"Generate" the Future of Work through AI: Empirical Evidence from Online Labor Markets

Jin Liu*

School of Management, University of Science and Technology of China, liujin07@mail.ustc.edu.cn

Xingchen (Cedric) Xu*

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

Xi Nan

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

Yongjun Li

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

Yong Tan

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

Large Language Model (LLM) based generative AI, such as ChatGPT, is considered the first generation of Artificial General Intelligence (AGI), exhibiting zero-shot learning abilities for a wide variety of downstream tasks. Due to its general-purpose and emergent nature, its impact on labor dynamics becomes complex and difficult to anticipate. Leveraging an extensive dataset from a prominent online labor market, we uncover a post-ChatGPT decline in labor demand, supply, and transactions for submarkets pertaining to textrelated and programming-related jobs, in comparison to those not directly exposed to ChatGPT's core functionalities. Meanwhile, these affected submarkets exhibit a discernible increase in the complexity of the remaining jobs and a heightened level of competition among freelancers. Intriguingly, our findings indicate that the diminution in the labor supply pertaining to programming is comparatively less pronounced, a phenomenon ascribed to the transition of freelancers previously engaged in text-related tasks now bidding for programming-related opportunities. Although the per-period job diversity freelancers apply for tends to be more limited, those who successfully navigate skill transitions from text to programming demonstrate greater resilience to ChatGPT's overall market contraction impact. As AI becomes increasingly versatile and potent, our paper offers crucial insights into AI's influence on labor markets and individuals' reactions, underscoring the necessity for proactive interventions to address the challenges and opportunities presented by this transformative technology.

Key words: Generative AI, Large Language Models, Artificial General Intelligence, ChatGPT, Online Labor Markets, Gig Economy, Labor Dynamics, Skill Transition

"It is not the most intellectual of the species that survives; it is not the strongest that survives; but the species that survives is the one that is able best to adapt and adjust to the changing environment in which it finds itself." - Megginson (1963)

1. Introduction

Online labor markets, particularly gig work platforms, have emerged as a pivotal component of the global labor landscape. According to the World Bank, as of 2023, the gig economy constitutes up to 12% of the global labor market and holds particular promise for vulnerable groups in developing nations¹. The work-from-home and work-from-anywhere trends have continued to proliferate during and after the pandemic. As per Upwork's report, freelancers contributed \$1.35 trillion to the U.S. economy in annual earnings in 2022, representing a \$50 billion increase from the previous year, with 60 million American freelancers participating in this burgeoning sector². Given their substantial impact on the economy, freelancer platforms have garnered significant attention from economic and information systems (IS) researchers (Liang et al. 2024, 2022, Benson et al. 2020, Horton 2019).

Nevertheless, the burgeoning trends in the online labor market may have undergone a profound transformation following the public release of ChatGPT on November 30, 2022. Conceptually, we posit that ChatGPT and other LLM-based Generative AI possess distinctive technical characteristics that distinguish them from prior technologies. These characteristics serve as unique antecedents and may introduce undetermined and potentially unparalleled consequences in online labor markets for the following reasons.

Focusing on the technological aspects (the *antecedents* in our context), large language models are distinguished by an unprecedented number of parameters and are trained on vast troves of data, endowing them with capabilities that far surpass those of their predecessors. Although there is no single consensus on what constitutes a "large" language model, we adopt the general-purpose

^{*} These authors contributed equally to the manuscript and are listed alphabetically.

¹ Please refer to https://openknowledge.worldbank.org/handle/10986/40066 for additional details.

² Please refer to https://www.businesswire.com/news/home/20221213005068/en/ for additional details.

criteria that the language models should be sufficiently large to achieve zero-shot learning ability on a wide range of tasks for which the model has not been specifically trained (Kojima et al. 2022). Prior to the development of such large language models, fine-tuning was the sole means by which a model, such as BERT, could be adapted to perform specific tasks (Brown et al. 2020). Consequently, past pre-trained models necessitated engineers to fine-tune the model and develop corresponding applications with clearly defined functions, which were then made available to users. However, the general-purpose abilities of LLMs enable the built-on generative AI application, exemplified by ChatGPT, to directly interact with end users in natural language for a wide variety of tasks, the scope of which is not deliberately defined by designers (Kosinski 2023). Moreover, a recently proposed taxonomy by Google DeepMind classifies LLM-based generative AI, such as ChatGPT and Llama 2, as the first generation of artificial general intelligence (AGI) due to their generalpurpose nature and exhibition of metacognitive capabilities (Morris et al. 2023). Recent research has uncovered emergent phenomena in LLMs, rendering their behavior unpredictable even to their designers: for example, generative AI can make rational deductions and exhibit personalities (Mei et al. 2024, Chen et al. 2023), but may also display hallucinations and be unsuitable for tasks requiring very high accuracy (Xu et al. 2024, Rawte et al. 2023). The combination of generalpurpose capabilities and emergent properties makes LLM-based generative AI differentiated, and their potential impact challenging to predict.

The impact of LLM-based generative AI on online labor markets (the *consequence* in our context) is multifaceted and challenging to deduce from literature without empirical investigations. Figure 1 summarizes the main actions related to online labor markets, with *potentially direct* influence mechanisms highlighted by red lines. The first salient challenge lies in the ambiguity surrounding which jobs can be fully automated, assisted, or remain unaffected by LLM-based generative AI, given its general-purpose and emergent properties (Morris et al. 2023). For a specific category of work, if it can be fully automated by LLM-based generative AI, potential clients may satisfy their demand using AI directly, eliminating the need to post job requirements in online labor

markets (red line 1). In cases where AI does not provide complete automation but instead aids humans in task completion, it can enable already skilled individuals to finish jobs more efficiently in terms of speed and/or quality, and/or increase the pool of people capable of performing the job. These effects can manifest on both the demand and supply sides: On the demand side (also red line 1), if service prices in the market remain constant, more potential clients might refrain from engaging in online labor markets due to improving outside options (self-completion); On the supply side, if service prices in the market remain constant, more people may have increased incentives or the ability to participate in online labor markets (red line 2). Past empirical studies on AI and labor typically have a clear boundary on the designed functions of a specific technology (Gallego and Kurer 2022), and sometimes even know who has access to it for the enterprise usage scenarios (Jia et al. 2024). However, this clarity is no longer present in the era of general-purpose AI, where the boundaries and access to technology are more fluid and uncertain.

The second challenge arises from the nature of online labor markets, where labor supply and demand form the matching process and are regulated by expected outcomes from this process (Liang et al. 2022). Freelancers who participate must strategically select which jobs to bid on, determine appropriate prices, and compete against each other. Given the pool of bidding freelancers, clients make choices based on prices and their expectations of job completion quality, with the option to select no one. If a freelancer is chosen, completion efficiency may also be directly affected (red line 4), as reports indicate that 20% of freelancers have regularly used Generative AI tools for work³. Beyond this final step, if Generative AI alters the pool of available jobs and participating freelancers, it can indirectly impact competition and prices (both bidding and transaction) through the matching process (Hong et al. 2016).

Thirdly, the general-purpose nature of LLM-based Generative AI and its potential to impact a wide range of tasks may not only enable individuals to perform a specific job or do so more efficiently but also provide existing freelancers with opportunities to undertake jobs across different

³ See https://www.upwork.com/research/freelance-forward-2023-research-report for additional details.

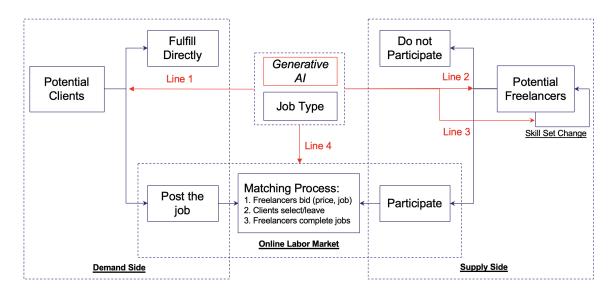


Figure 1 Research Framework.

categories and potentially facilitate transitions within the freelancer ecosystem (red line 3). This skill transition process can influence the competition and matching landscape, as freelancers with newly acquired or enhanced abilities may enter new job categories, altering the supply-demand dynamics. Conversely, the skill transition process can also be affected by market conditions, such as the relative demand for different types of jobs and the prevailing wage rates in various categories.

Given the importance of online labor markets and the complex dynamics introduced by LLM-based generative AI, as previously outlined, we aim to uncover the comprehensive influence of generative AI on online labor markets by investigating the following research questions:

- 1. How does the adoption of LLM-based generative AI affect labor demand, supply, and matching outcomes in online labor markets?
- 2. To what extent does LLM-based generative AI influence competition, job complexity, and market efficiency within online labor markets?
- 3. How do the effects of LLM-based generative AI vary across different job categories in online labor markets?
- 4. What impact does LLM-based generative AI have on the participation of new freelancers, the skill transitions of existing freelancers, and their wage outcomes in online labor markets?

To tackle these research questions, we harness a comprehensive dataset encompassing the entirety of job posts, bids, and final transactions, along with freelancers' information from one of the most prominent freelancer platforms, spanning from September 2021 to August 2023. We regard the launch of ChatGPT as an exogenous shock to online labor markets due to its abrupt introduction and largely unforeseen triumph. Given the versatile and emergent nature of ChatGPT, predetermining the degree to which jobs are impacted is challenging and constitutes the crux of this research. Rather than assigning specific exposure weights for each job category to ChatGPT, we employ a more conservative and exploratory approach. We first segment the entire online labor market into a series of submarkets, within each of which require the same set of skill categories. As ChatGPT's primary function is to generate text, including code, we consider all the submarkets which involve programming-related and/or text-related skills as the treated units. To circumvent the influence of alternative shocks, such as the introduction of stable-diffusion and other image generative AI technologies during the observational window, we exclude image-related submarkets from our main analyses and assess their impact independently. We utilize the remaining submarkets that are unrelated to programming, text, and image as the control group, such as fully manual local jobs like manufacturing. After ensuring parallel trends across a wide array of variables, we further substantiate the validity of this natural experiment opportunity, empowering us to investigate the differential impact of ChatGPT on various facets of online labor markets and address our research questions.

Employing the standard DiD specification, complemented by additional econometric analyses at both the freelancer and market levels, we uncover a series of intriguing findings. First, compared to the control group, the labor demand for programming-related and text-related submarkets decreases significantly in terms of both the number of jobs posted and the number of participating clients. A downward trend also emerges on the labor supply side, especially the number of job applications (bids). With the contraction on both sides, the number of transactions and transactional values achieved for these jobs also decline. Second, we observe that the remaining

programming-related and text-related jobs on these affected submarkets tend to be more intricate, as indicated by higher initial budgets set by clients. However, due to the intensified competition among freelancers, measured by the average number of bids per project, the final transactional price remains stable. Moreover, the reduced number of available projects and the heightened competition among freelancers do not guarantee higher efficiency, as the matching rate does not increase significantly, and the average completion time does not exhibit a substantial decrease. Third, we discover intriguing heterogeneities among different submarkets. The programming-related submarkets incur a significantly smaller labor supply decrease compared to text-related (programming-free) submarkets, while there is no significant heterogeneity in labor demand change. This phenomenon opens up several potential explanations for freelancer behavior dynamics: i) ChatGPT enhances the programming efficiency of existing freelancers, encouraging them to remain active. ii) Chat-GPT lowers the barrier for other freelancers to engage in programming, allowing them to transition and compensate for the loss in programming labor supply. iii) ChatGPT reduces the programming barrier or increases the efficiency of individuals outside the online labor markets, leading to an influx of new programming freelancers. We provide related evidence when addressing the fourth research question. Overall, we observe a significant but homogeneous decrease in the number of new freelancers joining the treated submarkets. Simultaneously, freelancers who previously bid on programming-related jobs exhibit an even greater reduction in activity. Conversely, a substantial number of freelancers who were previously engaged in text-related jobs are now actively bidding on programming tasks. This skill transition enables these freelancers to remain resilient in the face of ChatGPT's launch, lending credence to the second explanation for supply-side heterogeneity. As a final remark, we also examine the labor dynamics for image-related submarkets and find no significant differences compared with the control submarkets. This outcome may be due to the limited capabilities of text-to-image generative AI, which does not align well with the requirements of many jobs.

Our paper makes significant contributions to several strands of literature. First, we advance the understanding of AI's impact on labor markets by providing comprehensive evidence on how Chat-GPT, the first generation of AGI, shapes the demand, supply, and matching process of online labor

markets. We also document intriguing heterogeneities when a single technology can affect a wide range of jobs on both sides, thereby enriching the discourse on the multifaceted nature of this technological disruption. Second, we furnish novel insights on online labor markets by illustrating how a general-purpose technology influences market dynamics and induces freelancers' responses, including skill transition, thus shedding light on the adaptability and resilience of market participants in the face of technological change. Third, our economic analyses of LLM-based generative AI provide a more comprehensive understanding of AI's societal impact, thereby underscoring the pressing need for more AI social alignment research as AI systems become increasingly sophisticated and powerful.

