CESifo WORKING PAPERS
10601
2023
August 2023

The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market

Xiang Hui, Oren Reshef, Luofeng Zhou

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10601*

The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market

Abstract

Generative Artificial Intelligence (AI) holds the potential to either complement knowledge workers by increasing their productivity or substitute them entirely. We examine the short-term effects of the recent release of the large language model (LLM), ChatGPT, on the employment outcomes of freelancers on a large online platform. We find that freelancers in highly affected occupations suffer from the introduction of generative AI, experiencing reductions in both employment and earnings. We find similar effects studying the release of other image-based, generative AI models. Exploring the heterogeneity by freelancers' employment history, we do not find evidence that high-quality service, measured by their past performance and employment, moderates the adverse effects on employment. In fact, we find suggestive evidence that top freelancers are disproportionately affected by AI. These results suggest that in the short term generative AI reduces overall demand for knowledge workers of all types, and may have the potential to narrow gaps among workers.

Keywords: generative AI, large language model (LLM), online labor market.

Xiang Hui
Washington University in St. Louis
St. Louis / MO / USA
hui@wustl.edu

Oren Reshef
Washington University in St. Louis
St. Louis / MO / USA
oren@wustl.edu

Luofeng Zhou
Leonard N. Stern School of Business
New York University / NY / USA
lz2198@stern.nyu.edu

1 Introduction

Recent developments, followed by the rapid adoption of generative Artificial Intelligence (AI) models, such as ChatGPT and Midjourney, have offered great promises and perils. Powerful AI tools have dramatically improved performance over previous versions and users are able to use them to complete a variety of tasks without requiring specialized knowledge. In that sense, they can be thought of as a general-purpose technology (Brynjolfsson and Mitchell, 2017), with the potential for far-ranging economic and societal effects.

The effects of this new technology on workers remain unclear. On the one hand, AI can complement human workers by increasing their (Noy and Zhang, 2023); while on the other hand it may substitute workers, leading to mass layoffs and unemployment (Acemoglu and Autor, 2011; Brynjolfsson et al., 2018a; Agrawal et al., 2019a). AI may also alter the composition of workers in the labor market by either exacerbating or mitigating wage inequality within and across occupations. Accordingly, previous literature has attempted to evaluate the effects on labor market outcomes in other, similar technologies (Autor et al., 2003; Acemoglu and Restrepo, 2020). The direct effects of generative AI remain yet unexplored and current research offers predictions on the future implications of generative AI on various tasks and occupations (Eloundou et al., 2023; Felten et al., 2023).

In this paper, we conduct an empirical investigation of the short-term effects of the introduction of generative AI on labor market outcomes. Our empirical analysis focuses on the introduction of ChatGPT in November 2022. The data are obtained from a large online labor market, Upwork, which matches freelancers with short-term projects. Due to the flexibility of this spot market compared to traditional, formal employment, it is a great setting to explore the short-term effects of ChatGPT. We begin by obtaining publicly available data on the full employment histories of freelancers on the platform. We then use a difference-in-differences research design to study the differential change in employment outcomes of freelancers in more affected occupations compared to freelancers in less affected occupations, following the release of ChatGPT. In particular, given results from previous research (Eloundou et al., 2023), anecdotal evidence, and the fact that ChatGPT is a large language model (LLM) specifically trained to predict and generate text, we focus on writing-related services as the main affected occupations.

We find that ChatGPT has a substantial adverse affected workers' employment outcomes. Freelancers in more affected occupations experienced a decrease of 2% in the number of monthly jobs and a decrease of 5.2% in monthly earnings on the platform, following the release of ChatGPT, compared to freelancers in less-affected occupations. These effects represent an economically meaningful and statistically significant (at 1% level) reduction in employment on the platform. We observe adverse effects on both the intensive and extensive margins: freelancers are 1.2% less likely to receive any employment in a given month and take 4.7% less jobs, conditional on employment.

