Who Is AI Replacing? The Impact of Generative AI on Online Freelancing Platforms[*]

Ozge Demirci[†] Jonas Hannane[‡] Xinrong Zhu[§]

Abstract

This paper studies the impact of Generative AI technologies on the demand for online freelancers using a large dataset from a leading global freelancing platform. We identify the types of jobs that are more affected by Generative AI and quantify the magnitude of the heterogeneous impact. Our findings indicate a 21% decrease in the number of job posts for automation-prone jobs related to writing and coding, compared to jobs requiring manual-intensive skills, within eight months after the introduction of ChatGPT. We show that the reduction in the number of job posts increases competition among freelancers while the remaining automation-prone jobs are of greater complexity and offer higher pay. We also find that the introduction of Image-generating AI technologies led to a 17% decrease in the number of job posts related to image creation. We use Google Trends to show that the more pronounced decline in the demand for freelancers within automation-prone jobs correlates with their higher public awareness of ChatGPT's substitutability.

Keywords: Generative AI, large language models, ChatGPT, digital freelancing platforms
JEL No: O33, E24, J21, J24

[*]The author names are listed in alphabetical order. The authors thank Tomaso Duso, Ayelet Israeli, Stephan Seiler, Hannes Ullrich, and Xu Zhang for their valuable discussions and suggestions.

[†]Harvard Business School, odemirci@hbs.edu
[‡]DIW Berlin and Technische Universität Berlin, jhannane@diw.de
[§]Imperial College London Business School, x.zhu@imperial.ac.uk

1 Introduction

Recent advancements in artificial intelligence (AI) and natural language processing have brought changes to many industries. Among the latest innovations is ChatGPT, a large language model developed by OpenAI, which has demonstrated a remarkable capacity to generate human-like text responses that are coherent and context-relevant (Dwivedi et al. 2023). These groundbreaking technologies could have a profound impact on online labor markets (OLM). Freelancer jobs, once solely reliant on human expertise, now face the growing influence of automation due to the emergence of AI tools.

This paper examines the short-term impacts of Generative AI (GenAI) technologies on the demand for freelance jobs in online labor markets. We identify the types of jobs that are more affected by GenAI and quantify the magnitude of the impact. Online freelancer markets offer an ideal setting to study the short-term impact of GenAI tools on labor markets. These markets are characterized by flexible, short-term, task-oriented, and remote jobs. Likewise, the typical tasks for which people use AI tools are small, flexible, and short-term. Despite the unique features of online labor markets compared to traditional ones, examining AI's effects on these markets provides an opportunity to glean insights into broader contexts, with implications potentially extending to sectors beyond contract employment (Agrawal et al. 2015).

We analyze data from a leading global online freelancing platform consisting of 1,388,711 job posts from July 2021 to July 2023. Using a network clustering algorithm and leveraging detailed job post descriptions on skill and software requirements, we categorize job posts into distinct clusters such as data and office management, writing, and engineering. Based on the AI Occupational Exposure Index (AIOE) constructed by Felten et al. (2021, 2023), these clusters of jobs exhibit different exposure levels to large language model AI tools.[1] Accordingly, the clusters can be classified into three types: manual-intensive jobs (e.g., data and office management, video services, and audio services), automation-prone jobs (e.g., writing, software, app, and web development, engineering), and image-generating jobs (e.g., graphic design and 3D modeling). Manual-intensive jobs have notably smaller AIOE compared to automation-prone jobs, indicating lower exposure to Large Language Models (LLMs). We study the differential impacts of the introduction of GenAI tools on demand across these different types of job clusters. Our empirical framework comprises different versions of difference-in-differences designs, including standard DiD and recent methodological advances

[1]AIOE measures the extent to which occupations are exposed to AI language modeling advances through a survey, with higher values indicating higher susceptibility. Occupations with high AIOE include writers, authors and engineers.

such as Synthetic DiD (Arkhangelsky et al. 2021), and doubly robust DiD (Sant'Anna and Zhao 2020, Callaway and Sant'Anna 2021).

Our first set of results focuses on the impact of the release of ChatGPT. Comparing automation-prone jobs with manual-intensive ones, we find that the number of job posts for automation-prone jobs decreased by 20.86% more than for manual-intensive jobs within eight months after the introduction of ChatGPT. This decline indicates a significant drop in demand for freelancer jobs involving more repetitive tasks (e.g., writing) and coding and automation (e.g., software, website/app development, and engineering). Writing jobs experienced the most significant decrease in demand (30.37%), followed by software, website/app development (20.62%), and engineering (10.42%). While this decrease in the number of job posts intensifies competition between freelancers, we find that the remaining automation-prone jobs are of greater complexity and offer higher pay. Second, we assess the impact of GenAI tools for image creation, specifically the release of Midjourney, Stable Diffusion, and DALL-E 2, on the demand for jobs related to image creation and graphic design. We find that the introduction of Image-generating AI technologies led to a 17.01% decrease in the number of job posts for graphic design (18.49%) and 3D modeling (15.57%) relative to manual-intensive jobs. These declines in demand are larger than the seasonal variation observed on the platform or the documented effect of automation in traditional labor markets (Acemoglu et al. 2020). Our findings are robust across all empirical models of DiD and various robustness tests using alternative reference groups.

To strengthen the causal link between the differential demand decrease and the introduction of ChatGPT, we incorporate an external index—Google Trends Search Volume Indices (Google SVI), constructed by using co-search key terms such as "ChatGPT" combined with the descriptions of job clusters (e.g., ChatGPT writing). We consider SVI as a proxy for interest and awareness of the potential substitutability of ChatGPT in certain tasks. The Google SVI for writing, engineering, software, app, and web development exhibited significant growth compared to other jobs after the introduction of ChatGPT. We find a negative relationship between changes in the number of job posts within a cluster and Google SVI. This indicates that in job clusters with higher awareness or interest in AI tools' substitutability, there was a greater decrease in demand for freelancers.[2]

Our study contributes to the extensive literature on the impact of automation on labor markets (Acemoglu and Autor 2011, Acemoglu and Restrepo 2018, 2019, 2020, Brynjolfsson et al. 2018, Agrawal et al. 2019). Previous research documented how automation, including robots and machine learning, displaces certain jobs while creating new opportunities, leading to a complex reallocation of labor. These studies highlight that tasks involving rou-

[2]For one standard deviation increase in SVI, we estimate a decrease of 8.01% in the number of job posts.

tine, repetitive actions are more vulnerable to automation, whereas those requiring complex problem-solving and creativity are less affected. Our research complements this literature by focusing on GenAI's distinct capabilities to automate tasks such as coding, writing, and image creation. Compared to previous technologies, GenAI stands out as a versatile technology with wide-ranging applications and ease of integration and adaptability. Its rapid advancement and broad applicability indicate a potentially deeper and more far-reaching impact on labor markets. We provide new insights into how GenAI reshapes demand for various human skills in the short term, with the effect not only persisting but also growing over the sample period. Given the growing awareness of GenAI, as evidenced by Google SVI, our short-term findings may serve as indicators of long-term impacts.

Our paper also contributes to the growing literature on the impact of GenAI on labor markets and economic dynamics. Some earlier work focuses on measuring the exposure of different occupations to AI, proposing methodologies to identify the industries, jobs, or regions most affected by AI technologies (Brynjolfsson et al. 2018, Felten et al. 2021, 2023). Another line of literature studies the impact of AI technologies on aspects of economic activity, such as worker productivity (Brynjolfsson et al. 2023, Peng et al. 2023, Noy and Zhang 2023), writing assistance (Wiles and Horton 2023), firm value (Eisfeldt et al. 2023), market research (Brand et al. 2023), digital public goods (Burtch et al. 2023, del Rio-Chanona et al. 2023, Shan and Qiu 2023, Yilmaz et al. 2023), user-generated content (Knight and Bart 2023) and labor markets (Eloundou et al. 2023, Hui et al. 2023).

To the best of our knowledge, we are among the first ones to utilize job post data to examine the impact of Generative AI on online labor markets, offering unique insights due to two key features. First, our dataset enables us to directly measure freelancer demand by tracking the number of job posts over time. Using this measure, we quantify substitution effects and analyze heterogeneity across job types, revealing nuanced trends such as initial declines followed by escalating reductions. This complements existing literature on labor market changes due to automation. Second, the rich information in our dataset about skill and job requirements allows us to analyze how the nature of job posts evolve after GenAI tools. Our findings suggest interesting labor market dynamics, including potential adaptation in the labor force (emerging skill requirements like "using ChatGPT"), as well as higher pay and greater complexity in the remaining jobs.

The aforementioned key features enable us to provide several unique and complementary perspectives relative to concurrent papers Hui et al. (2023), Liu et al. (2023) and Qiao et al. (2023), which use employment histories or freelancer profiles to examine the impact of GenAI on freelancers' employment outcomes.[3] Our data helps address the potential

[3]Employment histories or freelancer profiles record only the jobs that are acquired by freelancers and approved

underestimation of demand that can occur when relying solely on employment histories, which may not fully capture job market dynamics due to the infrequent job acquisition among freelancers.[4] Additionally, our dataset encompasses all job posts on the platform rather than a subsample of active freelancers, ensuring a more representative analysis of the online labor market.[5] We employ a data-driven approach to cluster similar job posts, classify these clusters based on existing literature (Eisfeldt et al. 2023, Felten et al. 2021, 2023), and validate our classifications using Google SVI. While Hui et al. (2023) explore job classifications by comparing writing tasks with other jobs, our study extends on their work by quantifying the heterogeneous impacts of Generative AI across a broader range of job types. This approach enriches our understanding of AI's impact in reshaping labor demand.