In the subsequent sections of this paper, we synthesize pertinent literature and underscore our theoretical contributions in Section 2. Section 3 explicates the empirical setting, data acquisition, and variable construction approach. We summarize the identification challenges, selection of econometric specification, and the comprehensive analyses plan in Section 4. Section 5 presents our estimates on the average treatment effect at the market level. Section 6 showcases our extended examinations at both the market and freelancer levels. Section 7 offers robustness checks. Finally, Section 8 provides a conclusion with practical implications and future research directions outlined.

2. Related Literature and Theoretical Contributions

Our research aligns with and enriches three streams of literature: i) Online Labor Markets; ii) The Impact of AI on Labor Markets and Productivity; iii) AI Alignment and Large Language Models.

2.1. Online Labor Markets

Contextually, our paper is situated within the realm of online labor markets, which have demonstrated a remarkable ability to effectively reconcile labor supply and demand across time and space with unparalleled flexibility (Chen, 2019). A wealth of scholarly inquiries has been conducted to elucidate the complex dynamics of these digital labor marketplaces. These investigations encompass the interplay between online labor markets and external socioeconomic fluctuations, the intricate relationship between user behavior patterns and market design, and the advancements in algorithms designed to optimize the matchmaking processes.

The first line of inquiry delineates how conditions in the offline environment, such as unemployment rates and national developmental stages, catalyze individual engagement in online labor markets (Huang et al. 2020, Kanat et al. 2018). Concurrently, related studies also scrutinize the manner in which online labor markets reciprocally influence the external environment, focusing on areas like entrepreneurship, housing, and transportation (Zhang et al. 2022, Burtch et al. 2018, Cramer and Krueger 2016). Secondly, historical research has also centered on modeling user behavior and formulating improved market designs within online labor markets. For instance, certain investigations elucidate how employers dynamically learn strategies for hiring suitable workers (Kokkodis and Ransbotham 2023, Leung 2018). Additional research assesses how variables such as reputation (Benson et al. 2020), capacity (Horton 2019), and wage regulations (Chen and Horton 2016) impact market equilibrium and social welfare in online labor markets, thereby proposing and executing superior market designs (Garg and Johari 2021). Pertaining to the third strand of literature, optimization algorithms have been continuously honed to facilitate job recommendations and streamline matchmaking processes within the online labor markets (Aouad and Saban 2023, Kokkodis and Ipeirotis 2023).

Our research primarily contributes to the first stream of literature by empirically examining the transformative impact of ChatGPT, widely regarded as the first-generation AGI (Morris et al. 2023), on the dynamics of online labor markets. Traditionally, it has been feasible to identify a specific category within online labor markets that might be affected by a certain technological advancement, such as the influence of Google Translate on translation-related tasks (Yilmaz et al. 2023). However, the emergence of LLM-based generative AI, characterized by its general-purpose nature and emergent capabilities, presents a unique challenge in predicting its far-reaching effects across a wide range of tasks on both labor supply and demand. Our study represents a pioneering endeavor to delve into the granular details of how this groundbreaking technology reshapes various categories and aspects of online labor markets. Notably, we uncover heterogeneous effects on the supply of labor in programming-related tasks compared to text-related tasks. Furthermore, our research provides a comprehensive analysis of the skill transition process undertaken by freelancers in response to the advent of this disruptive technology.

2.2. The Impact of AI on Labor Markets and Productivity

Our research also contributes to the growing body of literature exploring the multifaceted ways in which AI is reshaping labor markets and productivity, considering AI as a potent tool capable of proficiently accomplishing specific tasks inherent in human professions (Agrawal et al. 2019). Firstly, AI might serve as a direct substitute for human workers, particularly those engaged in routine tasks susceptible to full automation by AI (Gallego and Kurer 2022, Paolillo et al. 2022). Secondly, in certain occupations, AI may augment or refine human labor, thereby bolstering productivity and quality, functioning as a complementary asset to human labor (Felten et al. 2021, Staccioli and Virgillito 2021). Thirdly, AI has the potential to engender novel employment opportunities where human labor is necessitated for the creation, maintenance, or leverage of AI to accomplish tasks previously beyond human capabilities (Acemoglu et al. 2022).

Existing literature has extensively explored the impact of AI on various aspects of the economy; however, the majority of empirical research has focused on AI either at a general, macro level (McElheran et al. 2024, Alekseeva et al. 2021) or on the usage of a specific AI by a predetermined group of individuals, such as those in customer service (Jia et al. 2024, Brynjolfsson et al. 2023). Despite the extensive research in this field, there is a lack of empirical evidence concerning the underlying mechanisms beyond the ultimate employment outcomes for an AI that can directly impact both labor demand and supply across various job categories through diverse channels.

Our study contributes to this stream of literature by offering a comprehensive examination of the impact of ChatGPT, an AGI, on various aspects of the labor market. We investigate ChatGPT's influence on labor supply, demand, competition, and efficiency changes at the market level, as well as the heterogeneous skill transitions and wage outcomes experienced by freelancers at the individual level. By providing a holistic view of the effects of LLM-based Generative AI on the labor market, our research aims to bridge the gap in the existing literature and offer valuable insights into the complex dynamics at play.

2.3. AI Alignment and Large Language Models

Our research directly engages with the AI alignment literature, particularly in the context of LLMs. As AI technology progresses and transforms key sectors, the concept of AI alignment has become increasingly significant (Gabriel 2020). This principle focuses on ensuring that AI systems adhere to human intentions and values throughout their lifecycle, including development, evaluation, and deployment (Ji et al. 2023). Furthermore, post-deployment considerations involve a social alignment perspective, advocating for AI systems that not only meet the objectives of their designers and direct users, but also resonate with broader societal values (Korinek and Balwit 2022). This broader application also allows us to contextualize previous research on the economics of AI within the AI alignment framework (Agrawal et al. 2019).

The gradual progression towards AGI and the increasing prevalence of interactions among agents, both human and AI, introduce additional complexities in AI alignment (Kirk et al. 2024). A recent taxonomy proposed by Google DeepMind categorizes AI systems into ten classes based on their application scope (narrow or wide) and performance across five tiers, ranging from rule-based to superhuman AI (Morris et al. 2023). Traditional AI systems are typically specialized and fall under the narrow category. In contrast, LLMs such as ChatGPT exhibit broader metacognitive abilities, representing an initial foray into the wide category (i.e., AGI) (Morris et al. 2023). The unpredictable nature of AGI's impacts, stemming from its metacognitive abilities, presents challenges in forecasting the affected domains, such as social communities (Xue et al. 2023) and software development (Peng et al. 2023). Furthermore, interactions between multiple agents can lead to unforeseen outcomes (Xu et al. 2023, Korinek and Balwit 2022). As the adoption of AI by humans increases and autonomous systems interact more frequently within markets and society, complex macro phenomena may emerge (De Marzo et al. 2023).

Our research is deeply embedded within this body of literature as it grapples with both intricacies inherent in AI alignment. Firstly, we endeavor to elucidate the impact of ChatGPT, which may permeate a multitude of job categories without distinct boundaries. Secondly, in the context of

online labor markets, participants on both sides, spanning a wide array of categories, possess the potential to adopt AI and interact with others, thereby introducing additional layers of complexity. Our work makes a significant contribution to this stream of literature by illuminating the market-level and individual-level dynamics within online labor markets that are precipitated by ChatGPT, thereby enriching our understanding of the social implications of LLM-based generative AI and providing insights for alignment considerations.

3. Research Context, Data, and Variables

To systematically and comprehensively address our research questions, we gather data from a leading freelancer platform and construct variables for subsequent market-level and freelancer-level analyses. Subsection 3.1 elucidates the platform's fundamental information, our selection of the observation window, the raw dataset, the job classification process, and the formation of submarkets. In subsection 3.2, we delineate the establishment of treatment and control groups at both levels of investigation. Subsection 3.3 discusses the development of variables of interest.

3.1. Research Context, Job Classifications, and Submarket Construction

Our data is derived from a preeminent online labor platform that facilitates connections between freelancers and clients across a diverse array of jobs, such as programming, writing, and architecture. This platform enables clients to post jobs (i.e., projects), encompassing specific job content, budgetary constraints, and the essential skills sought for the successful completion of the project. Upon reviewing these posted requirements, freelancers can submit their bids, articulating their capabilities and providing their respective quotations. Subsequently, clients exercise the prerogative to select from the pool of bidders, finalizing the transaction with their chosen freelancers (For a detailed elucidation of the process, kindly refer to Appendix A).

To investigate the impact of LLM-based generative AI, with a particular emphasis on ChatGPT, our observation window must encompass the launch date of ChatGPT, which occurred on November 30, 2022. To ensure sufficiently long observation periods before and after the launch time, while mitigating the confounding end-of-year effects, we establish a two-year observation window

spanning from September 2021 to August 2023. Within this period, we collect all job listings, associated job bids, and the data of freelancers who placed at least one bid on these projects. It is important to acknowledge that Dall-E, Midjourney, and several other text-to-image generative AI systems also released their publicly accessible versions during this time frame. In Sections 3.2, 6.3, and Appendix H, we discuss the construction of image-related treatment variables, the process of isolating their effect from the main analyses, and the estimation of their effects.

Upon the publication of a job listing on the platform, pertinent skill tags are provided to facilitate efficient search and filtering processes. These tags are subsequently leveraged for job classification and the construction of specialized submarkets. To begin with, the platform's front page features 12 popular skill categories, such as software development, accompanied by exemplary tags that serve to illustrate the nature of each category (e.g., C++ software development). However, it is important to note that the platform does not maintain an exhaustive list of all job tags associated with these popular categories, and moreover, not all job tags fall within the purview of these 12 categories. Building upon the platform's existing skill categories and skill tags, we employ a snowball process to expand the categorization to a total of 29 categories. Each of the 2,719 tags that appear on the platform is assigned to one of these categories (Full category lists and tag examples are provided in Appendix C). Furthermore, given that a single job listing may be associated with multiple tags, we utilize the aforementioned category-tag dictionary to ascertain the category or categories to which each job belongs. Leveraging this information, we can identify jobs that share similar skill set requirements (i.e., the same skill category combination) and subsequently form submarkets where labor supply and demand are matched. After eliminating extremely inactive submarkets, we arrive at a total of 873 submarkets, each of which features a minimum of one job posting per month on average. During our observation window, these submarkets collectively encompass approximately 1.6 million job postings and attract 42 million bids (job applications). The aforementioned classification and submarket construction processes are elucidated in greater detail within Appendix B.

3.2. Treatment Group, Control Group, and Treatment Timing

In the preceding subsection, we have categorized each of the 2,719 skill tags into one of the 29 categories and further allocated each of the approximately 1.6 million jobs to one of the 873 submarkets. Utilizing these categories and submarkets, we now establish the treatment and control groups at both the market and freelancer levels.

As per our definitions in Section 1, ChatGPT spearheads the first generation of AGI and serves as the primary focus of our analyses. It is important to note that ChatGPT is general-purpose, and its potential downstream applications are extensive, even surpassing designers' expectations. Consequently, it may be inappropriate to predefine different weights for the exposure levels of various job categories to ChatGPT. Given that the core function of ChatGPT is to generate texts, including code, we consider all text-related or programming-related skill categories as categories potentially affected by ChatGPT.

Moreover, during the same observation period, another significant technological advancement is image-to-text generative AI. Although the functions of these AI systems are specifically limited to generating images (a multi-model application) and do not constitute AGI, the image quality reaches unprecedented levels and garners considerable attention. To account for their potential effect, we isolate them as image-related categories for further analysis. The remaining skill categories encompass architecture, manufacturing, logistics, etc. These categories are not directly associated with either LLM-based text generative AI (ChatGPT) or image-to-text generative AI, thereby serving as categories reserved for controls.

It is essential to recollect that a single job may encompass multiple skill tags and, consequently, multiple skill categories, necessitating the utilization of the previously constructed submarkets. To minimize the direct exposure of the control group submarkets to generative AI, we exclusively employ submarkets in which all included skills are classified as control categories. Furthermore, to mitigate the potential confounding influence of text-to-image generative AI when examining the impact of ChatGPT, we label submarkets containing at least one image-related skill with

 $Treat_{image} = 1$ and exclude them from the main analyses. After omitting the image-related and control submarkets, the remaining submarkets constitute the treatment group for the main analyses, denoted by the variable Treat = 1. This designation signifies that all skills within these submarkets are not image-related, and concurrently, each submarket contains at least one skill pertaining to either text or programming. When investigating the heterogeneous impact of the treatment, we further introduce the variable Programming to indicate whether the submarket comprises any programming-related skills. In Appendix C, we provide an overview of the relationships between skill category combination, the submarket type, and the treatment variable.

Regarding the treatment timing, it is crucial to acknowledge that numerous models and their associated applications were released to the public during the observational window. As such, we utilize the release date of ChatGPT, the first LLM-based text generative AI, which occurred on November 30, 2022, as the treatment timing. Given that this date falls at the end of November, when conducting month-level analyses, we consider December 2022 as the initial period post-treatment $(Post = 1)^4$. This principle is similarly applied to text-to-image generative AI, with the first several public beta versions being released in July 2022, specifically Midjourney on July 12, 2022, and Dall-E on July 20, 2022. When considering the week-level analyses, we directly utilize the week that encompasses the release date of the first model as the first post-treatment period.