We test the robustness of our findings to several alternative specifications. In particular, to assess both the generalizability as well as the validity of our design, we conduct an additional analysis in which we estimate the effect of the release of another generative AI at a different time and on another set of workers. Specifically, we study the effect of the April 2022 release of image-based LLMs, such as Dall-E 2 and Midjourney, on freelancers offering image- and design-related occupations. Reassuringly, we observe qualitatively similar effects of the release of image-based AI models on the performance of freelancers on the platform, in terms of both the number of jobs and total earnings.

Next, we study whether the quality of a freelancer mediates the effect on employment outcomes. We use several measures to determine freelancers' quality, including their past employment and earnings, the skill level necessitated by past jobs, past performance, and hourly rate. Across the board, we do not find evidence that being high-quality moderates the adverse effects of ChatGPT on a freelancer's employment. In fact, evidence are consistent with high-quality workers being disproportionately affected by the release of generative AI tools, especially in the image-focused occupations.

We interpret these findings as suggesting that generative AI models act as a substitute for knowledge workers of all quality types, at least in the short term, effectively reducing their employment and earnings. The heterogeneous effects are consistent with generative AI potentially reducing the productivity gap between low-quality and high-quality workers. We conclude by discussing the broader implications of these preliminary results on the effect of AI on the labor market.

1.1 Related Literature

Our results contribute to the growing literature on AI and labor markets. One common theme in the literature is that the effect of AI on the labor market is theoretically ambiguous: while AI can serve as a substitute for labor, it also has the potential to complement workers by increasing labor productivity

(Autor et al., 2003; Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2018; Brynjolfsson et al., 2018a; Agrawal et al., 2019a; Acemoglu and Restrepo, 2020).

One of the most recent and influential developments is the release of new generative AI models and LLMs. Several papers have aimed to quantify their effects on the labor market, particularly a recent set of papers (Eloundou et al., 2023; Felten et al., 2023) that attempts to predict which industries will be most affected by LLMs in the future by using a measure of compatibility between the required tasks in an industry and the AI capabilities—a methodology first seen in Felten et al. (2018); Brynjolfsson et al. (2018a) and Webb (2019). Another stream of papers conducts field experiments randomizing access to generative AI tools and studies the effect on worker performance (Noy and Zhang, 2023; Brynjolfsson et al., 2023; van Inwegen et al., 2023; Peng et al., 2023). Lastly, Yilmaz et al. (2023) find that the introduction of Google's machine translation reduces translators' employment, especially for tasks with analytical elements. Perhaps most similarly to this work, the paper also finds that the introduction of ChatGPT reduced the number of questions and answers on Stack Exchange. Our paper contributes to this strand of literature by studying the short-run effect of ChatGPT on labor market outcomes, such as worker employment and income. Our granular worker-level data allows us to study how the effects differ by workers quality-level within a given occupation.

More broadly, our results also contribute to the literature examining the effect of AI technologies on the economy. Past work has examined the consequences of deploying AI in areas such as machine translation (Brynjolfsson et al., 2019), customer service (Luo et al., 2019; Schanke et al., 2021), sales (Cao and Zhang, 2020; Luo et al., 2021), human resources (Tambe et al., 2019), online retail (Overby et al., 2010; Sun et al., 2019; Zhang et al., 2021), international trade (Goldfarb and Trefler, 2018), data market (Tucker, 2018; Jin, 2018), and programming (Cowgill et al., 2020; Cheng et al., 2022). As a general-purpose technology, AI has the potential to drastically change productivity (Brynjolfsson et al., 2018b; Agrawal et al., 2019b) as well as innovation (Cockburn et al., 2018) processes.

2 Setting

ChatGPT, released by OpenAI, is a powerful LLM that dramatically improves the performance of its predecessor, GPT-2. LLMs are a class of AI models that have a large number of parameters (175 billion for ChatGPT) and are trained on large datasets of text. The trained data and flexibility of the models enables the learning of the statistical relationships between words and phrases, which in turn allows the model to generate text that is both natural and informative. Immediate use cases of GPT include content writing, copyediting, answering questions, and language translation. Besides its capability, ChatGPT is also free to use and accessible to the public. Perhaps unsurprisingly, ChatGPT's user base grew rapidly, reaching 100 million since its November 2022 launch. At the time of writing this paper, LLMs are still in a phase of rapid development, with recent developments such as GPT-4 and Bard.