Despite being in its early stages, GenAI's effects on the online labor markets are becoming discernible, which might indicate potential shifts in long-term labor market dynamics. Our findings on AI's heterogeneous short-term impacts on online freelance jobs hold implications for managers and policymakers. By highlighting potentially more impacted jobs by AI in the evolving employment landscape, our findings provide insights into the responsible and effective implementation of AI tools in the workplace.

The structure of the paper is as follows: Section 2 introduces institutional details, including GenAI tools and online labor markets. Section 3 describes our data sources and sample construction. Section 4 presents our empirical analyses and results. Section 5 concludes.

2 Institutional Details

2.1 Generative AI

Generative AI involves the creation of content, such as images, text, and music, that closely resembles human creations. OpenAI launched its AI Conversationalist, ChatGPT, on November 30, 2022, and the platform rapidly gained attention. By January 2023, ChatGPT was estimated to have reached 100 million monthly active users.[6] The Google search

by employers.

[4]Competition among freelancers on OLMs is intense (Beerepoot and Lambregts 2015), particularly affecting new freelancers who lack reputation (Pallais 2014). Hui et al. (2023) mention that a freelancer starts a job once every three months on average. Similarly, Qiao et al. (2023) note that the number of jobs obtained per freelancer per month is 0.3 on average. In our sample, the acceptance rate of job posts is about 25%.

[5]Liu et al. (2023) and Qiao et al. (2023) select random subsamples of 4321 and 5000 freelancers from the platforms, respectively.

[6]Source: https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/. https://explodingtopics.com/blog/chatgpt-users.

volume for ChatGPT surpassed that of other major AI,[7] peaking in April 2023.[8] Earlier in 2022, other Image-generating AI tools like DALL-E 2, Midjourney, and Stable Diffusion, were also introduced. These tools generate realistic images based on text descriptions. The release dates of these image-generating tools vary over time, depending on their versions and accessibility to the public. Figure A1 provides a timeline of the release dates of each GenAI technology to the general public.

2.2 Online Labor Market

Online labor markets (OLM) are a digital hub where freelancers offer specialized skills to potential employers. Platforms such as Upwork, Freelancer.com, and Fiverr facilitate this connection, allowing employers to post job listings on which freelancers can bid. The online freelancer market has gained popularity in recent years due to its flexibility, global reach, and efficient matching between freelancers and employers (Kässi and Lehdonvirta 2018). Kässi et al. (2021) estimate that by 2020, 8.5 million freelancers worldwide had obtained work and 2.3 million freelancers had found full-time jobs on OLM platforms.

Jobs on OLM platforms vary in scope, ranging from short-term data entry assignments to relatively more complex software development. Furthermore, OLM platforms led to a fragmentation of work into smaller tasks, where employers do not develop long-term relationships with freelancers (Graham and Anwar 2019). Employers can easily terminate jobs or rehire different freelancers, resulting in more flexible hiring decisions compared to the offline labor market. A substitution effect may emerge as employers favor AI-driven solutions for their cost-effectiveness, accessibility, and efficiency in handling repetitive tasks. Therefore, OLM constitutes a good setting for studying early trends in the impact of GenAI on employment.

3 Data

3.1 Freelancing Platform Data

The data were collected from an undisclosed, globally leading OLM platform using its API. On this platform, employers post their jobs and their budget range, specifying both the maximum and minimum amounts. The scope and requisites of a job post are outlined in the job description, which includes a task description (e.g., creating a short video) and desired skills (e.g., Video Editing, Video Production, Final Cut Pro, and Adobe Premiere Pro). The platform uses skill tags to optimize the matching process between employers and freelancers.

[7] Source: https://trends.google.com/trends/explore?q=chatgpt,bing%20AI,google%20bard&hl=en

[8] Source: https://trends.google.com/trends/explore?q=chatgpt&hl=en

These tags, chosen from a standardized list or entered manually by the employer, are included in each job post. Freelancers indicate their skills on their profiles, and only those whose skills match the job are eligible to bid on it. Eligible freelancers submit bids with their proposed price and time frame or may be directly invited by the employer. Employers then review bids and select freelancers based on expertise and bid details.

The data spans from July 2021 to July 2023 and includes all job posts on the demand side of this online platform. For each job post, we observe its title, job descriptions (including skill tags and preferred software), maximum and minimum budget range set by the employer, whether the payment is fixed or hourly, whether the job needs to be done by local freelancers ("local jobs"), the number of bids and average bidding price per job post, the date, location (country and city) and employer id of the posts, and the final status (awarded, expired, etc).[9] The data contains 2,712 unique skill tags, which are used in the next subsection to categorize job posts into distinct clusters. In our empirical analysis, we also use the unique number of skill tags of a job post as a measure of the job's complexity.

Classification of Job Posts. Our empirical analysis examines demand changes across various job types after GenAI tools are introduced. We first cluster job posts based on skill co-occurrences, allowing for a finer categorization beyond platform-defined broad labels like "design" or "trades and services."[10] Specifically, we apply an unsupervised clustering algorithm, the Louvain method (Blondel et al. 2008), to detect skill clusters that frequently occur together in job posts. This method is widely used for finding hidden structures in large networks, such as in social network analysis and recommendation systems.

Our algorithm detects 42 different clusters of skills in our data, representing distinct skill sets or software requirements necessary to perform specific tasks. In the next step, we map each job post to the cluster with the greatest overlap in skills. We conduct data cleaning by focusing on highly prevalent clusters (prevalence equal to or greater than 0.12%, which drops about 0.25% of all job posts) and merging three similar clusters together. This process yields 15 distinct clusters (Table C1). The technical details and sample construction are presented in Appendix B. Examining the skill tags and detailed job post descriptions and drawing on previous literature, we further characterize the job clusters into the following types (see Table C2 for these job clusters and their top 10 skill tags):
1. Manual-intensive jobs, including data and office management, video services, and audio services. These jobs require a large proportion of manual tasks. For example, data and office management frequently require freelancers skilled in working with Excel to create or

[9] We observe the time when a job post was last updated through the API.

[10] Rather than relying on broad job categories provided on the platform, our data-driven categorization is important for capturing the heterogeneous impact of GenAI on various jobs (Felten et al. 2023).

edit spreadsheets; audio services involve tasks such as audio production and sound design, and video services typically involve video creation or editing. These are fields where human labor provides unique value.[11]

2. Automation-prone jobs, including writing, engineering, and software, app, and web development. These clusters often involve tasks that are susceptible to digitalization or automation. The writing cluster, which includes proofreading, ghostwriting, and editing, is identified as one of the occupations most vulnerable to ChatGPT according to the previous literature (Eloundou et al. 2023). The engineering cluster includes electrical engineering and circuit design tasks requiring proficiency in coding like Mathematica, Matlab, and C programming. LLM has demonstrated effectiveness in simplifying and accelerating circuit development (Blocklove et al. 2023). The software, app, and web development cluster primarily includes job posts for website or app developers, which also require coding skills. ChatGPT has been shown to perform well with easy and medium programming problems (Bucaioni et al. 2024, Coello et al. 2024).

3. Image-generating jobs such as graphic design and 3D modeling. They primarily involve creating and modifying visual content and virtual three-dimensional models. In Section 4, we examine the impact of Image-generating AI tools on demand in these job clusters.

Notably, these eight clusters exhibit distinct exposure to AI, according to the AI Occupational Exposure Index (AIOE) introduced by Felten et al. (2021) and Felten et al. (2023). This index measures the extent to which occupations are exposed to advances in AI language modeling capabilities, encompassing either substitution or augmentation effects.[12] A higher AIOE value indicates greater susceptibility to Large Language Models. Table C3 presents the AIOE index for manual-intensive and automation-prone clusters.[13] In particular, manual-intensive jobs exhibit significantly lower AIOE compared to automation-prone jobs, suggesting that the former are expected to be less exposed to LLMs.

Based on these discussions, we focus on these eight clusters in our main analysis.[14] We

[11] During our analysis period, the versions of ChatGPT (3.5 and 4) did not demonstrate effective functionalities for these tasks.

[12] The AIOE index is constructed through a survey among Amazon Mechanical Turk (mTurk) workers. The survey assesses the capability of LLMs to perform tasks related to 52 distinct human abilities (e.g., oral comprehension, inductive reasoning). These 52 human abilities align with the Occupational Information Network (O*NET) database developed by the US Department of Labor to describe the occupational makeup. Linking these data together, Felten et al. (2023) calculate the AIOE for each occupation. For public AIOE datasets, please see `https://github.com/AIOE-Data/AIOE`.

[13] The AIOE index is exclusively measured for Large Language Models, not Image-generating AI tools.