Lastly, we also construct treatment variables Treat at the freelancer level to facilitate our freelancer-level analyses, which aim to investigate heterogeneous responses and wage changes among freelancers and elucidate the underlying mechanism. We confine the data sample to individuals who place at least one bid in November, 2022. Adhering to the same principle, we categorize freelancers who solely bid on jobs within the "control submarkets" prior to the shock as "control freelancers" (Treat = 0), while those who also placed bids within the "treated submarkets" are classified as "treated freelancers" (Treat = 1). This results in a sample of 138,169 active freelancers, enabling us to examine the differential changes in their behavior following the launch of ChatGPT.

⁴ Our analyses are robust to using November 2022 as the first period, and the effects become more salient over time.

3.3. Variable Descriptions

In this section, we introduce the primary variables of interest for our analysis and elucidate the process of constructing these variables. Our market-level variables are all constructed at the submarket-month level. To characterize labor demand, supply, and matching outcomes, we construct the following variables: On the demand side, our analysis revolves around two main variables: i) Jobs, signifying the number of jobs posted within a specific submarket in a given month, and ii) Clients, denoting the number of unique clients who post jobs within the submarket during a given month. On the supply side, we focus on two indicators: i) Bids, encapsulating the total number of bids submitted to jobs within a submarket in a given month, and ii) Freelancers, representing the number of distinct freelancers who have placed at least one bid within a particular submarket during a month. To examine the matching outcomes, we emphasize two fundamental variables: i) Matches, epitomizing the number of successful matches made (i.e., successfully completed jobs) within a submarket in a given month, and ii) Transaction, encapsulating the associated total transaction values for these matches.

In addition to the primary outcomes of interest, we construct several supplementary variables to evaluate market conditions, such as competition, job complexity, and efficiency. Changes in job complexity are partially reflected by the average job budget (Budget), calculated by taking the mean of all job budgets posted in a submarket during a particular month. To measure changes in market competition, we employ two variables: i) the average number of bids per job (Avg_bids) and ii) the final transactional price (Match_price). These variables are also aggregated at the monthly level for each submarket. To assess market efficiency, we calculate two variables on a monthly basis for each submarket: i) the matching rate (Match_rate), which is the ratio of the number of successfully completed jobs to the total number of posted jobs, and ii) the average job completion time (Complete_time), derived by subtracting the job assignment time from the payment time for each completed job and then taking the mean of this duration. To examine the participation of new freelancers in the market, we introduce another variable: the number of newly registered freelancers (New_freelancers) who bid in a submarket in a given month.

Analogous to the market-level variables, we also construct several freelancer-level main outcomes, including the monthly number of bids $(Bids_{freelancer})$, matches (i.e., number of completed jobs $Matches_{freelancer})$, and transactional values $(Transaction_{freelancer})$ for each individual freelancer.

Due to the left-skewed nature of the data distribution, we apply a log transformation to variables in the subsequent analysis, with the exception of $Match_rate$, which ranges between 0 and 1. Appendix D presents the descriptive statistics of all these variables.

4. Identification Challenges and Econometric Specifications

To rigorously investigate the impact of LLM-based generative AI on online labor markets, several identification challenges must be discussed prior to conducting further analyses.

The primary challenge arises from the potential interactions among market participants. As is common in research involving markets, units can influence each other through various mechanisms, such as competition and selection (Jensen and Miller 2018). In online labor markets, freelancers have the autonomy to select jobs and compete with one another (Liang et al. 2022), rendering the comprehensive capture of market dynamics a methodological conundrum, even with randomized experiments (Barach et al. 2020). Developing innovative statistical methods for analyzing matching markets is a promising avenue for future methodological research. However, given the currently available tools and the aspiration to expand the frontiers of our knowledge, we employ the following strategies: i) Rather than solely focusing on freelancer-level final transactional outcomes, we construct a series of submarkets to elucidate temporal dynamics in online labor markets, as detailed in Sections 5 and 6.1. By utilizing market-level data, we can directly examine heterogeneous changes on the demand side, which is mainly on-demand and suffers less from cross-unit interactions (Hong et al. 2016). ii) However, the presence of freelancers who are active for relatively long periods and can transition across different submarkets raises concerns regarding the labor supply side analyses. Consequently, we exercise meticulous care in interpreting the effects. It is crucial to emphasize that our objective is not to capture the hypothetical labor supply change in the absence of freelancer transition. Instead, at the market level, we capture the relative change between submarkets whose required skills overlap with the functionality of ChatGPT and those that do not, allowing for transitions as a potential mechanism. iii) Furthermore, to uncover the detailed transition process on the supply side, we complement our main analyses by documenting freelancers' skill transitions and other strategic behaviors at the individual level in Section 6.2.

Another concern stems from the fact that all time-variant observables are potentially influenced by LLM-based generative AI, regardless of whether they exist within or outside the online labor markets. The inclusion of such post-treatment variables as controls can obstruct certain mechanisms and bias the overall effect estimation (Tafti and Shmueli 2020), a phenomenon known as the bad control problem (Cinelli et al. 2022). Therefore, we treat all observables as outcome variables and examine their changes, instead of presuming some to be exogenous control variables. In alignment with established practices (Braghieri et al. 2022), our primary specifications incorporate two-way fixed effects without the introduction of time-variant controls:

$$Outcome_{it} = \beta_1 * Treat_i * Post_t + u_i + T_t + \epsilon_{it}$$

$$\tag{1}$$

$$Outcome_{it} = \beta_1 * Treat_i * Post_t + \beta_2 * Post_t + u_i + Y_t + M_t + \epsilon_{it}$$
(2)

Here, we denote the outcome of interest for each unit i (submarket or freelancer) during month t as $Outcome_{it}$. The binary variable $Treat_i$ is defined to be 1 if unit i belongs to the treatment group, and 0 otherwise. Analogously, the binary variable $Post_t$ is assigned the value 1 if month t occurs after November 2022, and 0 otherwise. Furthermore, we incorporate u_i for the unit-level fixed effect and T_t for the year-month-level fixed effect. In an alternative specification, we also employ Y_t and M_t to denote year fixed effect and month fixed effect, respectively, to aggregate the seasonal effect. The error term ϵ_{it} encapsulates unobserved factors that influence the outcome.

5. Main Analyses

In this section, we address the first research question (RQ1) by examining the overall impact of LLM-based generative AI on the demand, supply, and matching outcomes in the online labor market. Specifically, we aim to determine how submarkets requiring skills related to ChatGPT's functionalities (text and code generation, hereafter referred to as treated submarkets) exhibit distinct changes compared to those that are not directly exposed to such capabilities.

5.1. The Impact on the Demand Side

We commence our analysis by examining the impact of ChatGPT's introduction on labor demand, with the results presented in Table 1. Columns (1) and (3) present the estimates with year-month level fixed effects, while columns (2) and (4) show the results using separate year and month level fixed effects, which aggregate seasonal effects but allow for more degrees of freedom. The consistency of the results across different specifications and outcome variables is remarkable ($\beta_1 = -0.239 \text{ or } -0.240, p < 0.01$), indicating a significant 24% decline in job postings and participating clients within treatment submarkets compared to control submarkets following the introduction of ChatGPT.

Table 1 The Impact of Generative AI on Labor Demand Volume

	ln(Jo	bs+1)	ln(Clients+1)		
	(1)	(2)	(3)	(4)	
Treat*Post	-0.2390***	-0.2390***	-0.2400***	-0.2400***	
	(0.0592)	(0.0592)	(0.0593)	(0.0593)	
Constant	2.5540***	2.4660***	2.4720***	2.3880***	
	(0.0216)	(0.0097)	(0.0217)	(0.0090)	
Category FE	Yes	Yes	Yes	Yes	
Year FE	No	Yes	No	Yes	
Month FE	No	Yes	No	Yes	
Year-Month FE	Yes	No	Yes	No	
Observations	17,247	17,247	17,247	17,247	
R^2	0.9090	0.9080	0.9160	0.9150	

Robust standard errors in parentheses *** p < 0.01, ** p < 0.05, * p < 0.1

The results presented above might stem from multiple mechanisms. For the direct impact of ChatGPT on the demand side (red line 1 in Figure 1), two paths might come into play. Firstly, ChatGPT possesses the potential to fully automate tasks that previously necessitated human labor. Moreover, it can empower individuals who were previously incapable of performing specific tasks to accomplish them independently or enable skilled individuals to complete tasks with heightened efficiency. Both of these mechanisms can directly reduce potential clients' need to engage freelancers' labor services. Given that ChatGPT's primary functions involve text generation, including code, the demand for submarkets whose skill requirements are associated with such functions experiences

a relative decline. One may question that this is a market where demand can also be shaped by supply. However, conceptually, as an additional tool to enhance job efficiency, ChatGPT is unlikely to exert a direct negative impact on the supply side. As a result, the potential for ChatGPT to initially affect supply negatively and subsequently for the shortage to transcend to the demand side is limited. Moreover, the skill requirements for a particular task are predetermined naturally before considering how to fulfill it, whether via online labor markets or outside options. This external determination constrains interactions among disparate submarkets' demand. Consequently, our results are less susceptible to indirect influence from the supply side and cross-submarket contamination, and can be primarily attributed to the direct demand substitution by ChatGPT. However, due to our approach's exploratory nature and the aggregation of multiple submarkets, the specific submarkets that predominantly drive this negative influence remain uncertain, motivating our heterogeneous effects estimation in Section 6.1.2.

5.2. The Impact on the Supply Side

We then assess the impact on the supply side of the market employing a similar specification but with different outcome variables: the number of job applications (Bids) and the number of active freelancers (Freelancers). The estimation results, reported in Table 2, document a significant decrease in job applications ($\beta_1 = -0.142$, p < 0.05) and a marginal decrease in the number of active freelancers ($\beta_1 = -0.0858$, p < 0.1) in treated submarkets compared to control submarkets.

ln(Bids+1)ln(Freelancers + 1)(2)(3)(4)(1)Treat * Post-0.1420** -0.1420** -0.0858* -0.0858* (0.0598)(0.0597)(0.0514)(0.0514)4.7200*** 4.6380*** 4.4540*** 4.3850*** Constant (0.0218)(0.0227)(0.0188)(0.0229)Category FE Yes Yes Yes Yes Year FE No Yes No Yes No Yes Month FE Yes No Year-Month FE YesNo YesNo Observations 17,247 17,247 17,247 17,247 R^2 0.81600.81600.77900.7790

Table 2 The Impact of Generative AI on Labor Supply Volume

Robust standard errors in parentheses *** p < 0.01, ** p < 0.05, * p < 0.1

The observed decrease in labor supply cannot be regarded as the direct impact of ChatGPT, as theoretically, ChatGPT shall enhance the efficiency of existing freelancers or enable more freelancers to participate in the market (red line 2 in Figure 1). If this were the sole mechanism at play and its effect was not pronounced, one would expect to observe a null effect rather than the negative effect that was actually observed. Consequently, this decline is more plausibly explained as a supply-side response to the contraction in demand, which may manifest in two ways: i) existing freelancers choosing to opt out of participation, and/or ii) a reduction in the number of new freelancers joining the platform. These two mechanisms will be examined in detail in Sections 6.2.1 for existing freelancers and 6.1.3 for new entrants, respectively.

Unlike the labor demand side, which is less susceptible to direct cross-unit interactions, free-lancers on the supply side, particularly full-time freelancers who maintain a long-term presence on the platform, have the ability to migrate between different submarkets in response to market condition dynamics. This mobility implies that the estimated effects on labor supply encompass both the hypothetical change that would occur in the absence of transitions and additional alterations resulting from such movements. The former scenario arises when each freelancer evaluates fixed submarkets solely against external job opportunities, without considering the possibility of shifting between submarkets. This theoretically "clean" effect does not capture the full dynamics instigated by ChatGPT on the supply side. Consequently, we acknowledge freelancers' migrations as a natural mechanism influencing supply differences across various submarkets. We delve deeper into these migrations by assessing the heterogeneous impacts on diverse submarkets in Section 6.1.2 and exploring the detailed freelancer-level transition process in Section 6.2.2.

5.3. The Impact on Matching Outcomes

In the previous section, we have demonstrated a decrease in both labor supply and demand. However, the overall economic significance of these changes remains uncertain. To explore this, we examine two variables: the number of successful matches (i.e., successfully completed jobs, *Matches*) and the total transaction values associated with these matches (*Transaction*). As shown in Table 3, both the number of successful matches and the transaction values in the programming-related and text-related submarkets exhibit a significant decline compared to the control submarkets following the launch of ChatGPT ($\beta_1 = -0.13$ or -0.394, p < 0.01).

