Tables X2 and X3. These results provide additional support for the robustness of our findings. Furthermore, the parallel trends assumption is supported for these variables, as illustrated in Figure X1.

Table X1 Variable Summary Statistics

Variable	Observation	Mean	Std	Min	Max
$\ln(Clients + 1)$	25,833	2.849	1.211	0	6.898
$\ln(Freelancers + 1)$	25,833	5.086	1.681	0	10.348
$\ln(Matches + 1)$	25,833	1.287	1.045	0	5.278

Table X2 Effects of Generative AI on Other Market Outcomes

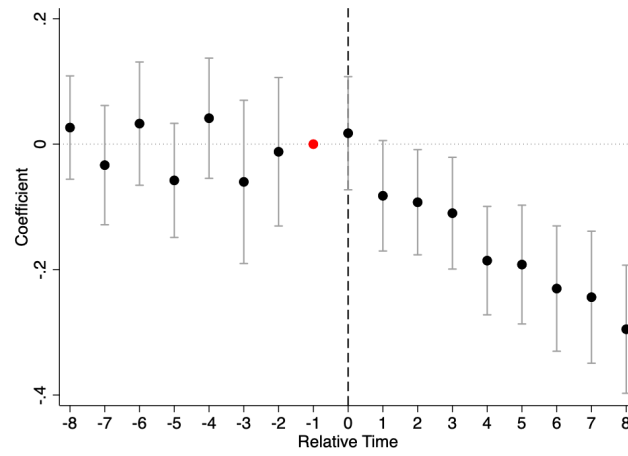
	$\ln(y + 1)$		
	<i>Clients</i>	<i>Freelancers</i>	<i>Matches</i>
	(1)	(2)	(3)
<i>Treat * Post</i>	-0.199*** (0.0292)	-0.141*** (0.0456)	-0.102*** (0.0252)
Constant	2.912*** (0.0093)	5.131*** (0.0146)	1.319*** (0.0081)
Submarket FE	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes
Observations	25,833	25,833	25,833
R^2	0.899	0.820	0.786

Note: Cluster-robust standard errors are reported at the submarket level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

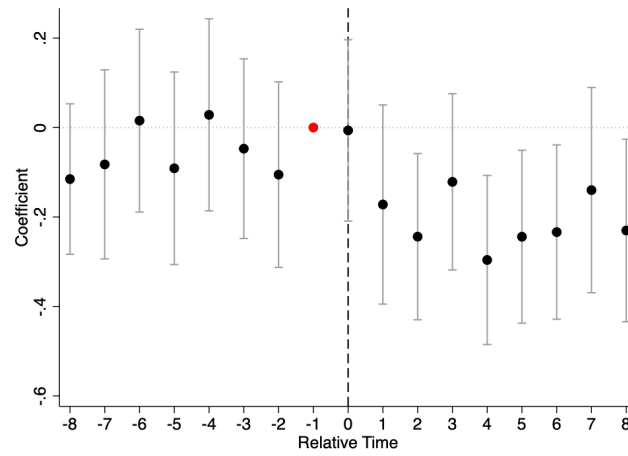
Table X3 Heterogeneous Effects of Generative AI on Other Market Outcomes

	$\ln(y + 1)$		
	<i>Clients</i>	<i>Freelancers</i>	<i>Matches</i>
	(1)	(2)	(3)
<i>Treat * Post</i>	-0.207*** (0.0317)	-0.194*** (0.0495)	-0.101*** (0.0278)
<i>Treat * Post * Programming</i>	0.0183 (0.0210)	0.117*** (0.0325)	-0.0007 (0.0216)
Constant	2.912*** (0.0093)	5.131*** (0.0146)	1.319*** (0.0081)
Submarket FE	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes
Observations	25,833	25,833	25,833
R^2	0.899	0.820	0.786

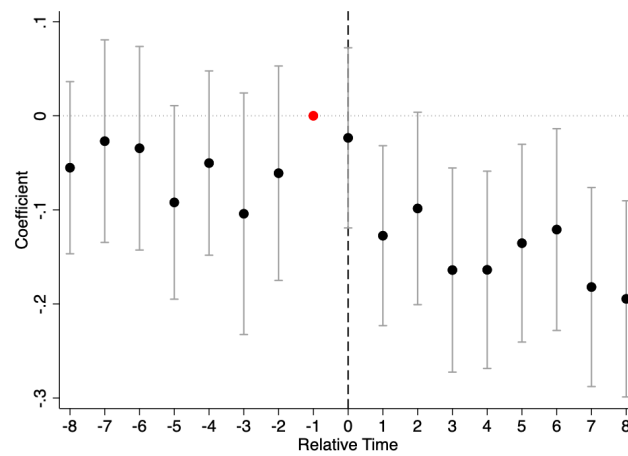
Note: Cluster-robust standard errors are reported at the submarket level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.