We study the effect of ChatGPT on labor market outcomes in the context of Upwork, one of the largest online labor markets in the world. The platform matches employers with independent freelancers for small to medium size tasks. The services on Upwork are typically remote jobs, ranging from data entry and graphic design to software development and marketing (Horton, 2010). On Upwork, a typical workflow starts with the creation of a job posting by a would-be employer (buyer). The job posting includes a description of the job, the category of the service (e.g., writing or administrative support), the required skills or qualifications for the job, and the expected outputs and timeline. Once a job is posted, workers may apply to the job by submitting a proposal, or may be directly invited by the buyer (Barach and Horton, 2021).

Upwork is a good setting to study the short-term effects of generative AI on labor outcomes. First, we are able to obtain the complete work histories of freelancers on the platform and to observe relevant freelancer attributes (or control for time invariant unobserved differences). Second, becasue this is a relatively-short-term, spot labor market, and employers are able renegotiate and rehire frequently, allowing for more flexibility compared to traditional, formal employment. Thus, while LLMs are still too young to replace full-time jobs, we may already be able to detect its effect in an online labor market.[1]

[1] A potential limitation is that we do not observe off-platform employment, and thus cannot say whether reduction is employment on the platform represents reduction in total employment, or is merely substituting other forms of employment. However, Horton (2017) provides evidence that multi-homing is rare and that online and offline hiring are not very substitutable. In addition, we are unable to evaluate whether high-skills Upwork workers are also highly skilled in the general population.

3 Empirical Framework

3.1 Data

Our primary data set is obtained from Upwork's freelancers API. We scraped all available information on all searchable freelancers' profile pages,[2] including their occupations, qualifications, and skills. We also observe metrics on past employment, such as total past earnings, number of previous jobs, success rate of completed jobs, and the reviews left by previous buyers. Notably, we are only able to observe a snapshot of the freelancer pool as it was in April 2023, and are unable to observe changes made in the profile page over time.[3]

We restrict our attention to the period from January 2022 through April 2023. Employment outcomes are aggregated to the freelancer-month level by the date at which the job started, focusing on monthly number of jobs and monthly income.[4] We winsorize both outcomes at the 1% and 99% levels. Because our research design uses variation across occupations, we restrict our attention to freelancers offering service in a single occupation (approximately 85% of freelancers). Our final data set consists of 92,547 freelancers. Table 1 presents the summary statistics for the variables used in the main analysis. We note that, on average, a freelancer starts a job once every three months, for an average monthly pay of $171. There exists, however, substantial variation across both months and freelancers.

Our secondary data set, which is used to provide descriptive motivation for our research design, consists of data obtained from Google Trends. The API provides relative interest in specific search terms on Google. For each one of the search terms, we obtain the change in interest in the United States over the duration of our sample period. Notably, Google only allows a comparison across five search terms at a time and always normalizes the results relative to highest value. Thus, when comparing across many search terms, we ran multiple queries, always including the highest value search term.

[2] We get data on freelancers that are available through the API. A subset of freelancers are not "searchable", and thus do not appear in our data. Unfortunately, we do not have an estimate of the share of non-searchable freelancers. In addition, in order to reduce bias from fraudulent jobs, which are generated to improve a freelancer's employment history, we omit freelancers whose first job's income is less than 5% or more than 500% the average of their first five jobs, which is approximately the 10th and 90th percentiles. We discuss our data sample restrictions in detail in Appendix B.

[3] For example, our sample only includes freelancers with active profiles at the time of obtaining the data, as we do not observe terminated accounts. Similarly, we are unable to observe changes made to freelancers' cover letters or qualifications over time.