Table 5 The impact of Generative Af on Matching Outcomes					
	ln(Mate	ches + 1)	ln(Transaction + 1)		
	(1)	(2)	(3)	(4)	
$\overline{Treat*Post}$	-0.1300***	-0.1300***	-0.3940***	-0.3940***	
	(0.0384)	(0.0383)	(0.1130)	(0.1130)	
Constant	1.3620***	1.3360***	3.7170***	3.6250***	
	(0.0140)	(0.0103)	(0.0412)	(0.0453)	
Category FE	Yes	Yes	Yes	Yes	
Year FE	No	Yes	No	Yes	
Month FE	No	Yes	No	Yes	
Year-Month FE	Yes	No	Yes	No	
Observations	17,247	17,247	17,247	17,247	
R^2	0.8940	0.8940	0.6520	0.6520	

Table 3 The Impact of Generative AI on Matching Outcomes

Robust standard errors in parentheses *** p < 0.01, ** p < 0.05, * p < 0.1

The effect is not only statistically significant but also economically meaningful. Given that the average monthly transaction volume for each submarket is approximately \$2,400 and the estimated relative decrease in transaction values is 39.40%, each AI-exposed submarket experiences a relative change of more than \$11,000 per year. Given the presence of hundreds of submarkets constructed in our analyses, this can translate to a million-level transactional value change for the focal freelancer platform. Although the estimated values cannot directly represent the "loss" for the entire platform, they serve as a strong indication of the transformative economic impact of ChatGPT on online labor markets. To gain a comprehensive understanding of the changes in different parts of the platform and different aspects of the matching process, we embark on a series of extended analyses in the following section.

6. Extended Analyses

In the preceding section, we have analyzed the average post-ChatGPT changes in labor demand, supply, and matching outcomes for treated submarkets relative to control submarkets (RQ1). Through this analysis, we also identify several avenues warranting further investigation, prompting

a deeper examination of these aspects using both market-level and freelancer-level data. Specifically, in light of the overall decline on both sides, Section 6.1.1 explores changes in the competition land-scape and provides insights into patterns concerning average job complexity and market efficiency (RQ2). In Section 6.1.2, acknowledging both theoretical and empirical evidence that different job types may respond differently, we directly investigate two-sided heterogeneity (RQ3). Detailed analyses of labor supply dynamics are then conducted as the mechanism check in Sections 6.1.3 and 6.2 (RQ4). Moreover, given the concurrent emergence of text-to-image generative AI technologies during the study period, which represent another external shock, Section 6.3 explores their impact and reports associate changes in image-related jobs.

6.1. Market-level Extended Analyses

6.1.1. The Impact on Job Complexity, Market Competition, and Efficiency

Given the observed declines in both the supply and demand aspects of the market, there persists an ambiguity regarding the remaining job opportunities on the demand side and the competitive landscape among freelancers on the supply side. To explore these dynamics, we analyze three key variables at the submarket-month level: the average budgets of posted jobs (Budget), the average number of bids per job (Avg_bid), and the average transaction price for completed jobs ($Match_price$). Utilizing the same specification as in our primary analysis⁵, we modify only the dependent variables and present the findings in columns (1)-(3) of Table 4.

Our analysis demonstrates a significant escalation in competition within the treatment submarkets compared with the control submarkets, as evidenced by a substantial increase in the average number of bids per job ($\beta_1 = 0.171, p < 0.01$). Conventionally, such heightened competition is expected to exert downward pressure on the budgets that clients are willing to offer. Contrary to these expectations, however, we observe an increase in the average budgets of jobs ($\beta_1 = 0.152, p < 0.01$). This anomaly suggests an opposing force at play: the contraction in demand ⁵ Due to space constraints, we present only the results of Equation (1) in the subsequent tables. However, it should

be noted that all the reported results remain robust when Equation (2) is employed as the model specification.

might predominantly involve lower-complexity jobs, while the remaining higher-complexity jobs are driving the increase in budgets. Furthermore, despite this rise in budgets, the average transaction price remains unchanged ($\beta_1 = -0.0128, p > 0.1$). This gap between budgets and transaction prices can also be attributed to the intensified competition on the supply side, which mitigates the upward force on transaction prices associated with higher budgets.

Additionally, the increase in freelancers competing for each job, coupled with potential efficiency improvements facilitated by ChatGPT (red line 4 of Figure 1), might suggest that a greater ratio of job postings could be fulfilled and completed more swiftly. To investigate these possibilities, we constructed two additional outcome variables: the matching rate ($Match_rate$) and the average completion time for matched jobs ($Complete_time$). We report the regression results for these variables in columns (4)-(5) of Table 4. However, the analysis does not provide strong evidence of efficiency improvements for either measurement ($\beta_1 = 0.0016$ or -0.0981, p > 0.1). This lack of significant efficiency gains could potentially stem from the intricacies of the matching process and the heightened complexity of the remaining jobs.

Efficiency Complexity Competition $ln(Budget) ln(Avg_bid + 1)$ $ln(Match_price) \ Match_rate \ ln(Complete_time)$ (1)(3)(4)(2)(5)0.1710*** 0.1520*** Treat * Post-0.0016-0.0128-0.0981(0.0522)(0.0323)(0.0815)(0.0083)(0.1300)5.6880*** 2.5650***Constant 4.6580*** 0.2170*** 4.8550*** (0.0185)(0.0115)(0.0282)(0.0029)(0.0450)Category FE Yes Yes Yes Yes Yes Year-Month FE Yes Yes Yes Yes Yes Observations 16,265 16,265 10.032 16,265 10.032 R^2 0.3070 0.24500.1710 0.2200 0.5120

Table 4 The Impact on Job Complexity, Market Competition, and Efficiency

Robust standard errors in parentheses *** p < 0.01, ** p < 0.05, * p < 0.1

6.1.2. The Heterogeneous Impact on Different Job Types

In Section 5, we document significant decreases in supply, demand, and transaction volumes within treatment submarkets. Given the general-purpose nature of ChatGPT, which is applied

across a broad spectrum of submarkets, our findings represent an average of effects across diverse submarkets with varying skill requirements. However, there might exist notable heterogeneity among the treated submarkets. This inquiry gains further relevance as preliminary evidence presented in Section 6.1.1 indicates that the effects may differ across various jobs. Additionally, there are widespread concerns about ChatGPT's tendency to generate semantically relevant but incorrect outputs, known as the "hallucination" issue (Xu et al. 2024), which is particularly problematic in tasks that require high levels of accuracy and standardization, such as programming (Liu et al. 2024). To explore the potential heterogeneity, we introduce a binary variable *Programming*, which is assigned a value of 1 for submarkets requiring programming skills and 0 for those involving text but not programming (still referred to as "text-related" submarkets for simplicity). We employ the triple difference (DDD) specification (Equation 3), as detailed in Appendix E, to explore the heterogeneous effects on programming-related versus text-related submarkets, with the findings presented in Table 5.

$$Outcome_{it} = \beta_1 * Treat_i * Post_t + \beta_2 * Treat_i * Post_t * Programming_i + u_i + T_t + \epsilon_{it}$$
 (3)

Using control submarkets as a baseline, the parameter β_1 now captures the relative changes in text-related submarkets following the introduction of ChatGPT, while β_2 represents the additional changes observed in programming-related submarkets beyond β_1 . Interestingly, despite no apparent heterogeneous impact on labor demand between programming-related and text-related submarkets being detected ($\beta_2 = -0.0224$ or -0.0295, p > 0.1), we observe a significantly smaller reduction in labor supply within the programming-related submarkets ($\beta_2 = 0.121$, p < 0.05). This heterogeneity could be driven by several mechanisms, which will be systematically explored in subsequent sections: (i) ChatGPT enhances programming efficiency for existing freelancers, thus encouraging their continued participation in the market; (ii) it lowers the barriers for freelancers to transition into programming roles, thereby compensating for losses in labor supply; (iii) it reduces entry barriers and enhances efficiency for individuals outside the online labor markets to engage in programming work, potentially attracting new market entrants.

Furthermore, a marginal heterogeneous decrease in transactional value is observed in submarkets related to programming ($\beta_2 = -0.18, p < 0.1$). Combining this finding with the absence of significant heterogeneous changes in demand and number of matches, the heterogeneity in transactional value can be attributed to relatively smaller reduction in labor supply. Consequently, the supply-side competition within these programming-related submarkets is relatively higher, thereby exerting downward pressure on the transactional prices of completed jobs.

Table 5 The Heterogeneous Effect of Generative AI on Demand, Supply, and Matching Outcomes

	ln(y+1)					
	Jobs	Clients	Bids	Free lancers	Matches	Transaction
	(1)	(2)	(3)	(4)	(5)	(6)
$\overline{Treat*Post}$	-0.2230***	-0.2190***	-0.2280***	-0.1730***	-0.1340***	-0.2640**
	(0.0625)	(0.0622)	(0.0693)	(0.0619)	(0.0421)	(0.1290)
Treat*Post	-0.0224	-0.0295	0.1210**	0.1210**	0.0063	-0.1800*
*Programming	(0.0293)	(0.0270)	(0.0527)	(0.0514)	(0.0239)	(0.0925)
Constant	2.5540***	2.4720***	4.7190***	4.4540***	1.3620***	3.7170***
	(0.0216)	(0.0217)	(0.0218)	(0.0188)	(0.0140)	(0.0411)
Category FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17,247	17,247	17,247	17,247	17,247	17,247
R^2	0.9090	0.9160	0.8160	0.7790	0.8940	0.6520

Robust standard errors in parentheses *** p < 0.01, ** p < 0.05, * p < 0.1

6.1.3. The Impact on New Freelancers' Market Entry Decisions

Given that some freelancers can be active in online labor markets for an extended period and work for diverse jobs, several potential mechanisms may explain the average and heterogeneous impacts on the labor supply side, as documented in Sections 5.2 and 6.1.2. 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 as Equations (1) and (3) to perform DiD and DDD analyses, with the results presented in Table 6. The estimates suggest that the decline in new participants is one contributing mechanism to the observed decrease in average labor supply

across treatment submarkets ($\beta_1 = -0.205$, p < 0.01). However, there is no significant heterogeneity in the number of new freelancers between programming-related and text-related submarkets ($\beta_2 = -0.0357$, p > 0.1), effectively ruling out differences in new participation as the driving mechanism for heterogeneous effects on labor supply. In Section 6.2, we will further investigate the behavior of existing freelancers at a more granular level to elucidate the remaining mechanisms.

Table 6 The Impact of Generative AI on Market Entry Decisions of New Freelancers

	$ln(New_Freelancer + 1)$		
	(1)	(2)	
$\overline{Treat*Post}$	-0.2050***	-0.1800***	
	(0.0409)	(0.0493)	
Treat*Post*Programming		-0.0357	
		(0.0388)	
Constant	3.7110***	3.7110***	
	(0.0149)	(0.0149)	
Category FE	Yes	Yes	
Year-Month FE	Yes	Yes	
Observations	17,247	17,247	
R^2	0.9250	0.9250	

Robust standard errors in parentheses *** p < 0.01, ** p < 0.05, * p < 0.1

6.2. Freelancer-level Extended Analyses

In Sections 5 and 6.1, we have documented a relative reduction in labor supply within the treated submarkets and emphasized that this decline is less pronounced in programming-related submarkets. As a mechanism check, we also find that the decrease in new freelancers contributes to the average effect but cannot explain the heterogeneous differences within the treated submarkets. Given that freelancers can engage with the platform for an extended period, we examine the behavior of existing freelancers as additional mechanism checks: In Section 6.2.1, we investigate whether the reduction in existing freelancers' participation can be a driver for the average effect and if there exists heterogeneity in engagement among freelancers. In Section 6.2.2, we further analyze how freelancers' transitions across submarkets might serve as the mechanism for the heterogeneous labor supply change. In both sections, beyond the changes in bidding behavior, we also estimate the variations in freelancers' revenue and test their relative resilience to market trends.

Considering that any behavior occurring post-shock could be influenced by the introduction of ChatGPT, we construct the freelancer-level treatment variables based on units' behavior **prior to** the shock (Tafti and Shmueli 2020). In line with our market-level analysis, we employ a conservative methodology: among the active freelancers who bid on at least one job in November 2022, we categorize those who exclusively bid on jobs within the control submarkets before the shock as control freelancers (Treat = 0), while those who participate in the treated submarkets are classified as treated freelancers (Treat = 1). Furthermore, within the treated freelancers, we identify those who engaged in programming tasks before the shock with Programming = 1 to explore heterogeneity. Additional model specification details can be found in Appendix F.

6.2.1. The Impact on Freelancers' Bid Behaviors, Revenue, and Skill Scope

We begin by investigating how freelancers who bid at least partially on treated submarkets before the shock (treated freelancers) behave differently compared to those who do not (control freelancers) after the introduction of ChatGPT. We employ the same specification as Equation (1), using each freelancer's monthly bids ($Bid_{freelancer}$) as the outcome variable. The results are presented in Column (1) of Table 7, indicating that treated freelancers, in comparison to control freelancers, become relatively less active on the platform post-ChatGPT ($\beta_1 = -0.1180$, p < 0.01). This reduction in activity could serve as an additional mechanism for the average labor decrease observed in treated submarkets, complementing the reduced new participation mechanism documented in the previous section. Meanwhile, we also adopt the DDD specification following Equation (3) and show the result in Column (2). Interestingly, we find that freelancers who bid on programming tasks before the shock are even less active compared to other treated freelancers ($\beta_2 = -0.1160$, p < 0.01). This finding contrasts with our market-level heterogeneity analysis, where programming-related submarkets experience a smaller loss in labor supply, thereby ruling out another potential mechanism. Consequently, it is crucial to emphasize the role of skill transition as a significant factor in supply-side heterogeneity. This mechanism will be evaluated in depth in Section 6.2.2.