[14] To ensure a clean comparison, we exclude legal, accounting, and finance, given that some of the job posts in these clusters require specific credentials (e.g., attorneys and CPAs). We also do not include social media marketing, internet marketing, and statistical analysis clusters due to non-parallel pre-trends. These clusters constitute only 9.34% of the entire sample, and our robustness checks in Appendix E confirm that their exclusion does not significantly affect our estimates. Additionally, we do not examine labor demand changes in translation, blockchain, smart contracts,

additionally exclude job posts with outlier maximum budget in the top 1% and restrict our sample to the 61 largest countries, which accounts for 95% of all job posts. We focus specifically on fixed-payment jobs, which constitute around 80% of the remaining job posts. The final sample includes 1,218,463 job posts from 541,828 employers. Table C4 provides summary statistics for key outcome variables. Finally, to capture overall demand on the platform, we aggregate the sample to the cluster-week-country level. We calculate the number of job posts and balance the sample by filling in zeros for cluster-week-country combinations with no job posts during a specific week. Table 1 summarizes the prevalence of the clusters in our analysis and provides summary statistics of the log number of posts at the cluster-week-country level before and after the GenAI tools. It shows a more prominent decline in the average number of job posts in automation-prone and image-generating clusters compared to manual-intensive ones after the introduction of ChatGPT and Image-generating AI.

Table 1: Cluster Summary Statistics

	Before ChatGPT		After ChatGPT	
	Log # of Posts	Percent (%)	Log # of Posts	Percent (%)
Manual Intensive				
Data and Office Management	2.08 (1.18)	8.59	1.84 (1.16)	8.64
Audio Services	0.63 (0.81)	0.9	0.56 (0.79)	1.07
Video Services	1.26 (1.04)	2.92	1.19 (1.04)	3.93
Automation Prone				
Writing	2.23 (1.21)	10.02	1.74 (1.16)	7.87
Software, App and Web Development	3.59 (1.11)	35.32	3.23 (1.08)	33.68
Engineering	1.1 (1.02)	2.16	0.86 (0.91)	1.91
	Before Image-generating AI		After Image-generating AI	
Manual Intensive				
Data and Office Management	2.13 (1.17)	8.45	1.88 (1.17)	8.82
Audio Services	0.64 (0.81)	0.87	0.57 (0.79)	1.06
Video Services	1.31 (1.04)	2.86	1.17 (1.04)	3.63
Image Generating				
Graphic Design	3.05 (1.16)	22.15	2.69 (1.21)	24.25
3D Modelling	1.81 (1.13)	5.45	1.49 (1.15)	5.94

Notes: This table reports the log number of job posts in each cluster for pre- and post-periods of ChatGPT and Image-generating AI, respectively. The sample is at the week-cluster-country level. The percentage column refers to the percentage of each job cluster in the sample before and after ChatGPT/Image-generating AI, respectively. Standard deviations are in the parentheses.

3.2 Google Search Volume Index Data

We gauge the evolving interest in and awareness of ChatGPT across job clusters using the Google Search Volume Index (SVI). The index is constructed by combining co-searches of ChatGPT with cluster descriptions, such as "ChatGPT writing." Thus, the co-search indices

and crypto clusters. Translation jobs have been affected by automated tools like Google Translate. Labor demand changes in blockchain, smart contracts, and crypto clusters are mainly impacted by industry downturns.

serve as a measure of interest and information intensity associated with using ChatGPT for certain tasks. Figure C1(a) presents the average search volume index (SVI) after the ChatGPT introduction for automation-prone and manual-intensive clusters, with automation-prone and manual-intensive jobs highlighted in red and blue, respectively. Figure C1(b) plots the monthly SVI over time. The figures show that the manual-intensive jobs have an almost zero SVI index throughout the sample period. In contrast, the automation-prone categories, frequently searched after the introduction of ChatGPT, experienced a significant increase.

4 Impacts of Generative AI on Online Labor Market

In this section, we analyze the short-term impact of GenAI tools on demand for different freelance jobs, using the manual-intensive cluster as the reference group based on the collective evidence from AIOE, Google SVI, and previous literature.

4.1 Empirical Strategy

As a baseline specification, we estimate the following two-way fixed-effect (TWFE) DiD model that compares the before-after difference in outcomes between job clusters:

$$y_{ctl} = \beta \text{Post}_t * \text{T}_c + \gamma_{cl} + \gamma_t + \epsilon_{ctl} \tag{1}$$

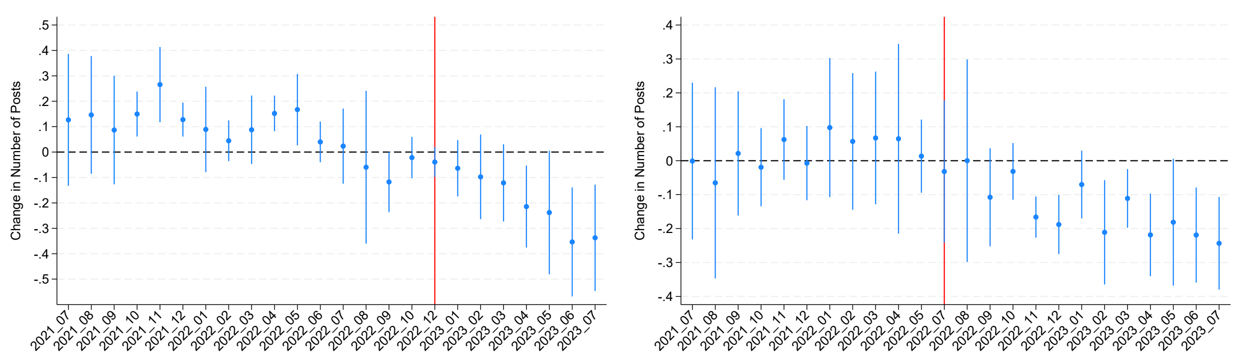
The unit of observation is a week t–country l for a given cluster c. y_{ctl} represents the outcome variable in week t in cluster c in country l. To measure the demand for freelance jobs, we operationalize y_{ctl} as the logarithm of the number of job posts. Post_t is a dummy variable that takes on a value of one following the release of GenAI tools (the week of Nov 30, 2022, for ChatGPT and the week of July 20, 2022, for Image-generating AI). T_c takes the value of zero for manual-intensive job clusters, while it takes a value of one for automation-prone job clusters in the context of ChatGPT (or for image-generating job clusters in the context of Image-generating AI). We also include country-cluster fixed effects (γ_{cl}) to control for country-cluster specific labor demand differences and week fixed effects (γ_t) to control for possible time trends and seasonality on the platform. Standard errors are clustered at the job cluster level.

To the extent that, in the absence of the AI tool introductions, the demand for freelancers evolved along parallel trends, and assuming job cluster-level average treatment effects are homogeneous across clusters and over time, the coefficient of interest β identifies the average treatment effect on the treated (ATT) of the introduction of GenAI tools on online labor market demand. To assess the validity of this assumption, we employ a difference-in-differences event-study framework:

$$y_{ctl} = \sum_{j=-2}^{T_0} \beta_j \text{Pre}_j \times T_c + \sum_{k=0}^{T_1} \beta_k \text{Post}_k \times T_c + \gamma_{cl} + \varepsilon_{ctl} \tag{2}$$

where Pre_j and Post_k is a set of indicator variables equal to 1 when an observation is j months before or k months after the release of GenAI tools (December 2022 for ChatGPT and July 2022 for Image-generating AI, respectively).[15] We plot the estimated coefficients β along with their confidence intervals in Figure 1. Panel (a) plots βs comparing automation-prone clusters and manual-intensive clusters, and Panel (b) plots βs comparing image-generating clusters and the manual-intensive clusters. Both figures show that the data are consistent with the assumption of parallel trends: the coefficients prior to the introduction of the GenAI tools (indicated by the red vertical lines) are close to zero.[16] Furthermore, following the introduction of the GenAI tools, the automation-prone and image-generating clusters began to exhibit a more pronounced decline in demand relative to the manual-intensive clusters.

Figure 1: Changes in Number of Job Posts



(a) ChatGPT **(b)** Image-generating AI

Notes: The figures plot β_k and β_j estimated from Equation 2. The red vertical line in Panel (a) marks December 2022, the month following the release of ChatGPT. In Panel (b), it marks July 2022, the month when the first Image-generating AI tools were released. Standard errors are clustered at the job cluster level.

Although TWFE regressions similar to Equation 1 are the workhorse model for evaluating causal effects, they have been shown to deliver consistent estimates only under relatively strong assumptions about homogeneity in treatment effects across treated groups

[15]For the event study, we aggregate the sample up to cluster-country-month level.

[16]A joint F-test of the β_js in the pre-period of ChatGPT yields a p-value of 0.1759, and the joint F-test of the β_js in the pre-period of Image-generating AI yields a p-value of 0.7323, not rejecting the hypothesis that they are zero.

and across time (De Chaisemartin and d'Haultfoeuille 2020, Borusyak et al. 2021, Callaway and Sant'Anna 2021, Goodman-Bacon 2021, Sun and Abraham 2021). We address concerns about the reliability of the TWFE estimator by replicating our results using the robust estimators introduced in Callaway and Sant'Anna (2021) (CS DiD) and Arkhangelsky et al. (2021) (Synthetic DiD). The CS DiD method provides a consistent estimate for ATT in DiD setups with multiple time periods and in the presence of heterogeneous treatment effects across time and/or treated units. The Synthetic DiD method uses a weighted average of outcomes from reference groups to predict the outcomes of the treated group as if the treatment did not happen. Both methods provide flexibility by relaxing the requirements of parallel pre-trends. Based on recent discussions about the log-transformation of count variables (Chen and Roth 2023), we also estimate the treatment effect using a negative binomial regression to better account for the over-dispersion in the number of job posts.