[4] We also observe the total number of hours for a subset of jobs. This measure, however, is poorly populated and is only available for per-hour jobs.

3.2 Empirical Strategy

The release of ChatGPT3 on November 30, 2022[5] provides a shock to both the salience and accessibility of generative AI and LLMs. Our primary goal is to estimate the direct short-term effect of this new technology on employment outcomes. Naturally, the release of ChatGPT has the potential to disrupt multiple industries and affect a large share of the workforce in the long term, similar to what has been observed for pre-GPT AI technologies (Acemoglu et al., 2022). Focusing on the short-, and even immediate-term effects of the release of generative AI, we focus on occupations that are most prominently affected by the current rather than future capabilities of the model. One such susceptible type of industry are writing-related occupations, such as content writing, editing, and proofreading.

As an LLM, dialogue-based AI, GPT is specifically trained using large amount of text to predict and generate text. It particularly excels in understanding and generating human-like responses to a wide range of queries. Though similar to other general-purpose technologies, ChatGPT has the potential to transform multiple industries in the long term (Helpman and Trajtenberg, 1994; Eloundou et al., 2023), the tangibility of tasks in the writing industry allows for a straightforward application of GPT's text-generation capabilities even in the the short term. For example, users who are looking for copy-editing services can easily copy and paste paragraphs into the GPT prompt and immediately evaluate its output. As stated by the developers of ChatGPT, OpenAI, one of their main goals "*is to improve their ability to understand and generate natural language text, particularly in more complex and nuanced scenarios*" (OpenAI, 2023).

Previous research has offered strong evidence that writing-related occupations are among the most affected. According to Eloundou et al. (2023), writing is consistently ranked among the top five occupations most exposed to GPT, across different measurements based on models and human experts.[6] Finally, the relevance of ChatGPT to writing-related tasks is often highlighted in popular media, which is perhaps more accessible to casual ChatGPT users.[7]

[5] https://openai.com/blog/chatgpt

[6] Specifically, measurements are based on either human experts' opinions or GPT-4's ratings, and they measure how exposed an occupation is to either GPT or GPT-related software. More detailed results can be found in Table 4 of Eloundou et al. (2023) at https://arxiv.org/pdf/2303.10130.pdf, and the measurements are "Human α", "Human β","Human ξ", "Model α", "Model β","Model ξ". (Accessed 7/4/2023).

[7] For instance, https://www.businessinsider.com/chatgpt-jobs-at-risk-replacement-artificial-intelligence-ai-labor-trends-2023-02#media-jobs-advertising-content-creation-technical-writing-journalism-2 and https://www.nytimes.com/2022/12/06/opinion/chatgpt-ai-skilled-jobs-automation.html.

Finally, in order to examine more directly the differential interest in ChatGPT by application type, we take advantage of Google Trends data. As presented in Panel A of Figure 1, we begin by examining the relative interest in the search terms "ChatGPT," "GPT," and "AI" over time. Consistent with the rapid growth in ChatGPT's user base, we document a sharp increase in the relative interest in ChatGPT-related queries following its launch on November 2022.

We then examine the relative interest in various occupational domains combined with GPT, as is captured by Google search queries. We choose domains as they appear on the Upwork platform, and combine them with the term GPT, e.g., "GPT Translation," "GPT Mobile Development," etc. The results are presented in Panel B, which provides descriptive evidence to support our research design and the choice of writing-related jobs. "GPT Writing" is by far the most commonly search term compared to other domains, such as "GPT translation" or "GPT software development." This suggests that the public's attention and interest are skewed towards the implementation of AI technologies in the writing industry. We interpret these patterns as suggestive evidence, consistent with the notion that while multiple industries are affected by the introduction of ChatGPT, the short-term implications are much more pronounced in the writing-related tasks.