Beyond bidding activity, we also evaluate the economic consequences using monthly revenue as the outcome variable $(Transaction_{freelancer})$. The estimation results in Columns (3) and (4)

suggest that freelancers who have previously participated in treated markets incur relative income losses compared to those who exclusively participate in the control submarket ($\beta_1 = -0.0374$, p < 0.01, Column (3)). This effect is primarily driven by those who bid on programming tasks ($\beta_2 = -0.0469$, p < 0.01, Column (4)). Notably, the income loss for treated freelancers who did not bid on programming tasks before the shock is minimal ($\beta_1 = -0.0005$, p > 0.1, Column (4)), suggesting that these freelancers may be applying for more profitable tasks after the introduction of ChatGPT, with further discussion provided in Appendix G. This outcome also aligns with the remaining mechanism for supply-side heterogeneity: the transition of freelancers from text-related tasks to programming, as programming jobs generally command higher rates on the platform.

Table 7 The Impact of Generative AI on Freelancers' Bid Behaviors, Revenue and Skill Convergence

	ln(y+1)					
	$\overline{Bids_{fr}}$	reelancer	$Transaction_{free lancer}$		$Skill_Scope_{freelancer}$	
	(1)	(2)	(3)	(4)	(5)	(6)
Treat*Post	-0.1180***	-0.0260***	-0.0374***	-0.0005	-0.0429***	-0.0399***
	(0.0059)	(0.0060)	(0.0061)	(0.0061)	(0.0073)	(0.0079)
Treat*Post		-0.1160***		-0.0469***		-0.0037
*Programming		(0.0043)		(0.0041)		(0.0051)
$Bids_{freelancer}$					0.0062***	0.0062***
v					(0.0001)	(0.0001)
Constant	0.6370***	0.6330***	0.1960***	0.1940***	0.7170***	0.7170***
	(0.0039)	(0.0038)	(0.0040)	(0.0040)	(0.0050)	(0.0050)
Category FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,748,338	1,748,338	1,748,338	1,748,338	1,748,338	1,748,338
R^2	0.7490	0.7490	0.4070	0.4070	0.7310	0.7310

Robust standard errors in parentheses *** p < 0.01, ** p < 0.05, * p < 0.1

Note: Estimates in Columns (5) and (6) are robust without the control variable.

Additionally, we examine whether treated freelancers bid for a more diverse range of jobs with the assistance of ChatGPT. We construct a variable, $Skill_Scope_{freelancers}$, which measures the number of unique skills required for the jobs each freelancer applies for in a month. To account for the natural increase in this variable with the number of bids, we include $Bids_{freelancers}$ as a control variable in the regression. Our results, shown in Columns (5) and (6), indicate that treated freelancers tend to leverage a smaller set of skills per period post-GPT ($\beta_1 = -0.0429$ or -0.0399,

p < 0.01). However, the skills utilized before and after the introduction of ChatGPT (across periods) could vary substantially due to skill transition, a phenomenon that will be investigated in the following subsection.

6.2.2. The Impact on Job Transition Decisions of Freelancers

Building on the previous empirical evidence and conceptual deductions, we now formally test whether freelancers transition from text-related jobs to programming tasks (red line 3 of Figure 1), which represents the final potential mechanism for labor supply heterogeneity. To achieve this, we focus on a subset of freelancers who exclusively bid on text-related jobs compared to those who bid on jobs in the control submarkets before the shock.

We construct two additional outcome variables: $Is_Bid_{Programming}$, a binary indicator of whether a freelancer bids on any programming-related jobs in a given month, and $Bids_{Programming}$, the number of programming-related jobs bid on by a freelancer. We then apply the standard DiD regression as specified in Equation (1), with the results presented in Table 8. The first column employs a linear probability model focusing on observations where a freelancer (unit) bid on at least one project in that month (period). The results indicate that, within the active period, freelancers who typically engage in text-related tasks before the shock show a 14.7% increased likelihood of bidding on programming-related jobs, compared to control group freelancers. Such skill transition behaviors effectively compensates for the labor supply loss in the programming-related submarkets.

Table 8 The Impact On Freelancers' New Programming Participation Behaviors

	$Is_Bid_{Programming}$	$Is_Bid_{Programming}$	$ln(Bids_{Programming} + 1)$
Treat*Post	0.1470***	0.0096***	0.0093***
	(0.0060)	(0.0007)	(0.0009)
Constant	0.0457***	0.0042***	0.0037***
	(0.0017)	(0.0004)	(0.0005)
Category FE	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes
Active Only	Yes	No	No
Observations	58,261	505,531	505,531
R^2	0.5090	0.1770	0.2480

Robust standard errors in parentheses *** p < 0.01, ** p < 0.05, * p < 0.1

The strategic behavior of these freelancers can be rationalized. ChatGPT can assist with coding, thereby reducing the time and energy costs associated with learning or performing programming-related jobs. Additionally, programming-related jobs tend to be more lucrative than text-related jobs. Consequently, this skill transition can be seen as a rational balancing of costs and revenues. To further assess the behavior and resilience of these strategic freelancers, we identify two subgroups among those who participate in text-related submarkets before the shock: one subgroup bid on at least one programming job afterward, while the other continue to bid exclusively on text-related jobs. We compare these subgroups to the same baseline (control freelancers) and conduct separate regressions following Equation (1). The results, presented in Table 9, show that freelancers who shift from text- to programming-related jobs significantly increase their bids (0.2, p < 0.01) and their transaction volumes remain relatively stable (0.0065, p > 0.1), compared with control freelancers. In contrast, those who persist with text-related jobs see no significant change in bid quantities (-0.0024, p > 0.1) but experience a significant decrease in transaction volume (-0.0470, p < 0.01) due to reduced market demand.

Table 9 The Impact On Freelancers' Transition Behaviors

	$Text_to$.Programming	$Text_remain_Text$		
•	ln(Bids+1)	ln(Transaction + 1)	ln(Bids+1)	ln(Transaction + 1)	
Treat*Post	0.2000***	0.0065	-0.0024	-0.0470***	
	(0.0145)	(0.0137)	(0.0099)	(0.0108)	
Constant	0.2110***	0.0405***	0.2420***	0.0554***	
	(0.0026)	(0.0025)	(0.0022)	(0.0024)	
Category FE	Yes	Yes	Yes	Yes	
Year-Month FE	Yes	Yes	Yes	Yes	
Observations	163,210	163,210	180,571	180,571	
R^2	0.5120	0.3160	0.5240	0.3120	

Robust standard errors in parentheses *** p < 0.01, ** p < 0.05, * p < 0.1

Note: Limited by space, we omit the "freeelancer" subscript in the outcome variable.

6.3. Extended Analyses on Image-related jobs

Given the emergence of many text-to-image generative AI tools during the observational window, we are also interested in how image-related submarkets change in response to the launch of these generative AI technologies. We use the launch time of the first publicly available version as the treatment time (*Post_image*), set at July 2022 on a monthly level. The control submarkets serve as the baseline, while the treatment group consists of the image-related submarkets. Furthermore, the significant success of ChatGPT has popularized generative AI, potentially increasing awareness of text-to-image generative AI applications. To fully understand the relative impact of these image-generative AI tools on image-related submarkets and the potential publicity spillover effect introduced by ChatGPT, we conduct two distinct sets of analyses.

Firstly, we examine the changes in demand, supply, and matching outcomes in image-related submarkets compared to control submarkets due to the introduction of text-to-image generative AI. This analysis covers the observation window until November 2022, with the results presented in Table H1 of Appendix H. Secondly, we extend the observation window to include subsequent periods and use the original *Post* variable from the main analyses to investigate the relative changes in image-related submarkets following the introduction of ChatGPT. The results of this spillover effect analysis are presented in Table H2. Interestingly, we find no significant effects across all measures for both sets of analyses. 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. Unlike ChatGPT, the outputs of these applications are typically well-confined and do not yet qualify as AGI (Morris et al. 2023).

7. Robustness Checks

Based on the analyses presented in Sections 5 and 6, we have comprehensively examined the influence of LLM-based generative AI on online labor markets. To ensure the robustness of our findings, we conduct a series of additional checks, summarized as follows. First, we re-estimate the model using weekly data to demonstrate robustness under an alternative data aggregation process, as detailed in Appendix I. Next, we employ a relative time model to estimate the effects and plot their dynamics. Figure J1 in Appendix J shows no significant pre-existing trends and indicates that the effects can strengthen over time. We also conduct placebo tests to strengthen the rigor of our analysis, including randomly reassigning treatment (Appendix K) and using the same period

from the previous year (November 2021) as the fake treatment time (Appendix L). Furthermore, we exclude the top submarkets and rerun the estimation to verify that our results are not driven by a few submarkets, as shown in Appendix M. Moreover, we also include additional control variables in the freelancer-level analysis to enhance robustness, as discussed in Appendix N. Finally, we employ the Negative Binomial models for count variables, as detailed in Appendix O. All these comprehensive checks and additional analyses confirm the robustness of our findings.

8. Discussion and Conclusion

In this study, we conduct an in-depth examination of the impact of LLM-based generative AI on the online labor marketplace. Utilizing a comprehensive dataset from a leading freelancer platform, we uncover a multitude of labor dynamics, visually represented in Appendix P. Our analysis indicates that submarkets exposed to skills directly intersecting with ChatGPT's capabilities (text or code generation) experience a significant decline in demand compared to submarkets whose skill requirements do not overlap with these abilities. As a result, the remaining jobs in affected submarkets are relatively more complex, and competition among freelancers intensifies. In response to the evolving market conditions, labor supply and the final transactional volume of these impacted submarkets also diminish, and this decrease is not offset by improvements in market efficiency (match rate or completion time). Notably, while no heterogeneous impact on labor demand is observed between programming-related and text-related submarkets, we identify a significantly smaller labor demand loss for programming-related submarkets. We employ further analyses to elucidate the underlying mechanisms at play. Specifically, we discern a significant skill transition effect from freelancers who previously focused exclusively on text-related submarkets to engagement in programming-related job bidding. Despite the overall market contraction and the converging skill scope among freelancers on average, freelancers who undertake such transitions exhibit greater economic resilience.

Our research presents both theoretical and empirical contributions. From a theoretical perspective, we are among the pioneering studies to comprehensively assess the impact of LLM-based generative AI on labor market dynamics, providing evidence on the differential changes in twosided participation, matching processes, and outcomes across different submarkets. Moreover, we meticulously document freelancers' responses and skill transition behaviors, thereby enriching our understanding of online labor markets. Additionally, we contribute to the AI social alignment research by introducing the labor market perspective, which could be of great significance as AI evolves to become more general-purpose and powerful.

From a practical standpoint, our study underscores the heterogeneities across various job categories, suggesting that platforms may benefit from offering guidance, such as complementary training sessions, to facilitate labor supply shifts and expansions based on observed demand data. As AI continues to evolve, freelancers should consider strategically positioning themselves by cultivating their skills in a forward-looking manner. Furthermore, governments and tech companies developing AI should recognize the profound impact of these technologies on the labor market, and incorporate such considerations into AI alignment and governance principles.

Our study opens up several promising avenues for future research. First, we observe the interplay between the two sides through the matching process and the potential for freelancers to transition across different submarkets by acquiring new skills. Such market structures are also applicable to other markets, such as marriage and college admissions. However, there remains a paucity of causal inference methods suitable for matching markets, which presents a propitious opportunity for statisticians and econometricians to develop novel methodologies. Furthermore, while online labor markets encompass a wide range of jobs, expanding the scope of future research to include the entire labor market could yield a more comprehensive understanding of AI's far-reaching impact. Finally, given the potential for freelancers' skill transitions to serve as a means of adapting to technological change, future research can explore possible market designs or policies that can effectively guide and facilitate transitions from both theoretical and empirical perspectives.

References

- Acemoglu D, Autor D, Hazell J, Restrepo P (2022) Artificial intelligence and jobs: evidence from online vacancies. *Journal of Labor Economics* 40(S1):S293–S340.
- Agrawal A, Gans J, Goldfarb A (2019) The economics of artificial intelligence: an agenda (University of Chicago Press).
- Alekseeva L, Azar J, Gine M, Samila S, Taska B (2021) The demand for ai skills in the labor market. *Labour economics* 71:102002.
- Aouad A, Saban D (2023) Online assortment optimization for two-sided matching platforms. *Management Science* 69(4):2069–2087.
- Barach MA, Golden JM, Horton JJ (2020) Steering in online markets: the role of platform incentives and credibility. *Management Science* 66(9):4047–4070.
- Benson A, Sojourner A, Umyarov A (2020) Can reputation discipline the gig economy? experimental evidence from an online labor market. *Management Science* 66(5):1802–1825.
- Braghieri L, Levy R, Makarin A (2022) Social media and mental health. *American Economic Review* 112(11):3660–3693.
- Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, et al. (2020) Language models are few-shot learners. *Advances in neural information processing* systems 33:1877–1901.
- Brynjolfsson E, Li D, Raymond L (2023) Generative ai at work. NBER Working Paper (w31161).
- Burtch G, Carnahan S, Greenwood BN (2018) Can you gig it? an empirical examination of the gig economy and entrepreneurial activity. *Management Science* 64(12):5497–5520.
- Chen DL, Horton JJ (2016) Research note—are online labor markets spot markets for tasks? a field experiment on the behavioral response to wage cuts. *Information Systems Research* 27(2):403–423.
- Chen Y, Liu TX, Shan Y, Zhong S (2023) The emergence of economic rationality of gpt. Proceedings of the National Academy of Sciences 120(51):e2316205120.
- Cinelli C, Forney A, Pearl J (2022) A crash course in good and bad controls. Sociological Methods & Research 00491241221099552.