(a) Clients



(b) Freelancers



(c) Matches

Figure X1 Pretrends and Temporal Dynamics

Appendix Y: Robustness Check - Removing Outliers

Given that submarkets can vary significantly in size, one might question whether the estimates are disproportionately affected by the largest and the smallest submarkets. To address this concern, we first exclude the top 1% of submarkets, followed by the removal of submarkets with an average job post count of less than one. The re-estimated model results, as presented in Tables Y1 and Y2, remain consistent with our main findings. This indicates that our conclusions are not driven by a few large submarkets.

Table Y1 The Impact of Generative AI after Excluding Outliers (Drop Top Submarkets)

	$\ln(y + 1)$			
	<i>Jobs</i>	<i>Bids</i>	<i>Transaction</i>	<i>Avg_Bids_job</i>
	(1)	(2)	(3)	(4)
<i>Treat * Post</i>	-0.222*** (0.0319)	-0.173*** (0.0479)	-0.318*** (0.0902)	0.135*** (0.0358)
Constant	2.989*** (0.0102)	5.371*** (0.0153)	5.580*** (0.0288)	2.086*** (0.0114)
Submarket FE	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes
Observations	25,569	25,569	25,569	25,569
R^2	0.884	0.835	0.603	0.540

Note: Cluster-robust standard errors are reported at the submarket level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table Y2 The Impact of Generative AI after Excluding Outliers (Drop Bottom Submarkets)

	$\ln(y + 1)$			
	<i>Jobs</i>	<i>Bids</i>	<i>Transaction</i>	<i>Avg_Bids_job</i>
	(1)	(2)	(3)	(4)
<i>Treat * Post</i>	-0.219*** (0.0339)	-0.160*** (0.0480)	-0.299*** (0.0960)	0.137*** (0.0348)
Constant	3.074*** (0.0110)	5.495*** (0.0155)	5.727*** (0.0310)	2.115*** (0.0112)
Submarket FE	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes
Observations	25,306	25,306	25,306	25,306
R^2	0.882	0.829	0.589	0.540

Note: Cluster-robust standard errors are reported at the submarket level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix Z: Robustness Check - Continuous Treatment

To further verify the robustness of our main results, we re-estimate the main models using each submarket's LM-AIOE score as a continuous treatment measure, replacing the median-split binary classification. Specifically, we estimate the following specification:

$$Outcome_{it} = \beta_0 + \beta_1 * LM_AIOE_i * Post_t + u_i + T_t + \epsilon_{it} \quad (Z1)$$

where LM_AIOE_i denotes the continuous LM-AIOE score of submarket i , capturing its degree of exposure to LLM-based technologies, and all other variables are defined as before. The coefficient β_1 captures the differential effect of ChatGPT across submarkets with varying levels of LM-AIOE exposure. The estimation results are reported in Tables Z1 and Z2. Across all outcomes, the estimates remain consistent with our baseline findings in terms of direction and statistical significance, further supporting the robustness of our main conclusions.

Table Z1 The Impact of Generative AI at the Market Level (Continuous Treatment)

	$\ln(y + 1)$			
	<i>Jobs</i>	<i>Bids</i>	<i>Transaction</i>	<i>Avg_Bids_job</i>
	(1)	(2)	(3)	(4)
<i>LM_AIOE * Post</i>	-0.292*** (0.0357)	-0.221*** (0.0535)	-0.525*** (0.114)	0.187*** (0.0410)
Constant	3.030*** (0.0100)	5.414*** (0.0150)	5.671*** (0.0318)	2.080*** (0.0114)
Submarket FE	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes
Observations	25,833	25,833	25,833	25,833
R^2	0.891	0.841	0.612	0.541

Note: Cluster-robust standard errors are reported at the submarket level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table Z2 The Heterogeneous Impact of Generative AI at the Market Level (Continuous Treatment)