3.3 Research Design

The main specification uses the panel structure of the data to employ a difference-in-differences research design. In particular, we estimate the following specification:

$$Y_{it} = \beta \text{Post}_t \times \text{T}_i + X_{it} + \alpha_i + \delta_t + \varepsilon_{it} \tag{1}$$

where i and t are indices for freelancer and month, respectively, *Post* is an indicator for post November 2022, and T denotes an occupations that were potentially more affected by ChatGPT, namely writing, proofreading, and copy-editing. X_{it} is a vector of time-varying freelancer-level characteristics, such as the average feedback score (and an indicator for whether a score is missing), logged past monthly income, and number of previous jobs completed.[8] Y denotes our main outcomes of interest, monthly

[8]As presented in Appendix Table A1, we experiment with alternative control variables and clustering. Across the board, the results remain qualitatively similar and statistically significant.

number of jobs and income, which we also decompose into the extensive and intensive margins. In particular, we first estimate the total effect on (logged) monthly number of jobs and income; second, on the intensive margin, as measured by the logged number of jobs and income conditional on working; and third, on the extensive margin, as measured by indicator for whether the freelancer accepted a new job at a given month. All regressions also include freelancer and month fixed effects. Finally, we cluster the standard errors for the error term ε_{it} at the occupation-level, corresponding to the level at which treatment is assigned (Bertrand et al., 2004; Abadie et al., 2023).

The key identifying assumption of the difference-in-differences design is that freelancers operating in more affected and less affected occupations would have evolved similarly, absent the release of GPT. To evaluate this assumption, we estimate a more flexible difference-in-differences event-study specification of the following form:

$$Y_{it} = \sum_{j=T_0}^{-1} \beta_j Pre_j \times T_i + \sum_{k=1}^{T_1} \beta_k Post_k \times T_i + X_{it} + \alpha_i + \delta_t + \varepsilon_{it} \tag{2}$$

where Pre_j and $Post_k$ are dummy variables equal to 1 when an observation is j months before or k months after the addition of the label.

4 Results

4.1 Main Results

We begin by estimating the effect of the release of ChatGPT on the employment and compensation of freelancers in more affected occupations, as described in Equation 1. Our main results are presented in Table 2. Across specifications, we find significant negative effects of the release of ChatGPT on freelancer employment on the platform. Columns (1) and (2) present the total effect on monthly employment and earnings of freelancers in writing-related occupations, compared to freelancers in less affected occupations. Following the release, the monthly number of jobs on the platform for freelancers in more affected occupations decreases by 2% (s.e.=0.004), and total monthly compensation decreases by 5.2% (s.e.=0.016).

In Columns (3) through (5), we decompose the total effect into the extensive and intensive margins. Column (3) presents the effect on the linear probability that a freelancer receives at least one job in a given month, which we interpret as the extensive margin of working on the platform. Following the release of GPT, the probability that a freelancer offering writing-related services starts a new job in a given month decreases by 1.2 percentage points, which is approximately a 10% drop compared to baseline employment. In Columns (4) and (5), we estimate the effect on the (log) number of jobs and monthly income, conditional on receiving at least one job in that month. The number of jobs decreases by approximately 4.7% (s.e.=0.011), which is substantially larger than the total effect, and the effect of income is approximately the same size, a decrease of 5.1%, though it is much noisier (s.e.=0.037).

To alleviate concerns regarding unobserved secular trends in freelancers' employment driving the main results, in Figure 2 we present the event time for our main specification, as detailed in Equation 2. Reassuringly, we observe similar pre-trends between freelancers in eventually treated and untreated occupations prior to the release of ChatGPT, which provides suggestive evidence in favor of the parallel trends assumption. Following its release, we see a persistent and growing decrease in both the monthly number of jobs (Panel A) and monthly compensation (Panel B) on the platform, suggesting a causal relationship.