- Cramer J, Krueger AB (2016) Disruptive change in the taxi business: The case of uber. *American Economic Review* 106(5):177–182.
- De Marzo G, Pietronero L, Garcia D (2023) Emergence of scale-free networks in social interactions among large language models. arXiv preprint arXiv:2312.06619.
- 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.
- Gabriel I (2020) Artificial intelligence, values, and alignment. Minds and machines 30(3):411-437.
- Gallego A, Kurer T (2022) Automation, digitalization, and artificial intelligence in the workplace: implications for political behavior. *Annual Review of Political Science* 25:463–484.
- Garg N, Johari R (2021) Designing informative rating systems: Evidence from an online labor market.

 Manufacturing & Service Operations Management 23(3):589–605.
- Hong Y, Wang C, Pavlou PA (2016) Comparing open and sealed bid auctions: Evidence from online labor markets. *Information Systems Research* 27(1):49–69.
- Horton JJ (2019) Buyer uncertainty about seller capacity: Causes, consequences, and a partial solution.

 Management Science 65(8):3518–3540.
- Huang N, Burtch G, Hong Y, Pavlou PA (2020) Unemployment and worker participation in the gig economy: Evidence from an online labor market. *Information Systems Research* 31(2):431–448.
- Jensen R, Miller NH (2018) Market integration, demand, and the growth of firms: Evidence from a natural experiment in india. *American Economic Review* 108(12):3583–3625.
- Ji J, Qiu T, Chen B, Zhang B, Lou H, Wang K, Duan Y, He Z, Zhou J, Zhang Z, et al. (2023) Ai alignment:

 A comprehensive survey. arXiv preprint arXiv:2310.19852.
- Jia N, Luo X, Fang Z, Liao C (2024) When and how artificial intelligence augments employee creativity.

 *Academy of Management Journal 67(1):5–32.
- Kanat I, Hong Y, Raghu T (2018) Surviving in global online labor markets for it services: A geo-economic analysis. *Information systems research* 29(4):893–909.
- Kirk HR, Vidgen B, Röttger P, Hale SA (2024) The benefits, risks and bounds of personalizing the alignment of large language models to individuals. *Nature Machine Intelligence* 1–10.

- Kojima T, Gu SS, Reid M, Matsuo Y, Iwasawa Y (2022) Large language models are zero-shot reasoners.

 Advances in neural information processing systems 35:22199–22213.
- Kokkodis M, Ipeirotis PG (2023) The good, the bad, and the unhirable: Recommending job applicants in online labor markets. *Management Science*.
- Kokkodis M, Ransbotham S (2023) Learning to successfully hire in online labor markets. *Management Science* 69(3):1597–1614.
- Korinek A, Balwit A (2022) Aligned with whom? direct and social goals for ai systems. Technical report, National Bureau of Economic Research.
- Kosinski M (2023) Theory of mind may have spontaneously emerged in large language models. $arXiv\ preprint$ arXiv:2302.02083.
- Leung MD (2018) Learning to hire? hiring as a dynamic experiential learning process in an online market for contract labor. *Management Science* 64(12):5651–5668.
- Liang C, Hong Y, Chen PY, Shao BB (2022) The screening role of design parameters for service procurement auctions in online service outsourcing platforms. *Information Systems Research* 33(4):1324–1343.
- Liang C, Hong Y, Gu B (2024) Monitoring and the cold start problem in digital platforms: Theory and evidence from online labor markets. *Information Systems Research* .
- Liu J, Xia CS, Wang Y, Zhang L (2024) Is your code generated by chatgpt really correct? rigorous evaluation of large language models for code generation. Advances in Neural Information Processing Systems 36.
- McElheran K, Li JF, Brynjolfsson E, Kroff Z, Dinlersoz E, Foster L, Zolas N (2024) Ai adoption in america: Who, what, and where. Journal of Economics & Management Strategy .
- Megginson LC (1963) Lessons from europe for american business. The Southwestern Social Science Quarterly 3–13.
- Mei Q, Xie Y, Yuan W, Jackson MO (2024) A turing test of whether ai chatbots are behaviorally similar to humans. *Proceedings of the National Academy of Sciences* 121(9):e2313925121.
- Morris MR, Sohl-dickstein J, Fiedel N, Warkentin T, Dafoe A, Faust A, Farabet C, Legg S (2023) Levels of agi: Operationalizing progress on the path to agi. arXiv preprint arXiv:2311.02462.

- Paolillo A, Colella F, Nosengo N, Schiano F, Stewart W, Zambrano D, Chappuis I, Lalive R, Floreano D (2022) How to compete with robots by assessing job automation risks and resilient alternatives. *Science Robotics* 7(65):eabg5561.
- Peng S, Kalliamvakou E, Cihon P, Demirer M (2023) The impact of ai on developer productivity: Evidence from github copilot. $arXiv\ preprint\ arXiv:2302.06590$.
- Perez-Truglia R (2020) The effects of income transparency on well-being: Evidence from a natural experiment. $American\ Economic\ Review\ 110(4):1019-1054.$
- Rawte V, Chakraborty S, Pathak A, Sarkar A, Tonmoy S, Chadha A, Sheth AP, Das A (2023) The troubling emergence of hallucination in large language models—an extensive definition, quantification, and prescriptive remediations. arXiv preprint arXiv:2310.04988.
- Staccioli J, Virgillito ME (2021) The present, past, and future of labor-saving technologies. *Handbook of Labor, Human Resources and Population Economics*, 1–16 (Springer).
- Tafti A, Shmueli G (2020) Beyond overall treatment effects: Leveraging covariates in randomized experiments guided by causal structure. *Information Systems Research* 31(4):1183–1199.
- Xu X, Lee S, Tan Y (2023) Algorithmic collusion or competition: the role of platforms' recommender systems. $arXiv\ preprint\ arXiv:2309.14548$.
- Xu Z, Jain S, Kankanhalli M (2024) Hallucination is inevitable: An innate limitation of large language models. arXiv preprint arXiv:2401.11817.
- Xue J, Wang L, Zheng J, Li Y, Tan Y (2023) Can chatgpt kill user-generated q&a platforms? Available at $SSRN\ 4448938$.
- Yilmaz ED, Naumovska I, Aggarwal VA (2023) Ai-driven labor substitution: Evidence from google translate and chatgpt. $Available\ at\ SSRN$.
- Zhang S, Lee D, Singh P, Mukhopadhyay T (2022) Demand interactions in sharing economies: Evidence from a natural experiment involving airbnb and uber/lyft. *Journal of Marketing Research* 59(2):374–391.

Appendix A: 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 accompanying flowchart:

- 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.

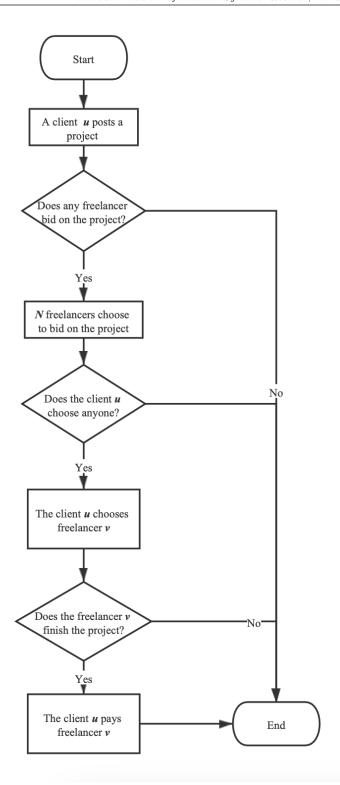


Figure A1 Matching Process

Appendix B: Detailed Job Classification and Submarket Construction Process

In this Appendix, we elucidate the details of job classification and submarket construction process. Our raw data encompasses all jobs along with their associated skill tags JT and a platform-generated dictionary comprising 12 prevalent skill categories and related tags CT. However, the platform does not maintain a comprehensive list of all skill tags associated with these prevalent categories, and furthermore, not all skill tags fall within the scope of these 12 categories.

To classify all jobs based on their tags and subsequently form submarkets, we need a comprehensive classification (category-tag) dictionary that assigns a single category to each skill tag firstly. This process is illustrated in the pseudocode Algorithm 1. Initially, we retrieve all unique skill tags that appear on the platform and generate a tag list without duplicates T. Subsequently, we examine these tags individually: if the platform has classified the tag t to category c, we retain that category. Otherwise, we determine if there exists a suitable job category c for the tag and append the tag to the category's maintained list CT_c if one is found. If no existing category is suitable, we create a new category with a newly initiated tag list maintained for that category. Through this process, we obtain an updated dictionary CT with 29 categories, assigning each of the 2,719 tags to one of these categories. This is done by two researchers of the author team manually to ensure consistent understanding.

Based on the comprehensive category-tag dictionary CT derived from the aforementioned process and the original job-tag dictionary JT of our raw dataset, we further determine the skill categories related to each job and subsequently form the submarkets. Following the pseudocode Algorithm 2, for each job with one or multiple skill tags, we employ CT to identify the corresponding category for each tag and ascertain the skill category or categories related to this job, storing the information in a new job-category dictionary JC. Finally, based on JC, we utilize jobs with identical combinations of categories to form independent submarkets, as jobs within each submarket require similar skills, and labor supply and demand for these skills match each other. After removing inactive submarkets, we retain a total of 873 submarkets, each of which has a minimum of one job posting per month on average.

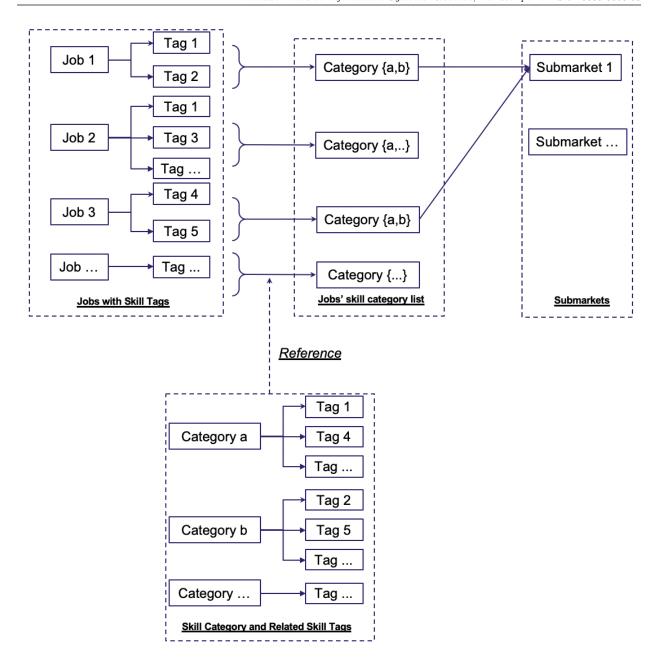


Figure B1 The Relationship between Jobs, Tags, Skill Categories, and Submarkets.

Algorithm 1 Category-Tag Dictionary Construction Process

Require: Job-Tag Dictionary JT, Platform-Provided Category-Tag Dictionary CT

1. Get all the tags on the labor market

```
Initialize Tag List T
for each job j \in JT: do
   Retrieve all the tags this job has JT_j.
   for each tag t: do
       Append t to T if t \notin T
   end for
end for
2. Construct the full category-tag dictionary CT
for each tag t \in T: do
   if \exists c \in CT, t \in CT_c then
       continue
   else
       if \exists c, t is suitable for CT_c then
           Append t to c's list of tags CT_c
       else
           Add one more key (category) c to CT and initialize its item (list of tags) with CT_c = [t]
       end if
   end if
end for
```

end for

Algorithm 2 Job-Category Dictionary and Submarket Construction Process

Require: Job-Tag Dictionary JT, Category-Tag Dictionary CT Returned from Algorithm 1

3. Construct the job-category dictionary JC

```
Initialize job-category dictionary JC
for each job j \in JT: do
   Retrieve all the tags this job has JT_j.
   for each tag t \in JT_j: do
       Search CT to find the category c where t \in CT_c
      Append c to JC_j if c \notin JC_j
   end for
end for
4. Construct the Submarket dictionary M
Initialize Submarket dictionary M
for each job j \in JC: do
   Retrieve all the job categories for this job JC_j.
   if JC_j \in M then
      Append job j to the submarket M_{JC_j}
   else
      Add a key (submarket) JC_j to M, and initialize its item as M_{JC_j}=[j]
   end if
```

Appendix C: Submarket Types, Skill Examples, and Treatment Variables

Following the process outlined in Appendix B, we categorized all skill tags into 29 distinct categories. As demonstrated in Table C1 - Panel A, the majority of these skill categories are related to either programming or text generation, which fall within the capabilities of text-generative AI, or to image generation, which aligns with the functions of text-to-image generative AI. As the extent and nature of Generative AI's impact on these jobs remain unknown and form the core of this research, we adopt an exploratory approach and assign types to these skill categories based solely on their relation (binary indicator instead of weights) to programming, text, and images. The remaining skill categories that lack a direct relationship to these functions are considered as the control type, serving as a baseline for comparison in our analysis.