4.2 Results—Impacts of GenAI Tools

Impact of ChatGPT Introduction. We estimate our baseline and robustness specifications to examine the impact of ChatGPT released on November 30, 2022. The result for all treated groups is presented in Column (1) of Table 2. The DiD coefficient (β) in Equation 1 is significantly negative (-0.234**), which corresponds to a 20.86% decrease in the weekly number of posts in automation-prone jobs compared to manual-intensive ones. Next, we examine which specific job cluster within the automation-prone category is most impacted by ChatGPT. We estimate our DiD models separately for each cluster in the automation-prone group. The results are presented in Columns (2) to (4) of Table 2. Writing jobs exhibit the largest decrease (30.37% from the DiD model), followed by software, app, and web development (20.62%), and engineering (10.42%). Importantly, this ranking corresponds to the relative increase in SVI, our ChatGPT awareness measure, shown in Figure C1. Rows two to four present estimation results from the Negative Binomial, CS DiD, and Synthetic DiD models. The estimates from all four models are highly comparable, with only minor discrepancies observed in a few cases.[17] Our estimated effect is substantial. It is larger in magnitude than both the seasonal demand variation on the platform over time and the

[17] The difference in results for the Engineering cluster between the DiD and Negative Binomial model can be attributed to the prevalence of zeros within that cluster. In the pre-period, 48% of all observations in this cluster are equal to zero, which increased to around 55% in the post-period. This suggests a substantial decline in demand occurred at the extensive margin, better captured by a Negative Binomial model than by OLS with log-transformed dependent variables (Chen and Roth 2023).

impact of automation in traditional labor markets.[18,19]

Table 2: Changes in Demand for Freelancers after ChatGPT Introduction

	All Treated Groups	Writing	Software, App and Web Development	Engineering
DiD	-0.234**	-0.362***	-0.231**	-0.11
	(0.0837)	(0.0543)	(0.0543)	(0.0577)
Negative Binomial	-0.241***	-0.379***	-0.170***	-0.235***
	(0.0916)	(0.0666)	(0.0701)	(0.0665)
CS DiD	-0.174***	-0.233***	-0.187***	-0.1016***
	(0.0364)	(0.0183)	(0.0183)	(0.0183)
Synthetic DiD	-0.176***	-0.280***	-0.165***	-0.0798**
	(0.0271)	(0.0338)	(0.0338)	(0.0338)

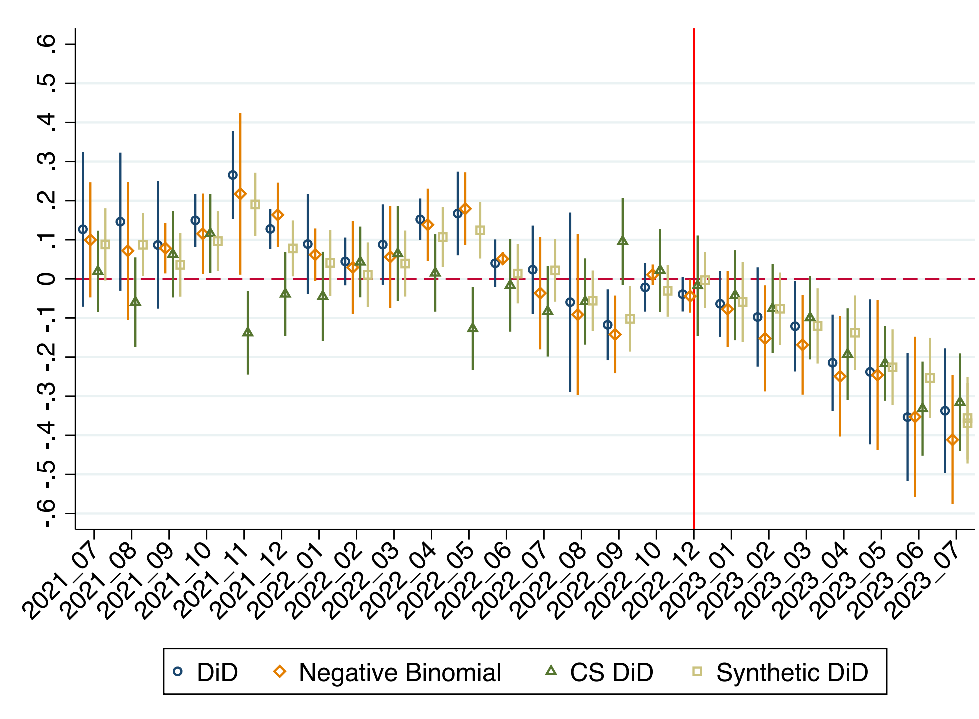
Notes: Each row corresponds to an estimation method. The first column reports the estimation results for all treated groups. The second to fourth columns report results for writing, software, app, and web development, and engineering, respectively. The number of observations is 39,528 for Column (1) and 26,352 for Columns (2) to (4). The number of job clusters is eight in the full sample. R^2 of DiD are higher than 0.85. Standard errors in parentheses are clustered at the job cluster level, and they are estimated using bootstrap for CS DiD and Synthetic DiD. *** p<0.01, ** p<0.05, * p<0.1.

Figure 2 plots the event-study figures using all four methods and shows that the data are consistent with the parallel trends assumption and the estimates align with each other.

[18] During the pre-period, the average number of automation-prone job posts is equal to 6,696.5 per week, and the standard deviation across weeks is 1,329, which is equal to 19.84% of the mean (i.e. coefficient of variation, also known as relative standard deviation).

[19] Acemoglu et al. (2020) find that a 20 percentage point increase in robot adoption within the French manufacturing sector is associated with a 3.2% decline in industry employment.

Figure 2: Event Study Estimators — Impact of ChatGPT



Notes: The figure overlays event-study plots using DiD, Negative Binomial, CS DiD, and Synthetic DiD. The bars represent 95% confidence intervals. The red vertical line marks December 2022. Standard errors are clustered at the job cluster level.

Figure 2 also shows that the decrease in demand has increased over time since the introduction of ChatGPT. In Appendix D, we estimate two sets of regressions to examine and quantify this change over time. The results, reported in Table D1, suggest that the effect not only persists but also grows, with a 10% greater decrease observed every three months post-ChatGPT. Additionally, the decrease becomes more pronounced following the introduction of more advanced versions of ChatGPT.

We also examine changes in other outcome variables, focusing on employers who posted jobs in both pre- and post-periods using Equation 1.[20] Since all DiD models deliver similar results (Table 2), we use the baseline model in this analysis. First, we examine changes in maximum budget, number of bids per job post, and the complexity of jobs (measured by the number of skill tags in the job post).[21] The results are reported in Columns (1) to (3) in Table 3. Following the introduction of ChatGPT, we observe a 5.71% increase in the maximum budget for automation-prone job clusters compared to manual-intensive jobs.

[20]We focus on this subsample (35.45% of total observations) to alleviate potential selection bias arising from employers leaving the platform due to substitution effects. In the regressions, we control for employer fixed effects.

[21]For the number of bids per job post, we only consider job posts that are open for bidding. Around 28.45% of the job posts in our sample are direct invitations to specific freelancers and hence do not have freelancers bidding on them.

Additionally, the average number of bids per job post rose by 8.57%, and job complexity increased by 2.18%. These findings suggest that after ChatGPT's release, there is a slight increase in job complexity, budget, and competition in automation-prone jobs.

Next, we analyze changes in job posting frequency among employers who remain on the platform. To measure posting frequency, we construct three metrics: (a) the number of weeks an employer posts within a cluster in the pre- and post-periods, with the sample at the employer-prepost periods level; and (b) the number of posts an employer makes per month, both conditional on posting in that month and unconditional, with the sample at the employer-month level. We apply the same DiD model to these outcome variables, comparing the differential changes between automation-prone and manual-intensive clusters. The results, reported in Columns (4) to (6) in Table 3, indicate a significantly larger decrease in posting frequency in automation-prone job clusters.

Table 3: Changes in Other Job Market Outcomes

	(1) Budget	(2) # of Bids	(3) Complexity	(4) # of Weeks	(5) # of Posts / Month (conditional)	(6) # of Posts / Month (unconditional)
$\text{Post}_t * \text{T}_c$	12.67***	0.0822**	0.103***	-0.692**	-0.191***	-0.00543**
	(2.987)	(0.0229)	(0.0157)	(0.206)	(0.0352)	(0.00205)
Observations	296,368	211,740	296,368	97,257	180,427	6,552,000
R-squared	0.423	0.498	0.479	0.048	0.006	0.013
Pre-Mean	221.66	3.37	4.74	1.89	1.49	0.02
Percentage Change (%)	5.71	8.57	2.18	-36.64	-12.82	-27.16
Employer FE	Yes	Yes	Yes	No	No	No
Cluster FE	Yes	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	No	No	No
Month FE	No	No	No	No	Yes	Yes
Post-period FE	No	No	No	Yes	No	No
Observation Level	Job Posts	Job Posts	Job Posts	Employer-PrePost	Employer-Month	Employer-Month

Notes: This table reports estimation results of Equation 1 for other outcome variables. Budget refers to the maximum budget (USD) of the job post. The number of bids is logged. Complexity is measured using the number of skill tags of a job post. # of weeks refers to the number of weeks an employer posts jobs in the pre and post periods. Standard errors in parentheses are clustered at the job cluster level. *** p<0.01, ** p<0.05, * p<0.1.