We test the robustness of our results to multiple alternative specifications, such as omitting our control variables Also, due to the small number of clusters, we also use wild bootstrap Cameron et al. (2008) to calculate our standard errors. Alternatively, we cluster at the industry rather than occupation level. The results are presented in Appendix Table A1. A related concern is that there exist unobserved differences between freelancers in affected and unaffected occupations that are driving the main results, even without the introduction of generative AI. We thus conduct an additional analysis in which we match freelancers based on pre-release characteristics to account for observable, and potentially unobservable, differences,[9] matching coarsely on the freelancer's past employment, past income, education, badges, success rate, and hourly rate on the platform. As presented in Appendix Table A2, using the matched sample improves the precision of, and generally strengthens, the main results.

[9]Specifically, we use a coarsened exact matching (CEM) algorithm, where we match coarsely on the freelancer's past employment, past income, education, badges, success rate, and hourly rate on the platform.

Finally, in order to both evaluate the internal validity of our design and assess the generalizability of our findings, we conduct an additional analysis in which we estimate the effect of the release of another generative AI, at a different time, and on another set of workers. In particular we study the effect of the release of DALL-E 2 and Midjourney, two models designed to generate images from natural language descriptions. Similar to our main specification, we define the affected group as the subset of freelancers who offer design, image editing, and art services.[10] Both models were released at about the same time, with DALL-E 2 released in April 2022, and Midjourney in July. To be conservative, we code the post period as post-April 2022. The results are presented in Appendix Table A3, which has the same structure as the main analysis in Table 2. The estimate of the effect of image-focused generative AI echo our main results: we find consistent negative effects of the release of the new technology on the performance of freelancers on the platform, in terms of both number of jobs and total compensation.

4.2 Heterogeneous Treatment Effects

Having documented robust evidence of adverse effects of the introduction of a LLM on employment outcomes, we turn to explore the heterogeneity in treatment effects in order to shed light on the drivers of the main estimates. We are particularly interested in understanding whether freelancer's quality or experience can mitigate (or potentially exacerbate) the effects on employment outcomes. Our granular, freelancer-level data allow us to examine the heterogeneity in treatment effects by the attributes of workers performing similar jobs and tasks.

Previous work suggests that high-quality suppliers may be generally less threatened by adverse market shocks (e.g., Syverson, 2004; Reshef, 2023). If that is the case, then AI may only hinder the performance of relatively low-quality freelancers. On the other hand, several papers also find that similar technology changes may differentially benefit low-skilled workers (Autor and Dorn, 2013; Noy and Zhang, 2023). Understanding the treatment effect heterogeneity by worker type can thus shed light on the nature of the technological shock and how pervasive are the implications of AI to the economy.

To this end, we interact our main estimator, $Post \times T$, with several measures of freelancers' past

[10] Another advantage of this analysis is that we have a well-defined treatment group, as the control groups are not directly affected the release of DALL-E 2 and Midjourney.

performance: the number of past jobs on the platform, the total past income on the platform, and skill level necessitated by past jobs.[11] We define these heterogeneity measures based on their November 2022 levels, right before the new LLM was launched. In addition, we also test for heterogeneity by the freelancers' attributes that are most salient on their freelancer page: whether they received a "top rated" badge from Upwork, their past success rate (as detailed by past employers), and their stated hourly rate.[12] Notably, we only observe these three measures at the time of collecting the data and are unable to assess whether those changed over our sample period. For ease of interpretation, we create an indicator for above- or below-median values for each one of our six of our heterogeneity measures.[13]

The results are presented in Table 3. Panel A presents the heterogeneous treatment effects on the monthly number of jobs, and Panel B present the effects on monthly income. In general, across the various measures of freelancers' quality, we can generally rule out that freelancers' quality moderates the adverse employment effects of generative AI. We mostly observe negative point estimates on the interaction terms, suggesting that, "top" freelancers are actually the ones most adversely affected by AI. These effect, however, are often not statistically significant, especially for monthly income, which tends to be substantially noisier. As presented in Appendix Table A4, we observe more negative and statistically significant effects when estimating the heterogeneous effects on image-based AI models. For example, observing Column 2 we find that freelancers with above median prior earnings suffer an additional reduction of 1.2% in the number of jobs (s.e.=0.007) and a 2.9% reduction in monthly income, though the result is not statistically significant. For image-based services, we estimate reductions of 7% and 14% for freelancers with above median prior earnings, both significant at the 1% level. One potential explanation for the stronger results on image-based LLM is the longer post-period time window, which may lead to higher adoption rates as well as technological improvements.