Moreover, as a single job might require multiple skills, sometimes spanning different skill categories and even types, a job's (or submarket's) category list may reflect several types. For instance, if a job post requires freelancers to edit a video and write a report, this project necessitates both text-related skills and control skills. In keeping with the exploratory spirit, we still consider this job as belonging to the text-related submarkets. Consequently, as shown in Panel B of Table C1, only when all skill categories within a submarket belong to the control type, the submarket is classified as a control submarket. This approach allows for a more conservative analysis of the potential impact of Generative AI on various skill categories and submarkets, while acknowledging the complexity of job requirements in the online labor market.

Similarly, as our primary interest lies in ChatGPT, a general-purpose text-generative AI, we aim to minimize the influence of text-to-image generative AI, which was also launched during the observation period. Therefore, if even a single skill category out of a submarket's entire skill category list is related to image output, we attribute the submarket to the image-related submarkets, and exclude it from the main analyses.

Therefore, when conducting the main analyses to examine the post-GPT relative changes in submarkets related to text and/or programming versus the control submarkets, we exclude the image-related submarkets and assign Treat = 0 to control submarkets and Treat = 1 to the other (treatment) submarkets. We also introduce an additional indicator variable, Programming, to denote whether a treatment submarket includes at least one skill category related to programming. When studying the relative impact on image-related submarkets compared to control submarkets, we define a separate treatment variable, $Treat_{image}$, to distinguish the effect of text-to-image generative AI from that of text-generative AI.

Table C1 Skill Type, Skill Categories, and Tags Examples

	Table C1 Skill Type,	Daniel A
	C1:11 C +	Panel A
Type	Skill Categories	Skill Tag Examples
Programmin	g Statistics Programming	SPSS, Matlab and Mathematica, SAS
	Software Development	C, C++, Python, Java, Software Architecture
	Web Development	HTML, CSS, Web Application
	Mobile App Development	iPhone, Android, App Design
	Data Processing and Mining	Data Visualization, Data Mining, Data Analysis
	Database Programming	MySQL, MySQL, Oracle
	SEO	Google Analytics, SEO, Link Building
	Data Scraping and Entry	Data Entry, Web Scraping, Data Scraping
	Artificial Intelligence	AI/RPA development, Robotic Process Automation
	Other Coding	Algorithm Analysis, Process Engineering
Text	Translation	French Translator, English Translator
	Internet Marketing	SocialMedia Post Design, Content Strategy
	Other Marketing	Email Marketing, Email Campaign, Telemarketing
	Writing	Ghostwriting, Speech Writing, Content Writing
	Microsoft	Excel, Word, Powerpoint, Google Docs
	Proofreading	Document Checking, English Grammar, Spelling
	Business Analysis	Business Plans, Business Strategy, Sales
	Customer Service	Customer Support, Customer Experience
	Education	Education & Tutoring, Teaching/Lecturing
	Legal	Contracts, Legal Review, Legal Assistance
Image	Graphic Design	Logo Design, Illustrator, Icon Design
	Photography	Photography, Photoshop, Photo Editing
	Graphic Art	Visual Arts, Graphic Art, Image Processing
Control	Architecture	Civil Engineering, Building Architecture
	3D & Animation Modeling	3D Animation, 3D Rendering, 3D Design
	Video Services	Videography, Video Services, Video Editing
	Audio Services	Audio Engineering, Audio Editing, Audio Services
	Manufacturing	Mechanical Engineering, Electronics
	Logistics	Shipping, Delivery, Packing & Shipping
		Panel B
Skill Type C	Combination	Submarket Type
Programmin	g	Programming-related Submarkets
Programmin	g, Text	Programming-related Submarkets
Programmin	g, Control	Programming-related Submarkets
Programmin	g, Text, Control	Programming-related Submarkets
Text		Text-related Submarkets
Text, Contro	ol	Text-related Submarkets
Image		Image-related Submarkets
Image, Text		Image-related Submarkets
Image, Prog	ramming	Image-related Submarkets
Image, Cont	rol	Image-related Submarkets
Image, Prog	ramming, Text	Image-related Submarkets
Image, Prog	ramming, Control	Image-related Submarkets
Image, Text,		Image-related Submarkets
All		Image-related Submarkets
Control		Control Submarkets

Table C2 Submarket Type and Treatment Variables

Submarket Type	Treat	Programming	$Treat_{image}$
Programming-related Submarkets	1	1	NA
Text-related Submarkets	1	0	NA
Image-related Submarkets	NA	NA	1
Control Submarkets	0	0	0

Appendix D: Variable Summary Statistics

This section presents the summary statistics of our main variables of interest. It is crucial to note that budget, Avg_bids, and Match_rate are only measured when at least one job is posted to the submarket in a given month. Moreover, Match_price and Complete_time are only measured when there is at least one match (a job successfully completed) in the submarket in a given month. As a result, the number of observations may vary across these variables. Whenever they are measured, there will be a positive value. Some jobs are relatively simple, requiring less than one hour and costing less than one dollar, resulting in negative values for the variables after log-transformation.

Since each submarket emerges on the platform at different times, the number of observations varies across submarkets. To illustrate, consider the number of job posts as an example. If a submarket has its first job posting in January 2022, we treat this as the submarket's inception. In subsequent months, the number of job posts in the submarket is recorded, with months lacking job posts marked as 0. Prior to January 2022, no observations are available for this submarket. As for freelancers, we treat their registration time as the moment they first enter the market. This approach is consistently applied to all the other variables.

It is also noteworthy that the variables exhibit a high degree of skewness. Only after log transformation do the mean and standard deviation of these variables become comparable in scale, rendering count models inappropriate for our scenarios. Furthermore, the mean of the log-transformed variables and the log-transformed mean differ considerably, particularly when the data is highly skewed. For instance, the mean of transaction is approximately 2,400, whose log transformation exceeds 7, while the mean of ln(transaction + 1) is less than half of this value. This phenomenon also applies to the freelancer-level data, as it is not uncommon for multiple workers to share a single account and collectively bid on a wide range of jobs. In such cases, a single account can complete hundreds of tasks across different categories within one month.

Table D1 Variable Summary Statistics

Variable	Observation	Mean	Std	Min	Max
Demand-Side					
ln(Jobs+1)	17,247	2.4667	1.6214	0	8.6087
ln(Clients+1)	17,247	2.3844	1.5776	0	8.2244
Supply-Side					
ln(Bids+1)	17,247	4.6677	2.3960	0	11.8199
ln(Freelancers + 1)	17,247	4.4227	2.1672	0	11.1208
Market Matching					
$\overline{ln(Matches+1)}$	17,247	1.3148	1.4165	0	7.3199
ln(Transaction + 1)	17,247	3.5736	3.4603	0	12.4231
Other Market Variables					
Complexity: ln(Budget)	16,265	5.7415	1.0251	0	11.8068
$Competition: ln(Avg_bids + 1)$	$16,\!265$	2.6254	0.9423	0	5.6490
$Competition: ln(Match_price)$	10,041	4.6531	1.3589	-8.5172	10.4015
$Efficiency: Match_rate$	$16,\!265$	0.2167	0.2072	0	1
$Efficiency: ln(Complete_time)$	10,041	4.8185	2.3216	-6.8024	9.6151
$ln(New_free lancers + 1)$	17,247	3.6357	1.7483	0	10.5407
Freelancer-Level					
$ln(Bids_{freelancer} + 1)$	1,748,338	0.5601	1.1830	0	8.3102
$ln(Matches_{freelancer} + 1)$	1,748,338	0.0821	0.3515	0	6.1003
$ln(Transaction_{freelancer} + 1)$	1,748,338	0.1715	0.9572	0	11.4084

Appendix E: Notes on Difference-in-difference-in-differences Specification

In Section 6.1.2, 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 (3). This phenomenon arises because, in our analysis, any submarket categorized under Programming = 1 is simultaneously considered a treated submarket (Treat = 1). In other words, programming-related 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 (3) is formulated, wherein Treat*Post captures the post-ChatGPT changes in text-only treated submarkets relative to the control submarkets. Meanwhile, Treat*Programming*Post captures additional differences in programming-related submarkets compared to text-only submarkets, highlighting the heterogeneity of interest.

Appendix F: Notes on Freelancer-level Model Specification

Similar to the market-level analyses, we categorize freelancers into three groups to conduct the main analyses in Section 6.2.1, based on their bidding behavior prior to the shock. Specifically, among freelancers who submitted a bid on at least one job in November 2022, we classify those who exclusively bid on jobs within the control submarkets before the shock as control freelancers (Treat = 0), and those who participate in the treated submarkets as treated freelancers (Treat = 1). Freelancers who bid on jobs in image-related submarkets are excluded to minimize the confounding effects of text-to-image generative AI. Additionally, if a freelancer bid on at least one job in a programming-related submarket, they are categorized as a programming-related freelancer (Programming = 1) to capture their ability to bid on such jobs before the shock. Thus, similar to our market-level analyses, programming-related freelancers are a subset of treated freelancers and absorb the effect of Treat * Programming * Post.

Consequently, most model specifications remain consistent: Columns (1) and (3) of Table 7 and all columns in Tables 8 and 9 follow Equation (1), while Columns (2) and (4) of Table 7 follow Equation (3). Additionally, in Columns (5) and (6) of Table 7, we evaluate the diversity of skill requirements among the jobs a freelancer bids on during a month. Since the observation of skill requirements depends on bids submitted, the average decrease in bids can naturally lead to a smaller skill scope. Therefore, we control for the number of bids to determine if the skill scope still shrinks among the treated freelancers. When we add such time-variant controls, the model specifications are as follows:

$$Outcome_{it} = \beta_1 * Treat_i * Post_t + Control_{it} + u_i + T_t + \epsilon_{it}$$

$$\tag{4}$$

$$Outcome_{it} = \beta_1 * Treat_i * Post_t + \beta_2 * Treat_i * Post_t * Programming_i + Control_{it} + u_i + T_t + \epsilon_{it}$$
 (5)

Such specifications are also used in Appendix G and Appendix N for various reasons.

Appendix G: Control Variable Inclusion for Profitability Check

In our freelancer-level analyses, we find that treated freelancers tend to incur a wage loss compared to control freelancers after the launch of ChatGPT. However, this might simply be due to their decreased bidding activity. Therefore, we aim to evaluate the earning efficiency of different freelancers' bids by controlling for the number of bids they make. Using the model specifications shown in Equations (4) and (5), we rerun the regression. Interestingly, we find that although the coefficients are smaller in magnitude compared to those reported in Table 7, they remain significant. The results indicate that the decrease in freelancer income is not only due to fewer bids but also because the average gain from each bid has decreased. This aligns with the increased competition among freelancers, leading to a lower chance of securing a job.

It is also noteworthy that the earning efficiency of freelancers who previously bid on text-related jobs does not significantly decrease compared to control freelancers. This supports their rational decision to transition skills: although competition generally increases, the higher revenue rates in programming tasks can compensate for their lower probability of winning job opportunities.

Table G1 The Impact of Generative AI on Freelancers' Earning Efficiency

	$ln(Transaction_{freelancer} + 1)$			
	(1)	(2)		
Treat*Post	-0.0267***	5.39e-05		
	(0.0060)	(0.0060)		
Treat*Post*Programming		-0.0340***		
		(0.0040)		
$Bids_{freelancer}$	0.0045***	0.0045***		
·	(0.0001)	(0.0001)		
Constant	0.1440***	0.1430***		
	(0.0041)	(0.0041)		
Category FE	Yes	Yes		
Year-Month FE	Yes	Yes		
Observations	1,748,338	1,748,338		
$=$ $\frac{R^2}{R^2}$	0.4330	0.4330		

Appendix H: Image-Related Jobs

In both analyses, we use $Treat_{image,i}$ to indicate whether a unit i is an image-related submarket $(Treat_{image,i} = 1)$ or a control submarket $(Treat_{image,i} = 0)$. In the first analysis, we use the month that text-to-image generative AI became publicly available (July 2022) to denote $Post_{image,t}$, with observations extending only to November 2022, before the launch of ChatGPT. These analyses follow Equation (6) and examine whether image-related submarkets exhibit different trends compared to control submarkets following the introduction of text-to-image generative AI. The results are presented in Table H1.

For the second analysis, we aim to determine whether the success of ChatGPT has drawn attention to text-to-image generative AI applications, potentially motivating wider adoption and impacting labor markets. Thus, consistent with our main analyses, we use the launch time of ChatGPT to denote *Post* and include all subsequent periods. This analysis follows Equation (7), with results shown in Table H2.