	$\ln(y+1)$			
	<i>Jobs</i>	<i>Bids</i>	<i>Transaction</i>	<i>Avg_Bids_job</i>
	(1)	(2)	(3)	(4)
<i>LM_AIOE * Post</i>	-0.296*** (0.0363)	-0.264*** (0.0540)	-0.536*** (0.116)	0.146*** (0.0419)
<i>LM_AIOE * Post * Programming</i>	0.0144 (0.0266)	0.159*** (0.0413)	0.0413 (0.0798)	0.153*** (0.0316)
Constant	3.030*** (0.0100)	5.407*** (0.0151)	5.669*** (0.0320)	2.073*** (0.0114)
Submarket FE	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes
Observations	25,833	25,833	25,833	25,833
R^2	0.891	0.841	0.612	0.542

Note: Cluster-robust standard errors are reported at the submarket level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix AA: Robustness Check - Multi-Group Analysis

To further assess the robustness of our findings to the choice of treatment categorization, we replace the median-split binary classification with a tercile-based approach, dividing submarkets into three groups (high, medium, and low exposure) based on their LM-AIOE values. We use the low-exposure group as the baseline and estimate separate treatment effects for the high- and medium-exposure groups. Tables AA1 and AA2 report the results for the main analysis and heterogeneity analyses, respectively.

As shown in Table AA1, the results remain consistent with our baseline findings: relative to the low-exposure group, both the high- and medium-exposure groups experience significant declines in jobs, bids, and transactions following ChatGPT's introduction, with stronger effects for the high-exposure group. Results in Table AA2 also show a heterogeneity pattern broadly similar to that of the main analysis. Together, these results further support the robustness of our main findings.

Table AA1 The Impact of Generative AI at the Market Level (Three Groups)

Panel A				
	High vs. Low			
	$\ln(y + 1)$			
	<i>Jobs</i>	<i>Bids</i>	<i>Transaction</i>	<i>Avg.Bids_job</i>
	(1)	(2)	(3)	(4)
<i>Treat * Post</i>	-0.258*** (0.0345)	-0.201*** (0.0518)	-0.421*** (0.0982)	0.167*** (0.0387)
Constant	3.007*** (0.0096)	5.382*** (0.0144)	5.534*** (0.0272)	2.046*** (0.0107)
Submarket FE	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes
Observations	14,855	14,855	14,855	14,855
R ²	0.886	0.837	0.614	0.526
Panel B				
	Medium vs. Low			
	$\ln(y + 1)$			
	<i>Jobs</i>	<i>Bids</i>	<i>Transaction</i>	<i>Avg.Bids_job</i>
	(1)	(2)	(3)	(4)
<i>Treat * Post</i>	-0.183*** (0.0331)	-0.135*** (0.0499)	-0.209** (0.0934)	0.109*** (0.0378)
Constant	2.825*** (0.0092)	5.088*** (0.0139)	5.299*** (0.0260)	2.055*** (0.0105)
Submarket FE	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes
Observations	14,850	14,850	14,850	14,850
R ²	0.894	0.845	0.631	0.543

Note: Cluster-robust standard errors are reported at the submarket level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table AA2 The Heterogeneous Impact of Generative AI at the Market Level (Three Groups)

Panel A				
	High vs. Low			
	$\ln(y+1)$			
	<i>Jobs</i>	<i>Bids</i>	<i>Transaction</i>	<i>Avg.Bids_job</i>
	(1)	(2)	(3)	(4)
<i>Treat * Post</i>	-0.282*** (0.0378)	-0.289*** (0.0567)	-0.461*** (0.107)	0.0982** (0.0422)
<i>Treat * Post * Programming</i>	0.0649* (0.0348)	0.234*** (0.0530)	0.107 (0.107)	0.185*** (0.0395)
Constant	3.007*** (0.0096)	5.382*** (0.0143)	5.534*** (0.0272)	2.046*** (0.0107)
Submarket FE	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes
Observations	14,855	14,855	14,855	14,855
R ²	0.886	0.837	0.614	0.527
Panel B				
	Medium vs. Low			
	$\ln(y+1)$			
	<i>Jobs</i>	<i>Bids</i>	<i>Transaction</i>	<i>Avg.Bids_job</i>
	(1)	(2)	(3)	(4)
<i>Treat * Post</i>	-0.169*** (0.0393)	-0.169*** (0.0604)	-0.186* (0.108)	0.0667 (0.0450)
<i>Treat * Post * Programming</i>	-0.0251 (0.0305)	0.0624 (0.0483)	-0.0426 (0.0871)	0.0776** (0.0379)
Constant	2.825*** (0.0092)	5.088*** (0.0139)	5.299*** (0.0260)	2.055*** (0.0105)
Submarket FE	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes
Observations	14,850	14,850	14,850	14,850
R ²	0.894	0.845	0.631	0.543

Note: Cluster-robust standard errors are reported at the submarket level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix AB: Robustness Check - Alternative Measures of Generative AI Exposure

To further examine the robustness of our results, we adopt an alternative peer-reviewed measure of occupational AI exposure from Eloundou et al. (2024), who assess the labor market impact potential of large language models across occupations. We reconstruct the treatment classification and re-estimate the models. Tables AB1 and AB2 present the corresponding estimation results. The findings remain largely consistent with the main analysis, further supporting the robustness of our conclusions.