Lastly, we examine the skill tags to identify posts explicitly mentioning "ChatGPT" as a required skill. In the post-period, we find 903 job posts listing ChatGPT in the skill tags. Notably, more than 88% of these jobs fall into automation-prone categories, with the majority (744 job posts) specifically related to Software, App and Web Development. We then regress the number of job posts with the "ChatGPT" skill tag on a time variable that counts the number of weeks since the release of ChatGPT for the post-period. The results indicate a significant increase in the number of job posts requiring ChatGPT-related skills, with an average rise of 0.68 posts per week over time following ChatGPT's introduction.

Impact of Image-generating AI Introduction. In this subsection, we examine the effects of Image-generating AI technologies on demand for freelancer jobs in graphic design and 3D modeling clusters, using the baseline specification in Equation 1 and the robust DiD models. Specifically, we focus on three major Image-generating AI technologies, DALL-E 2, Midjourney, and Stable Diffusion, introduced between July and September 2022 (Figure A1). The release date for each of these technologies differs by a few weeks, and we assign the earliest public release as the treatment time. Therefore, Post_t is equal to one for weeks after July 20th, 2022. This specification ensures that effects from each of these Image-generating technologies are captured. The reference group is the manual-intensive clusters.

Table 4 presents the estimation results for Image-generating AI technologies. Column (1) shows a significant decrease in the number of job posts related to image creation compared to manual-intensive jobs. Specifically, within a year of the introduction of Image-generating AI, the number of job posts for graphic design and 3D modeling decreased by 17.01%. This effect is again larger than the seasonality variation on the platform or the effect of automation on traditional labor markets (Acemoglu et al. 2020).[22] The remaining rows report the estimation results from the Negative Binomial, CS DiD, and Synthetic DiD models, respectively. Each alternative model gives significant and comparable results to each other and provides further evidence for the robustness of the main effect. Since the post period in this regression includes the introduction of ChatGPT, we further restrict the post period to the period until the ChatGPT introduction date (November 2022). Column (2) provides the estimation results for this restricted period. In line with Column (1), it indicates a 12.90% larger decrease in the number of job posts for image creation.[23]

Columns (3) to (6) in Table 4 focus on the graphic design and 3D modeling clusters separately. The estimates from the baseline DiD regression in the first row indicate an 18.47% decline in the number of job posts for graphic design (Column (3)) and 15.52% for 3D modeling (Column (5)). Results from other estimation methods and the sample restricted to the "Pre-ChatGPT" period yield consistent findings.

[22] During the pre-period, the average number of image-generating job posts is equal to 4,158.8 per week, and the standard deviation across weeks is 514.4, which is equal to 12.37% of the mean (i.e. coefficient of variation).

[23] The magnitude of the coefficient in Column (2) is marginally smaller than that in Column (1). This is likely due to the gradual adoption of Image-generating AI technologies over time. With a longer post-period, we observe a bigger impact.

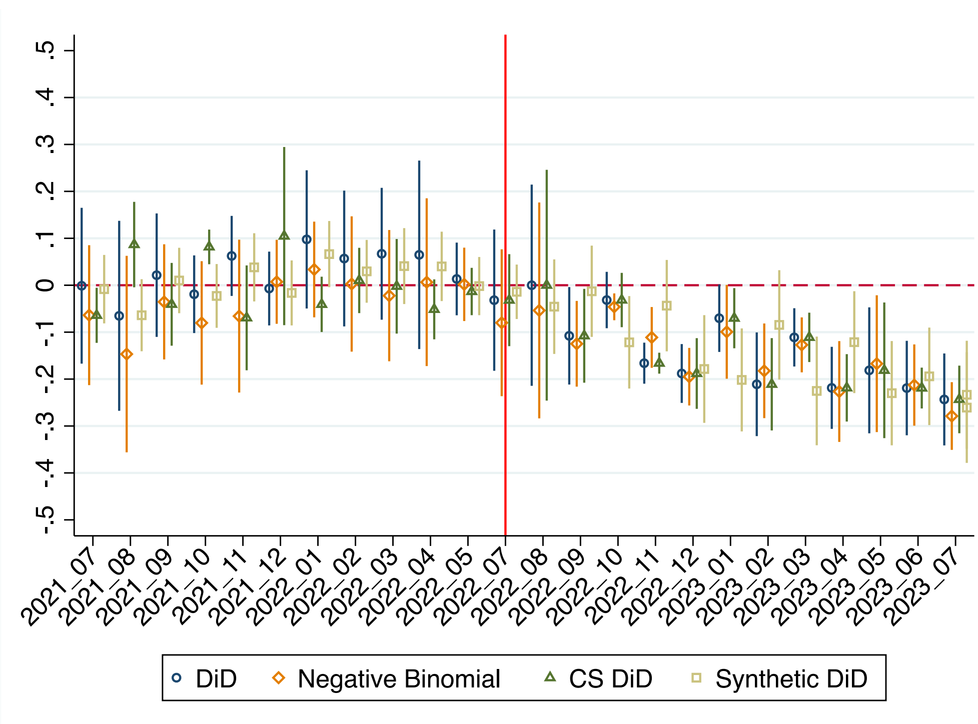
Table 4: Changes in Demand for Freelancers after Image-Generating AI Technology

	All Treated Groups		Graphic Design		3D Modeling	
	Entire period	Pre-ChatGPT	Entire period	Pre-ChatGPT	Entire period	Pre-ChatGPT
DiD	-0.1864**	-0.1381**	-0.2042**	-0.1677***	-0.1687**	-0.1083**
	(0.0488)	(0.042)	(0.0484)	(0.036)	(0.0484)	(0.0361)
Negative Binomial	-0.1244***	-0.0869***	-0.1232***	-0.1025***	-0.1319***	-0.0627***
	(0.0411)	(0.0186)	(0.0427)	(0.0111)	(0.0392)	(0.0125)
CS DiD	-0.1077*	-0.0577	-0.187***	-0.150***	-0.028	0.034
	(0.0615)	(0.077)	(0.0251)	(0.04088)	(0.0251)	(0.0408)
Synthetic DiD	-0.178***	-0.121***	-0.176***	-0.139***	-0.180***	-0.103***
	(0.0297)	(0.0335)	(0.0303)	(0.0312)	(0.0303)	(0.031)

Notes: Each row corresponds to an estimation method. The first two columns report estimation results for all treated groups compared to manual-intensive job clusters. The remaining columns report results for graphic design and 3D modeling, respectively. In columns labeled "Pre-ChatGPT," the post period is restricted to before the introduction of ChatGPT (November 2022), while in other columns, the post period spans from July 2022 to July 2023. The total number of observations is 32,940 in Column (1) and 22,265 in Column (2). Columns (3) and (5) have 26,352 observations, and Columns (4) and (6) have 17,812 observations. The number of job clusters is five in the full sample. R^2 of DiD are higher than 0.85. Standard errors in parentheses are clustered at the job cluster level and estimated using bootstrap for CS DiD and Synthetic DiD. *** p<0.01, ** p<0.05, * p<0.1.

Figure 3 plots the estimates from the event study analysis using all four models. The figure supports the assumption of parallel trends, showing a consistent decline in job posts related to image generation across all models.

Figure 3: Event Study Estimators — Impact of Image-generating AI



Notes: The figure shows event-study plots using DiD, Negative Binomial, CS DiD, and Synthetic DiD. The bars represent 95% confidence intervals. The red vertical line marks July 2022. Standard errors are clustered at the job cluster level.

Robustness Checks and Placebo Tests. We conduct robustness analyses and placebo tests to confirm that our results capture the substitution effects of GenAI tools.

First, we show that the variation in interest and awareness of using ChatGPT across job categories, proxied by Google SVI (Figure C1), predicts the incremental decline of demand in automation-prone jobs. We estimate the following specification, where SVI_{ct} is the weekly Google SVIs across job clusters:

$$y_{ctl} = \beta SVI_{ct} * \text{Post}_t + \gamma_{cl} + \gamma_t + \epsilon_{ctl} \tag{3}$$

The results of the regression are presented in Figure E1. Panel (a) shows the estimated baseline SVI effect, $\hat{\beta}\text{SVI}_{ct}$, plotted against Google SVI, and Panel (b) presents estimation results. Both panels highlight a significantly negative relationship between Google SVI and the short-term change in the number of job posts. An increase of one standard deviation in SVI corresponds to an 8.01% decrease in job posts.[24] This implies that job categories experiencing increased interest in using ChatGPT also experienced a more notable decline in demand for freelancers.