Both sets of analyses show insignificant results across all measurements.

$$Outcome_{it} = \beta_1 * Treat_{image,i} * Post_{image,t} + u_i + T_t + \epsilon_{it}$$

$$\tag{6}$$

$$Outcome_{it} = \beta_1 * Treat_{image,i} * Post_t + u_i + T_t + \epsilon_{it}$$

$$\tag{7}$$

	ln(y+1)						
	Jobs	Clients	Bids	Free lancers	Matches	Transaction	
$\overline{Treat_{image} * Post_{image}}$	0.0011	0.0037	-0.0437	-0.0625	0.0990	0.2500	
	(0.0626)	(0.0567)	(0.0890)	(0.0848)	(0.0774)	(0.2560)	
Constant	2.3760***	2.3070***	4.5460***	4.2490***	1.7970***	3.3150***	
	(0.0186)	(0.0168)	(0.0264)	(0.0252)	(0.0230)	(0.0759)	
Category FE	Yes	Yes	Yes	Yes	Yes	Yes	
Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes	
Observations	2,579	2,579	2,579	2,579	2,579	2,579	
R^2	0.9400	0.9440	0.8460	0.8060	0.9210	0.7560	

Table H1 The Effect of Generative AI on Image-related Labor Market

Robust standard errors in parentheses *** p < 0.01, ** p < 0.05, * p < 0.1

Table H2 The Spillover Effect of ChatGPT's Success on Image-related Labor Market

	ln(y+1)						
	Jobs	Clients	Bids	Free lancers	Matches	Transaction	
$\overline{Treat_{image} * Post}$	-0.0627	-0.0843	0.0128	0.0247	0.0312	0.0237	
	(0.0827)	(0.0775)	(0.1030)	(0.0846)	(0.0728)	(0.2100)	
Constant	2.3310***	2.2740***	4.5490***	4.2460***	1.7220***	3.3970***	
	(0.0275)	(0.0258)	(0.0342)	(0.0282)	(0.0242)	(0.0700)	
Category FE	Yes	Yes	Yes	Yes	Yes	Yes	
Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes	
Observations	4,136	4,136	4,136	4,136	4,136	4,136	
R^2	0.9380	0.9430	0.8310	0.7870	0.9210	0.7430	

Appendix I: 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 Tables I1 to I3. 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 will be a substantial number of zero observations, which makes it less appropriate to directly interpret the coefficients as proportional changes.

Table I1 The Impact of Generative AI on Labor Demand Volume (Week-level)

		(bs+1)	ln(Clients+1)		
Treat*Post	-0.1630***	-0.1630***	-0.1680***	-0.1680***	
	(0.0491)	(0.0491)	(0.0487)	(0.0487)	
Constant	1.4130***	1.3140***	1.3750***	1.2790***	
	(0.0186)	(0.0053)	(0.0184)	(0.0050)	
Category FE	Yes	Yes	Yes	Yes	
Year FE	No	Yes	No	Yes	
Week FE	No	Yes	No	Yes	
Year-Week FE	Yes	No	Yes	No	
Observations	74,017	74,017	74,017	74,017	
R^2	0.8790	0.8790	0.8850	0.8850	

Robust standard errors in parentheses *** p < 0.01, ** p < 0.05, * p < 0.1

Table I2 The Impact of Generative AI on Labor Supply Volume (Week-level)

	ln(Bids+1)		ln(Freela	ncers + 1)
Treat*Post	-0.1890***	-0.1890***	-0.1480***	-0.1480***
	(0.0510)	(0.0510)	(0.0458)	(0.0457)
Constant	3.0250***	2.8940***	2.8870***	2.7720***
	(0.0193)	(0.0142)	(0.0173)	(0.0141)
Category FE	Yes	Yes	Yes	Yes
Year FE	No	Yes	No	Yes
Week FE	No	Yes	No	Yes
Year-Week FE	Yes	No	Yes	No
Observations	74,017	74,017	74,017	74,017
R^2	0.7640	0.7630	0.7360	0.7350

Table I3 The Impact of Generative AI on Matching Outcomes (Week-level)

	ln(Matches + 1)			action + 1)
Treat*Post	-0.0647**	-0.0647**	-0.3490***	-0.3490***
	(0.0298)	(0.0298)	(0.0822)	(0.0821)
Constant	0.6590***	0.6260***	1.9890***	1.8530***
	(0.0113)	(0.0043)	(0.0311)	(0.0202)
Category FE	Yes	Yes	Yes	Yes
Year FE	No	Yes	No	Yes
Week FE	No	Yes	No	Yes
Year-Week FE	Yes	No	Yes	No
Observations	74,017	74,017	74,017	74,017
R^2	0.8630	0.8630	0.6300	0.6290

Appendix J: 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 trend 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).

The results are shown in Figure J1. 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 figure reveals 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.

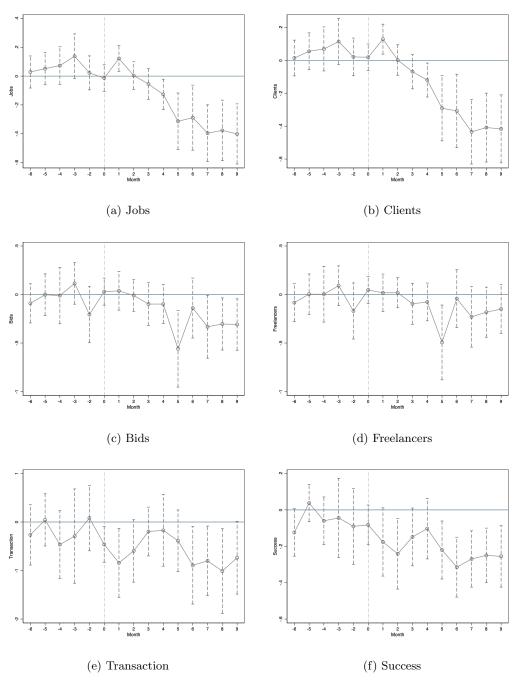


Figure J1 Pretrends and Temporal Dynamics

Appendix K: 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 six main outcome variables based on random treatment assignment. The distribution of placebo treatment effects resembles a normal distribution, as shown in Figure K1, 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.

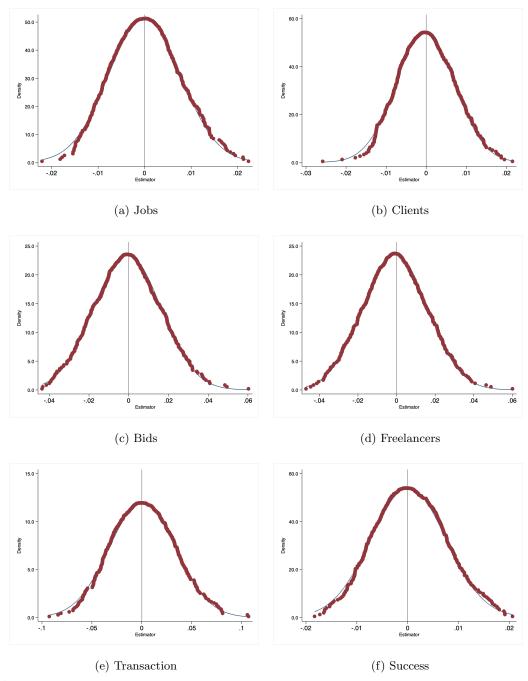


Figure K1 Placebo Tests

Appendix L: Robustness Check - Fake Treatment Time (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 fake shock date. We maintain the time period from September 2021 to November 2022 and rerun all regressions. As shown in Table L1, our main findings remain robust, suggesting that they are not driven by end-of-year effects or seasonal factors.

Table L1 Placebo Test with a Fake Shock Time

		ln(y+1)						
	Jobs	Clients	Bids	Free lancer	s Matches	$\overline{Transaction}$		
$Treat * Post_{fake}$	-0.0274	-0.0275	-0.0288	-0.0083	0.0372	0.1460		
	(0.0541)	(0.0526)	(0.0535)	(0.0407)	(0.0612)	(0.1540)		
Constant	2.6140***	2.5240***	4.7120***	4.4370***	1.3500***	3.5320***		
	(0.0414)	(0.0402)	(0.0409)	(0.0311)	(0.0468)	(0.1180)		
Category FE	Yes	Yes	Yes	Yes	Yes	Yes		
Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes		
Observations	10,047	10,047	10,047	10,047	10,047	10,047		
R^2	0.9140	0.9200	0.8370	0.8010	0.8970	0.6580		

Appendix M: Robustness Check - Removing Outliers

Given that submarkets can vary significantly in size, one might question whether the estimates are disproportionately affected by the largest submarkets. To address this concern, we exclude the top 1% of submarkets and re-estimate the model. The results, presented in Table M1, remain consistent with our main findings. This indicates that our conclusions are not driven by a few large submarkets.

Table M1 The Impact of Generative AI after Excluding Outliers

		ln(y+1)					
	Jobs	Clients	Bids	Freelancer	$rs\ Matches$	$\overline{Transaction}$	
Treat*Post	-0.2510***	-0.2500***	-0.1480**	-0.0938*	-0.1330***	-0.3980***	
	(0.0610)	(0.0615)	(0.0622)	(0.0532)	(0.0400)	(0.1180)	
Constant	2.5010***	2.4200***	4.6530***	4.4040***	1.3080***	3.6350***	
	(0.0223)	(0.0225)	(0.0228)	(0.0195)	(0.0146)	(0.0431)	
Category FE	Yes	Yes	Yes	Yes	Yes	Yes	
Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes	
Observations	17,055	17,055	17,055	17,055	17,055	17,055	
R^2	0.8970	0.9050	0.8020	0.7660	0.8770	0.6330	

Appendix N: Robustness Check - Active Periods as the Control Variable

At the freelancer level, experience in the online labor market can significantly influence decisions to participate and chances of winning bids. In a two-way fixed effects model, a freelancer's tenure, which naturally increases by one per period, is absorbed by the fixed effects. To further control for individual experience, we add a time-variant control variable (*Active_Period*), which measures the cumulative number of periods a freelancer has bid on jobs up to the each observation period. As shown in Table N1, our results remain robust with the inclusion of this additional control variable.

Table N1 The Impact of Generative AI on Freelancers' Bid Behaviors, Revenue (Add Controls)

	ln(y+1)					
	Bids		Trans	action		
	(1)	(2)	(3)	(4)		
Treat*Post	-0.1290***	-0.0265***	-0.0392***	-0.0006		
	(0.0058)	(0.0061)	(0.0060)	(0.0061)		
Treat*Post*Programming		-0.1330***		-0.0500***		
		(0.0041)		(0.0036)		
$Active_Period$	0.0113***	0.0131***	0.0017	0.0024		
	(0.0010)	(0.0010)	(0.0015)	(0.0015)		
Constant	0.6080***	0.5980***	0.1920***	0.1880***		
	(0.0051)	(0.0050)	(0.0063)	(0.0062)		
Category FE	Yes	Yes	Yes	Yes		
Year-Month FE	Yes	Yes	Yes	Yes		
Observations	1,748,338	1,748,338	1,748,338	1,748,338		
R^2	0.4330	0.4330	0.7490	0.7490		

Appendix O: Robustness Check - Alternative Model for Count Variables

It is noteworthy that several of our variables, including Jobs, Clients, Bids, Freelancers, and Matches, 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, we employ it to re-estimate the impact on these count variables. Tables O1 present 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-related or text-related submarkets, with no heterogeneity observed between these two types of treated submarkets. On the labor supply side, we still observe a significant decrease in text-related submarkets, while the programming-related submarkets incur a significantly smaller supply loss, particularly in terms of the number of active freelancers. Regarding the matching outcomes, the number of completed jobs also decreases. Notably, we observe that programming-related submarkets incur a smaller loss in this measure as well. Although in the main analyses, this heterogeneity is not significant, the coefficient is in the same direction. The positive coefficient is consistent with the supply-side heterogeneity and skill transition, as the freelancer pool available to clients in programming-related submarkets shrinks less compared to those in text-related submarkets.

Table O1 Alternative Model for Count Variables

		Negative Binominal Regressions				
	Jobs	Clients	Bids	Freelancers Matches		
Treat*Post	-0.2810***	-0.2650***	-0.1990***	-0.1440**	-0.2660***	
	(0.0668)	(0.0646)	(0.0609)	(0.0565)	(0.0461)	
Treat*Post	-0.0412	-0.0489	0.1130***	0.1110***	0.0770**	
*Programming	(0.0322)	(0.0304)	(0.0425)	(0.0408)	(0.0360)	
Constant	1.3880***	0.0705***	2.5850***	2.5530***	0.0038	
	(0.0154)	(0.0125)	(0.0396)	(0.0393)	(0.0182)	
Category FE	Yes	Yes	Yes	Yes	Yes	
Year-Month FE	Yes	Yes	Yes	Yes	Yes	
Observations	17,247	17,247	17,247	17,247	17,247	

Appendix P: Summary of Results

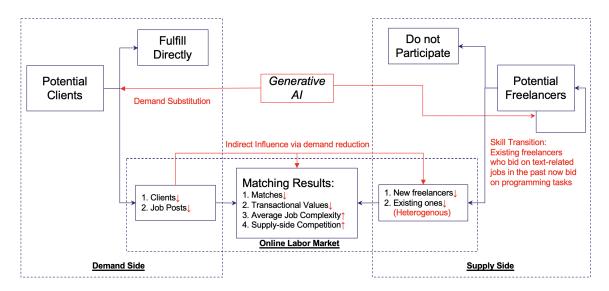


Figure P1 Visual Summary of Results