Table AB1 The Impact of Generative AI at the Market Level (Alternative Measures of Generative AI Exposure)

	$\ln(y + 1)$			
	<i>Jobs</i>	<i>Bids</i>	<i>Transaction</i>	<i>Avg_Bids_job</i>
	(1)	(2)	(3)	(4)
<i>Treat * Post</i>	-0.233*** (0.0317)	-0.117** (0.0508)	-0.339*** (0.0890)	0.219*** (0.0354)
Constant	3.020*** (0.0098)	5.388*** (0.0157)	5.629*** (0.0274)	2.065*** (0.0109)
Submarket FE	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes
Observations	25,833	25,833	25,833	25,833
R^2	0.891	0.841	0.612	0.542

Note: Cluster-robust standard errors are reported at the submarket level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table AB2 The Heterogeneous Impact of Generative AI at the Market Level (Alternative Measures of Generative AI Exposure)

	$\ln(y + 1)$			
	<i>Jobs</i>	<i>Bids</i>	<i>Transaction</i>	<i>Avg_Bids_job</i>
	(1)	(2)	(3)	(4)
<i>Treat * Post</i>	-0.246*** (0.0344)	-0.171*** (0.0548)	-0.367*** (0.0958)	0.183*** (0.0382)
<i>Treat * Post * Programming</i>	0.0297 (0.0221)	0.118*** (0.0338)	0.0607 (0.0670)	0.0765*** (0.0266)
Constant	3.020*** (0.0098)	5.388*** (0.0156)	5.629*** (0.0274)	2.065*** (0.0109)
Submarket FE	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes
Observations	25,833	25,833	25,833	25,833
R^2	0.891	0.841	0.612	0.542

Note: Cluster-robust standard errors are reported at the submarket level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix AC: Recent Literature on AI in Labor Markets

Table AC1: Recent Literature on AI in Labor Markets

Title (Authors)	Research Context and Focus	Type of Technology	Level of Analysis	Demand-side or Supply-side	Relevant Takeaways
Threatened by AI: Analyzing Users' Responses to the Introduction of AI in a Crowd-Sourcing Platform (Lysyakov and Viswanathan 2023)	Impact of AI Logo System on a Crowd-sourcing Platform	Specialized AI	Designer-level	Both Sides	• Demand displacement effect: The AI launch reduces the number of lower-tier simple logo contests. • Workers' heterogeneous responses: Low-ability designers exit the platform, while higher-ability designers migrate to higher-tier logo contests or non-logo contests.
The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market (Hui et al. 2024)	Impact of Generative AI on Online Labor Markets	General-Purpose AI	Freelancer-level	Supply Side	• Demand displacement effect: AI-exposed freelancers experienced fewer jobs and lower earnings after ChatGPT's release. • Heterogeneous displacement effect: Top-rated AI-exposed freelancers are hurt more.
Skill-Biased Technical Change, Again? Online Gig Platforms and Local Employment (Guo et al. 2025)	Impact of Gig Platform Entry (TaskRabbit) on Local Worker Participation	Digital Platform	Market-level	Supply Side	• (Heterogeneous) demand displacement effect: TaskRabbit entry reduces the employment of middle-skilled workers whose tasks overlap with platform matching algorithms. • Redistribution effect: Displaced middle-skilled workers shift into self-employment as small business owners.
AI, Skill, and Productivity: The Case of Taxi Drivers (Kanazawa et al. 2026)	Impact of Routing AI on Taxi Drivers' Productivity	Specialized AI	Employee-level	Supply Side	• Productivity effect: Routing AI reduces cruising search time among taxi drivers. • Deskilling effect (heterogeneity): The productivity improvement mainly happens among low-skilled drivers, narrowing the productivity gap between high- and low-skilled workers and suggesting that AI substitutes for worker skills.
Who Is AI Replacing? The Impact of Generative AI on Online Freelancing Platforms (Demirci et al. 2025)	Impact of Generative AI on Online Labor Markets	General-Purpose AI	Market-level	Demand Side	• Demand displacement effect: ChatGPT's release reduces the number of automation-prone job postings relative to manual-intensive jobs.
Reskilling the Workforce for AI: Domain Expertise and Algorithmic Literacy (Tambe 2026)	Firms' Hiring for Algorithmic Skill	AI Technology at Large	Firm-level	Demand Side	• Firms increasingly hire algorithmic skills for non-technical occupations, particularly in roles with high decision-making responsibilities. • Skill-transition effect of tools: The diffusion of algorithmic skills is accelerated by reductions in learning costs (e.g., no-code tools).