In Appendix E, we conduct several robustness checks. These include using alternative reference groups (e.g., audio and video job clusters only, an expanded reference group that includes other clusters not used in our main analysis, and local jobs), employing a more aggregated sample at the week-cluster level, and considering hourly-paid jobs. Our results are robust across all of these checks. Additionally, we conduct placebo tests by assigning "placebo" treatment time and find insignificant estimates in both the ChatGPT and Image-generating AI analyses.

5 Concluding Remarks

This paper documents the short-term impact of GenAI technologies on demand in the online labor market. Using data from a global freelancer platform, we quantify a 21% greater decline in demand for automation-prone jobs compared to manual-intensive jobs after ChatGPT introduction. Writing is the job category most affected by ChatGPT, followed by software, app and web development, and engineering. We also find a 17% more pronounced decrease in demand for graphic design and 3D modeling jobs following the release of Image-generating AI technologies. Our findings also suggest that freelancers with certain skills may face more competition after the introduction of GenAI tools. Given the already intense competition for job opportunities in online labor markets (Beerepoot and Lambregts 2015), the increased

[24] In other words, a 1% increase in SVI is associated with a 0.404% decrease in the number of job posts.

substitutability between freelancer jobs and GenAI could further decrease earnings in the short term.

GenAI's early impact on online labor markets suggests possible changes in broader labor dynamics. With the increasing adoption of GenAI, there will likely be a significant shift in the types of skills in demand. Tasks that can be easily automated by AI, such as routine and repetitive tasks, are expected to decline in demand. In contrast, as our findings suggest, demand for new skills may emerge for effectively incorporating GenAI tools into job tasks. Skills that complement AI, such as critical thinking, creativity, and emotional intelligence, may become more valuable and in demand. This shift in skill demand may lead to more pronounced differences across the labor market, with a growing divide between high-skill, high-wage jobs and low-skill, low-wage jobs.

Due to the time frame of technological shocks in our study, we focus on the short-term impact of AI on employment. However, in the long run, there might still be net job growth as a result of AI, potentially attributed to productivity effects and reinstatement effects. Some early studies show potential productivity benefits: a large-scale controlled trial found that consultants using GPT-4 have access to 12.2% more tasks, complete them 25.1% faster, and produce 40% higher quality results than those without the tool (Dell'Acqua et al. 2023). Another experimental evidence finds that ChatGPT decreased the time required for business writing work by 40%, with output quality rising by 18% (Noy and Zhang 2023). Our finding about the emergence of job posts specifically seeking "skills using ChatGPT", primarily in automation-prone jobs, also indicates early trends in new skills and job creation.

As a response to these changes, we are likely to observe dynamic adaptations in the labor market. Workers and firms need to invest in education, training, and technological innovation to remain competitive in an AI-impacted economy. This will require targeted upskilling and reskilling initiatives consistent with broader automation and labor market studies. Policymakers may need to promote equitable access to education and training opportunities, support displaced workers through social safety nets and reemployment programs, and encourage innovation and entrepreneurship to harness the benefits of GenAI while mitigating its adverse effects.

GenAI will also greatly impact managerial decision-making. Our findings highlight the need to consider the potential impacts of AI on various aspects of business operations. Managers should identify tasks that are more susceptible to automation and recognize roles requiring AI-complementary skills. Firms can leverage GenAI to improve efficiency and innovation but should simultaneously develop strategies to support workers transitioning from automated roles. Establishing partnerships with training institutions and online learning platforms can facilitate continuous learning opportunities. By taking these specific actions, companies can better prepare their workforce for the future, ensuring organizational resilience and employee growth in an AI-impacted economy.

References

Acemoglu D, Autor D (2011) Skills, tasks and technologies: Implications for employment and earnings. *Handbook of labor economics*, volume 4, 1043–1171 (Elsevier).

Acemoglu D, Lelarge C, Restrepo P (2020) Competing with robots: Firm-level evidence from france. *AEA papers and proceedings*, volume 110, 383–388 (American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203).

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

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

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

Agrawal A, Gans JS, Goldfarb A (2019) Artificial intelligence: the ambiguous labor market impact of automating prediction. *Journal of Economic Perspectives* 33(2):31–50.

Agrawal A, Horton J, Lacetera N, Lyons E (2015) Digitization and the contract labor market: A research agenda. *Economic analysis of the digital economy* 219–250.

Arkhangelsky D, Athey S, Hirshberg DA, Imbens GW, Wager S (2021) Synthetic difference-in-differences. *American Economic Review* 111(12):4088–4118.

Beerepoot N, Lambregts B (2015) Competition in online job marketplaces: towards a global labour market for outsourcing services? *Global Networks* 15(2):236–255.

Blocklove J, Garg S, Karri R, Pearce H (2023) Chip-chat: Challenges and opportunities in conversational hardware design. *2023 ACM/IEEE 5th Workshop on Machine Learning for CAD (MLCAD)*, 1–6 (IEEE).

Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* 2008(10):P10008.

Borusyak K, Jaravel X, Spiess J (2021) Revisiting event study designs: Robust and efficient estimation. *arXiv preprint arXiv:2108.12419* .

Brand J, Israeli A, Ngwe D (2023) Using gpt for market research. *Available at SSRN 4395751* .

Brynjolfsson E, Li D, Raymond LR (2023) Generative ai at work. Technical report, National Bureau of Economic Research.

Brynjolfsson E, Rock D, Syverson C (2018) Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics. *The economics of artificial intelligence: An agenda*, 23–57 (University of Chicago Press).

Bucaioni A, Ekedahl H, Helander V, Nguyen PT (2024) Programming with chatgpt: How far can we go? *Machine Learning with Applications* 15:100526.

Burtch G, Lee D, Chen Z (2023) The consequences of generative ai for ugc and online community engagement. *Available at SSRN 4521754* .

Callaway B, Sant'Anna PH (2021) Difference-in-differences with multiple time periods. *Journal of econometrics* 225(2):200–230.

Chen J, Roth J (2023) Logs with zeros? some problems and solutions. *The Quarterly Journal of Economics* .

Coello CEA, Alimam MN, Kouatly R (2024) Effectiveness of chatgpt in coding: A comparative analysis of popular large language models. *Digital* 4(1):114–125.

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

del Rio-Chanona M, Laurentsyeva N, Wachs J (2023) Are large language models a threat to digital public goods? evidence from activity on stack overflow. *arXiv preprint arXiv:2307.07367* .

Dell'Acqua F, McFowland E, Mollick ER, Lifshitz-Assaf H, Kellogg K, Rajendran S, Krayer L, Candelon F, Lakhani KR (2023) Navigating the jagged technological frontier: Field experimental evidence of the effects of ai on knowledge worker productivity and quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper* (24-013).

Dwivedi YK, Kshetri N, Hughes L, Slade EL, Jeyaraj A, Kar AK, Baabdullah AM, Koohang A, Raghavan V, Ahuja M, et al. (2023) "so what if chatgpt wrote it?" multidisciplinary perspectives on opportunities, challenges and implications of generative conversational ai for research, practice and policy. *International Journal of Information Management* 71:102642.

Eisfeldt AL, Schubert G, Zhang MB (2023) Generative ai and firm values. Technical report, National Bureau of Economic Research.

Eloundou T, Manning S, Mishkin P, Rock D (2023) Gpts are gpts: An early look at the labor market impact potential of large language models. *arXiv preprint arXiv:2303.10130* .

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* .

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

Graham M, Anwar MA (2019) The global gig economy: Towards a planetary labour market? *First Monday* 24(4), URL http://dx.doi.org/10.5210/fm.v24i4.9913.

Hui X, Reshef O, Zhou L (2023) The short-term effects of generative artificial intelligence on employment: Evidence from an online labor market. *Available at SSRN 4527336* .

Knight S, Bart Y (2023) Generative ai and user-generated content: Evidence from online reviews. *Available at SSRN 4621982* .

Kässi O, Lehdonvirta V (2018) Online labour index: Measuring the online gig economy for policy and research. *Technological Forecasting and Social Change* 137:241–248, ISSN 0040-1625, URL http://dx.doi.org/https://doi.org/10.1016/j.techfore.2018.07.056.

Kässi O, Lehdonvirta V, Stephany F (2021) How many online workers are there in the world? a data-driven assessment. *Open Research Europe* 1.

Liu J, Xu X, Li Y, Tan Y (2023) " generate" the future of work through ai: Empirical evidence from online labor markets. *arXiv preprint arXiv:2308.05201* .

Lukac M (2021) Two worlds of online labour markets: Exploring segmentation using finite mixture models and a network of skill co-occurrence URL http://dx.doi.org/10.31235/osf.io/m7ah4.

Newman MEJ (2006) Modularity and community structure in networks. *Proceedings of the National Academy of Sciences* 103(23):8577–8582, URL http://dx.doi.org/10.1073/pnas.0601602103.

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

Pallais A (2014) Inefficient hiring in entry-level labor markets. *The American Economic Review* 104(11):3565–3599, ISSN 00028282.

Peng S, Kalliamvakou E, Cihon P, Demirer M (2023) The impact of ai on developer productivity: Evidence from github copilot. *arXiv preprint arXiv:2302.06590* .

Qiao D, Rui H, Xiong Q (2023) Ai and jobs: Has the inflection point arrived? evidence from an online labor platform. *arXiv preprint arXiv:2312.04180* .