Continued on next page

Table AC1 continued

Title (Authors)	Research Context and Focus	Type of Technology	Level of Analysis	Demand-side or Supply-side	Relevant Takeaways
Generative AI at Work (Brynjolfsson et al. 2025)	Impact of Generative AI Assistant on Customer Service Productivity	Specialized AI (Fine-Tuned for Customer Service)	Employee-level	Supply Side	• Productivity effect: AI assistance increases worker productivity. • Deskilling effect (heterogeneity): Low-skill and less-experienced workers have larger improvements, compressing the productivity distribution.
This Paper	Impact of Generative AI on Online Labor Markets	General-Purpose AI	Market-level & Freelancer-level	Both Sides	• Demand displacement effect: Generative AI reduces job postings in exposed submarkets. • Competition intensification effect: Labor supply declines less than demand, raising the average number of bids per job. • Skill-transition effect: Freelancers transition toward programming-intensive submarkets, even though those submarkets experience a similar demand displacement. • Heterogeneous skill-transition effect: High-skilled freelancers are more likely to transition, consistent with differences in absorptive capacity.

Appendix References

- Al Balawi R, Hu Y, Qiu L (2023) Brand crisis and customer relationship management on social media: Evidence from a natural experiment from the airline industry. *Information Systems Research* 34(2):442–462.
- Altonji JG, Elder TE, Taber CR (2005) Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of political economy* 113(1):151–184.
- Ananthakrishnan UM, Xu X, Mitkina M, Tan Y (2025) Political consumerism and the emergence of rare information on user-generated content platforms. *Management Science*, *Forthcoming* .
- Angrist JD, Pischke JS (2009) *Mostly harmless econometrics: An empiricist's companion* (Princeton university press).
- Braghieri L, Levy R, Makarin A (2022) Social media and mental health. *American Economic Review* 112(11):3660–3693.
- Brynjolfsson E, Li D, Raymond L (2025) Generative ai at work. *The Quarterly Journal of Economics* 140(2):889–942.
- Callaway B, Goodman-Bacon A, Sant'Anna PH (2024) Difference-in-differences with a continuous treatment. Technical report, National Bureau of Economic Research.
- Campbell DT, Stanley JC (2015) *Experimental and quasi-experimental designs for research* (Ravenio books).
- Cassano F, Gouwar J, Nguyen D, Nguyen S, Phipps-Costin L, Pinckney D, Yee MH, Zi Y, Anderson CJ, Feldman MQ, Guha A, Greenberg M, Jangda A (2023) Multipl-e: A scalable and polyglot approach to benchmarking neural code generation. *IEEE Trans. Softw. Eng.* 49(7):3675–3691, ISSN 0098-5589, URL <http://dx.doi.org/10.1109/TSE.2023.3267446>.
- Chan J, Mojumder P, Ghose A (2019) The digital sin city: An empirical study of craigslist's impact on prostitution trends. *Information systems research* 30(1):219–238.
- Chen FF, Saffar W, Shang L (2025) Local agglomeration and household mortgage debt. *Management Science* 71(6):5001–5021.
- Demirci O, Hannane J, Zhu X (2025) Who is ai replacing? the impact of generative ai on online freelancing platforms. *Management science* 71(10):8097–8108, ISSN 0025-1909.
- Eloundou T, Manning S, Mishkin P, Rock D (2024) Gpts are gpts: Labor market impact potential of llms. *Science* 384(6702):1306–1308.
- Faccio M, O'Brien WJ (2021) Business groups and employment. *Management Science* 67(6):3468–3491.
- Felten E, Raj M, 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 EW, Raj M, Seamans R (2023) Occupational heterogeneity in exposure to generative ai. *Available at SSRN 4414065* .