Sant'Anna PH, Zhao J (2020) Doubly robust difference-in-differences estimators. *Journal of Econometrics* 219(1):101–122.

Shan G, Qiu L (2023) Examining the impact of generative ai on users' voluntary knowledge contribution: Evidence from a natural experiment on stack overflow. *Available at SSRN 4462976* .

Sun L, Abraham S (2021) Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics* 225(2):175–199.

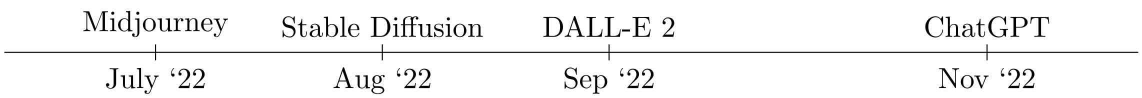
Wiles E, Horton JJ (2023) The impact of ai writing assistance on job posts and the supply of jobs on an online labor market .

Yilmaz ED, Naumovska I, Aggarwal VA (2023) Ai-driven labor substitution: Evidence from google translate and chatgpt. *Available at SSRN* .

Online Appendix

A GenAI Background

Figure A1: The Timeline of Release Dates for Different GenAI Technologies



Notes: The figure presents publicly release dates for Midjourney, Stable Diffusion, Dall-E 2, and ChatGPT. Information is obtained through the providers' official websites.

B Louvain Clustering Method and Sample Construction

The Louvain clustering method is an unsupervised algorithm used to identify communities or clusters within a network. The algorithm iteratively optimizes the partitioning of nodes into communities based on the density of connections within and between them, ultimately revealing cohesive groups of nodes with higher intra-community connectivity compared to inter-community connections.[25] The method involves two phases: first, nodes are iteratively moved to the community that results in the maximum increase in modularity.[26] Second, the network is coarsened by aggregating all nodes of a community together into one node, thus creating a new network. This second step reduces the complexity of the network while preserving the community structure found in the first phase. The two phases are performed iteratively until the maximum modularity is reached.

In our application, we consider all job posts to be constituting a complex hidden network composed of clusters that share similar skill requirements. Therefore, the skills become nodes, and the co-occurrence of skills in the job posts becomes edges. We aim to identify "communities" of skills (clusters) from the entire pool of posts based on the co-occurrence of skills. Specifically, similar to Lukac (2021), we build a skill co-occurrence network that reflects joint occurrences of required skills across job posts. Our network is represented by an association matrix A_{is} where

$$A_{is} = \begin{cases} 1 & \text{if job post } i \text{ requires skill } s \\ 0 & \text{otherwise} \end{cases}$$

[25] Nodes represent individual entities.

[26] In the context of network analysis, modularity is a measure that quantifies the relative density of edges (i.e., the ties between nodes) inside communities with respect to edges outside communities. It can be used as an objective to optimize in the context of community detection (Newman 2006, Blondel et al. 2008) to find the best possible grouping of nodes in a given network.

We construct the skill co-occurrence network by multiplying the association matrix A_{is} by its transpose: $N = A_{is}^{\intercal} A_{is}$. The resulting network N is a square matrix in which both rows and columns represent a skill. Thus, each element N_{qj} indicates how many times skill q and skill j are jointly required for a job post. The clustering method takes the matrix N as an input and identifies a unimodal network that is composed of 42 clusters. We then map each job post to a cluster with the largest overlap in skills. For example, if a job post includes three skill tags, and two of them belong to cluster A while one belongs to cluster B, we assign this job post to cluster A since the majority of its skills fall into that cluster.[27] This assignment ensures that each job post belongs to a single cluster, which facilitates the aggregation of our sample.[28] We name the clusters based on the skill tags they contain.

Finally, we proceed through the following steps to ensure the representativeness of the sample: (1) we keep the jobs that are the most prevalent on the platform. We exclude job posts belonging to less prevalent clusters (below 0.12%). This step drops 0.249% of job posts and results in 18 major clusters. The excluded clusters relate to niche job categories, such as Cartography, Amazon FBA, Fundraising, or Digital Forensics. In the remaining clusters, we merge three clusters that involve similar skills into one cluster. These three clusters mainly require programming and coding, specifically related to Software, Mobile Application, and Web Development. (2) We exclude job posts with maximum budgets in the top 1% and restrict to the 61 largest countries in the sample. The cleaning process results in a sample of 1,388,711 job posts belonging to 15 clusters (Table C1). For our main empirical analysis, we focus on 8 clusters with 1,218,463 job posts described in Section 3.

[27] The mean and median of the number of clusters per job post are 1.58 and 1.

[28] Among the 42 clusters identified by the Louvain algorithm, three of them do not have any project where a majority of skills belong to those clusters.

C More Details about the Sample

Table C1: Cluster Summary Statistics

Job Cluster	Total Number of Posts	Percentage of Total Posts	Mean Log Number of Posts	SD Log Number of Posts
3D Modelling	78,437	5.65 %	1.65	1.15
Accounting and Finance	10,308	0.74 %	0.49	0.76
Audio Services	13,120	0.94 %	0.61	0.80
Blockchain, Smart Contracts and Crypto	10,987	0.79 %	0.55	0.77
Data and Office Management	119,350	8.59 %	2.00	1.17
Engineering	29,009	2.09 %	1.02	0.99
Graphic Design	319,367	23.00 %	2.87	1.20
Legal	6,278	0.45 %	0.32	0.64
Internet Marketing	76,826	5.53 %	1.64	1.12
Social Media Marketing	25,119	1.81 %	0.92	0.93
Software, App and Web Development	483,898	34.85 %	3.47	1.11
Statistical Analysis	8,651	0.62 %	0.45	0.71
Translation	32,079	2.31 %	1.18	0.98
Video Services	44,035	3.17 %	1.24	1.04
Writing	131,247	9.45 %	2.07	1.22

Notes: This table presents the total number of job posts in each cluster throughout our sample period (Column 1) and their percentage in the sample (Column 2). Columns 3 and 4 summarize our main variable of interest, which is the logarithmized number of job posts aggregated at the week-cluster-country level.

Table C2: Job Clusters and their Most Frequent Skill Tags

Cluster	Most Frequent Skill Tags
3D Modelling	3D Modelling, 3D Rendering, AutoCAD, 3D Animation, Building Architecture, CAD/CAM, 3ds Max, Interior Design, 3D Design, Solidworks
Audio Services	Audio Services, Audio Production, Voice Talent, Music, Sound Design, Voice Artist, Voice Over, Audio Editing, Video Services, English (US) Translator
Data and Office Management	Data Entry, Excel, Data Processing, Web Search, Web Scraping, Copy Typing, Virtual Assistant, Word, PDF, Visual Basic
Engineering	Electrical Engineering, Electronics, Engineering, Microcontroller, Matlab and Mathematica, Arduino, Mathematics, PCB Layout, Circuit Design, C Programming
Graphic Design	Graphic Design, Photoshop, Logo Design, Illustrator, Website Design, Photoshop Design, WordPress, Illustration
Software, App and Web Development	PHP, HTML, Website Design, JavaScript, Software Architecture, Mobile App Development, MySQL, WordPress, Android, CSS
Video Services	Video Services, Video Editing, After Effects, Video Production, Animation, Videography, 3D Animation, Graphic Design, YouTube, 2D Animation
Writing	Article Writing, Content Writing, Research Writing, Copywriting, Article Rewriting, Ghostwriting, Report Writing, Technical Writing, Research, Blog

Notes: This table presents the most frequent skill tags from the job posts in each cluster used in our analysis.

Table C3: Job Clusters and Corresponding AIOE Index

Cluster Labels	Occupation Title	Language Modeling AIOE
Data and Office Management	Data Entry Keyers	0.172
Audio Services	Sound Engineering Technicians	0.338
Video Services	Film and Video Editors	0.657
Software, App and Web Development	Software Developers, Applications	0.882
Engineering	Electrical Engineers	0.901
Writing	Writers and Authors	1.170

Notes: This table presents the AIOE index for the six job clusters related to manual-intensive and automation-prone types. We manually map the job clusters with the AIOE index, associating each cluster with the "Occupation Title" in the AIOE database that it most closely relates to.

Table C4: Summary Statistics for Main Outcome Variables

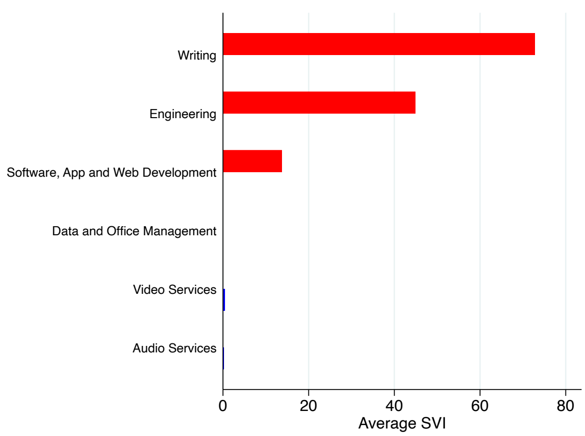
	Mean	SD	Median
Weekly Number of Job Posts	11,811.97	2,468.40	11,462.00
Maximum Budget (in USD)	337.17	596.23	168.31
Number of Bids per Job Post	26.43	36.29	13.00
Number of Skill Tags per Job Post	4.52	1.61	5.00

Notes: This table reports the summary statistics of the main outcome variables from our sample before aggregation. For rows 2 to 4, one unit of observation is a job post. The maximum budget is adjusted using country-specific inflation rates. The number of skill tags is used as a proxy for the complexity of the jobs.

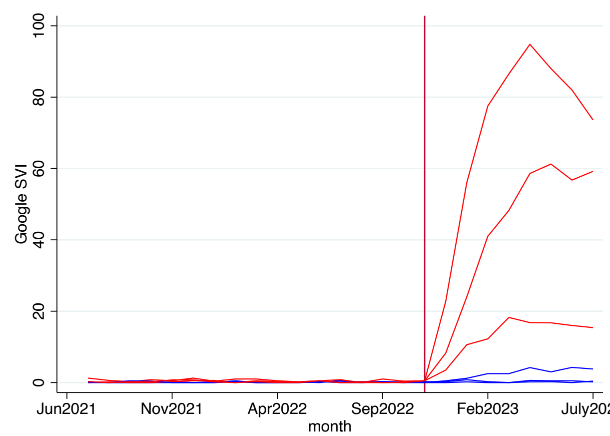
Figure C1 plots the average and monthly time trend of Google SVI. Google only allows for a comparison across five search terms at a time and normalizes the results relative to the highest value. Hence, during data collection, we conducted multiple queries while keeping the highest value search term constant (i.e. ChatGPT writing). The SVI for software, app, and web development is calculated as the sum of three individual SVI indices (software development, app development, and web development).

Figure C1: Google Trends SVI

(a) Average (Post-ChatGPT Period)



(b) Monthly Time Trend



Notes: Panel (a) plots the average Google Trends SVI over the months following the introduction of ChatGPT for the automation-prone (in red color) and manual-intensive (in blue color) clusters, and Panel (b) plots the monthly Google Trends SVI for each cluster. In Panel (b), the time lines from top to bottom are writing, engineering, software, app and web development, data and office management, video services, and audio services. The red vertical line marks December 2022.

D Evolution of the Effect Over Time

We investigate the effect of ChatGPT over time using two approaches. First, we interact dummies for 1 to 3 months post-ChatGPT, 3 to 6 months post-ChatGPT, and 6 to 8 months post-ChatGPT with dummies for the treated units. Second, we decompose the longer-run effects based on different versions of ChatGPT, using three post dummies interacted with the treated dummies.[29] We then estimate separate DiD regressions corresponding to these specifications. The results, presented in Table D1, align well with Figure 1. The coefficients become increasingly negative over time and with the release of more advanced versions of ChatGPT. These findings suggest that the effect not only persists but also grows over time.

Table D1: Evolution of the Effect Over Time — ChatGPT

	(1) Within Three Months	(2) Three to Six Months	(3) Six to Eight Months	(4) ChatGPT 3.5	(5) ChatGPT Plus	(6) ChatGPT 4
$Post_t \times T_c$	-0.145**	-0.239*	-0.364**	-0.130*	-0.186**	-0.295**
	(0.0563)	(0.0970)	(0.104)	(0.0505)	(0.0678)	(0.104)
Observations	31,476	32,574	30,012	30,012	29,280	34,404
R-squared	0.888	0.887	0.887	0.889	0.889	0.885
Week FEs	Yes	Yes	Yes	Yes	Yes	Yes
Cluster-Country FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: In Column (4), $Post_t$ is set to 1 if an observation falls between the week of November 31, 2022, and the week of February 1, 2023. In Column (5), $Post_t$ is set to 1 if an observation is between the week of February 1, 2023, and March 14, 2023. In Column (6), $Post_t$ is set to 1 if an observation is between the week of March 14, 2023, and July 1, 2023. Standard errors in parentheses are clustered at the job cluster level. *** p<0.01, ** p<0.05, * p<0.1.

E Robustness Checks and Placebo

In this section, we conduct a series of robustness checks using Equation 1. First, we examine demand changes in job clusters not included in our main ChatGPT analysis—namely legal, accounting and finance, social media marketing, internet marketing, and statistical analysis—relative to manual-intensive jobs. The estimated $\hat{\beta}$ is both statistically insignificant and of small magnitude (0.0272). This suggests that the more substantial decrease in demand is unique to the automation-prone categories, providing further evidence that automation-prone jobs are the most affected.

Second, we conduct robustness checks using alternative reference groups (Table E1). The first column reports results using audio and video services as the reference group, considering that ChatGPT may improve at data entry tasks over our post-period. The second column uses an expanded reference group that includes manual-intensive job clusters as well as clusters not utilized in our main analysis, such as legal, accounting and finance, social media marketing, internet marketing, and statistical analysis. In column (3), we run the regression with "local jobs" requiring physical presence, which comprise 1.06% of our sample, as the reference group, and find a slightly larger decrease. In addition to the results presented in Table E1, we conduct two robustness checks. First, we run an analysis focusing specifically on hourly-paid jobs on the platform and obtain a similar result (-0.150*). Second, we use a sample aggregated across countries at the cluster-week level and obtain an estimate $\hat{\beta}$ of -0.2909**.

[29]On February 1, 2023, OpenAI announced the launch of ChatGPT Plus, and on March 14, 2023, OpenAI released GPT-4 within the ChatGPT platform. https://www.searchenginejournal.com/history-of-chatgpt-timeline/488370/.

Table E1: Changes in Demand for Freelancers after ChatGPT Introduction (Robustness)

	(1) Audio and Video	(2) Expanded Reference	(3) Local Jobs
$\text{Post}_t * \text{T}_c$	-0.292**	-0.250***	-0.371**
	(0.0668)	(0.0660)	(0.0690)
Observations	32,940	72,468	26,244
R-squared	0.894	0.879	0.906
Week FEs	Yes	Yes	Yes
Cluster-Country FEs	Yes	Yes	Yes

Notes: Estimation results of Equation 1 using alternative reference groups. *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the job cluster level in parentheses.

Table E2 presents robustness checks for Image-generating AI, structured in the same way as Table E1. First, we estimate demand changes using audio and video jobs as the reference group in Column (1). In Column (2), we run a robustness test using the expanded reference group. Lastly, we estimate our regression using "local jobs" as the reference group in Column (3). The results are all consistent with the main results in Table 4.

Table E2: Changes in Demand for Freelancers after Image-generating AI Introduction (Robustness)

	(1) Audio and Video	(2) Expanded Reference	(3) Local Jobs
$\text{Post}_t * \text{T}_c$	-0.234***	-0.196***	-0.313***
	(0.0294)	(0.0259)	(0.0154)
Observations	26,352	65,880	19,656
R-squared	0.879	0.865	0.896
Week FEs	Yes	Yes	Yes
Cluster-Country FEs	Yes	Yes	Yes

Notes: Estimation results of Equation 1 using alternative reference groups. *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the job cluster level in parentheses.

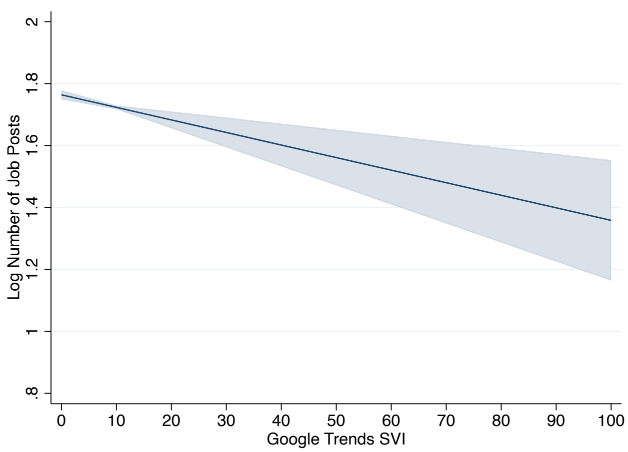
Finally, we conduct a series of placebo tests to ensure our results are not influenced by spurious correlations in the data. For ChatGPT analysis, we assign a placebo treatment in November 2021, one year before its introduction. The post-period is December 2021 to July 2022, and the pre-period is July 2021 to November 2021. The coefficient is insignificant (-0.068), indicating that the decrease in automation-prone jobs is unique to the period after ChatGPT's introduction. For Image-generating AI analysis, we perform a similar placebo test, assigning a treatment in January 2022, with the post-period from January 2022 to July 2022. The coefficient is also insignificant (-0.005).[30]

[30] We perform another placebo test by setting the post-period as January 2022 to April 2022 to avoid contamination of the treatment effect by earlier, limited versions of Image-generating GenAI tools. The estimated coefficient is also statistically insignificant (0.0325).

F Analysis using Google SVI

Figure E1: Google Trends SVI and Changes in Number of Job Posts

(a) Plot of Marginal Effect



(b) Estimation Results

	Log Number of Posts
$SVI_{ct} * \text{Post}_t$	-0.00405**
	(0.00107)
Observations	39,528
R-squared	0.885
Week FEs	Yes
Cluster-Country FEs	Yes

Notes: The figure plots the estimated marginal Google SVI effect ($\hat{\beta} SVI_{ct}$) reported in the right-hand-side table, with the corresponding 95% confidence interval. *** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered at the job cluster level are in parentheses.