- Galor O, Özak Ö (2016) The agricultural origins of time preference. *American economic review* 106(10):3064–3103.
- Gong J, Greenwood BN, Song Y (2023) An empirical investigation of ridesharing and new vehicle purchase. *Manufacturing & Service Operations Management* 25(3):884–902.
- Guo X, Cheng Z, Pavlou PA (2025) Skill-biased technical change, again? online gig platforms and local employment. *Information Systems Research* 36(3):1354–1374.
- Hui X, Reshef O, Zhou L (2024) The short-term effects of generative artificial intelligence on employment: Evidence from an online labor market. *Organization Science* 35(6):1977–1989.
- Jiang DL, Ye S, Zhao L, Gu B (2025) Do reductions in search costs for partial information on online platforms lead to better consumer decisions? evidence of cognitive miser behavior from a natural experiment. *Information Systems Research* 36(3):1780–1798.
- Kanazawa K, Kawaguchi D, Shigeoka H, Watanabe Y (2026) Ai, skill, and productivity: The case of taxi drivers. *Management Science* 72(2):1376–1388.
- Kocetkov D, Li R, allal LB, LI J, Mou C, Jernite Y, Mitchell M, Ferrandis CM, Hughes S, Wolf T, Bahdanau D, Werra LV, de Vries H (2023) The stack: 3 TB of permissively licensed source code. *Transactions on Machine Learning Research* ISSN 2835-8856, URL <https://openreview.net/forum?id=pxpbTdUEpD>.
- Kontopantelis E, Doran T, Springate DA, Buchan I, Reeves D (2015) Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. *bmj* 350.
- Li S, Zhan X (2019) Product market threats and stock crash risk. *Management Science* 65(9):4011–4031.
- Lokkila E, Christopoulos A, Laakso MJ (2023) A data-driven approach to compare the syntactic difficulty of programming languages. *Journal of Information Systems Education* 34(1):84–93.
- Lu Y, Yu L (2015) Trade liberalization and markup dispersion: evidence from china's wto accession. *American Economic Journal: Applied Economics* 7(4):221–253.
- Lysyakov M, Viswanathan S (2023) Threatened by ai: Analyzing users' responses to the introduction of ai in a crowd-sourcing platform. *Information Systems Research* 34(3):1191–1210.
- Nguyen D, Nam L, Dau A, Nguyen A, Nghiem K, Guo J, Bui N (2023) The vault: A comprehensive multilingual dataset for advancing code understanding and generation. Bouamor H, Pino J, Bali K, eds., *Findings of the Association for Computational Linguistics: EMNLP 2023*, 4763–4788 (Singapore: Association for Computational Linguistics), URL <http://dx.doi.org/10.18653/v1/2023.findings-emnlp.316>.
- Oster E (2019) Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics* 37(2):187–204.
- Ousterhout JK (1998) Scripting: Higher level programming for the 21st century. *Computer* 31(3):23–30.
- Perez-Truglia R (2020) The effects of income transparency on well-being: Evidence from a natural experiment. *American Economic Review* 110(4):1019–1054.

- Petrova M, Sen A, Yildirim P (2021) Social media and political contributions: The impact of new technology on political competition. *Management Science* 67(5):2997–3021.
- Prechelt L (2000) An empirical comparison of seven programming languages. *Computer* 33(10):23–29.
- Ran Z, Rau PR, Ziegler T (2025) Sometimes, always, never: Regulatory clarity and the development of digital financing. *Management Science* 71(9):8027–8071.
- Schmidheiny K, Sieglöcher S (2023) On event studies and distributed-lags in two-way fixed effects models: Identification, equivalence, and generalization. *Journal of Applied Econometrics* 38(5):695–713.
- Sen A, Grad T, Ferreira P, Claussen J (2024) (how) does user-generated content impact content generated by professionals? evidence from local news. *Management Science* 70(9):6045–6068.
- Stone JV (2013) Bayes' rule : a tutorial introduction to bayesian analysis.
- Tambe PB (2026) Reskilling the workforce for ai: Domain expertise and algorithmic literacy. *Management Science* 72(1):515–537.
- Wang H, Overby E (2023) Do political differences inhibit market transactions? an investigation in the context of online lending. *Management Science* 69(8):4685–4706.
- Ye T, Shankar V (2025) Close a store to open a pandora's box? the effects of store closure on sales, omnichannel shopping, and mobile app usage. *Marketing Science* 44(4):820–837.
- Zhang Y, Zheng Z, Gu B (2025) Omniscience or differentiation? an empirical study of knowledge structure and career development of it workers. *Information Systems Research* 36(2):1129–1146.