We interpret the heterogeneous effects by worker types through the lenses of the canonical skill-biased technological change (SBTC) model presented in Card and DiNardo (2002). Assuming that the labor

[11] For each job posting, the buyer must indicate the desired level of freelancer's experience: entry, intermediate, and expert, coded as 1,2, and 3, respectively. For each worker we calculate the average desired experience level for all the jobs performed on the platform prior to the introduction of ChatGPT. We interpret this measure as a proxy for a freelancer's skill level.

[12] In Appendix Table A5 we explore heterogeneity by several additional measures of freelancer quality: namely their average rating on the platform, based on past employers feedback; tenure, as measured by the number of years since their first job on the platform; and whether they obtained college education. Similar to the main dimensions of heterogeneity, we discretize all three measures.

[13] The results are qualitatively robust to using the continuous values, as presented in Appendix Table A6.

supply on Upwork does not change dramatically over the short run, the relative effect on both labor types would translate to changes in relative marginal product after the introduction of ChatGPT.[14] If one were to interpret the results in Tables 3 and A4 as suggestive evidence that high-quality workers are disproportionately hurt, this interpretation would be consistent with the growing experimental evidence that the adoption of LLM differentially benefits low-ability workers relative to high-ability ones (Noy and Zhang, 2023; Brynjolfsson et al., 2023; Peng et al., 2023).

5 Conclusion

This paper studies the short-term effects of generative AI and LLMs on labor outcomes by estimating the effect of ChatGPT on the employment of workers in a large online labor market. Across the board, we find that freelancers who offer services in occupations most affected by AI experienced reductions in both employment and earnings. The release of ChatGPT leads to a 2% drop in the number of jobs on the platform, and a 5.2% drop in monthly earnings. The results are robust to several alternative tests, including a similar reduction in the employment outcomes of freelancers offering design and image-editing services following the introduction of image-focused generative AI. In addition, we find that offering high-quality service does not mitigate the negative effect of AI on freelancers, and in fact present suggestive evidence that top employees are disproportionately hurt by AI.

The results have several implications to policymakers and business leaders. Aside from the many benefits of AI technology, it may also have substantial economic and societal ramifications. As the usage and development of generative AI continues to grow, there is thus room for additional scrutiny of the pervasive effects of the technology on various industries and the economy as a whole. Importantly, we find that top performance and high-quality service do not help mitigate or harness the introduction of generative AI. Businesses leaders must carefully examine how to adapt and whether to adopt these technologies, as AI threats to indiscriminately disrupt incumbents and erode their competitive advantage.

Notably, in this paper we provide novel, preliminary evidence on the short-term effects of generative AI. However, the long-term implications may be significantly different, and it is unclear how our findings extend to longer time horizons. On the one hand, the negative effects on labor outcomes

[14] Although our results, technically speaking, concern changes in the relative earnings rather than relative wages, the qualitative insight is similar.

may be exacerbated as the technology becomes more prevalent and penetrates additional industries and occupations. In contrast, as the development of AI capabilities continues it may become more integrated with various tools and begin to better complement the performance of human workers. Finally, our paper focuses entirely on the (negative) effects on workers. Assessing the total welfare implications to the introduction of generative AI technology to all stakeholders is beyond the scope of this paper and remains a promising direction for future work.

References

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2023): "When should you adjust standard errors for clustering?" *The Quarterly Journal of Economics*, 138, 1–35.

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

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

Acemoglu, D. and P. Restrepo (2018): "Artificial intelligence, automation, and work," in *The economics of artificial intelligence: An agenda*, University of Chicago Press, 197–236.

——— (2020): "Robots and jobs: Evidence from US labor markets," *Journal of political economy*, 128, 2188–2244.

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

——— (2019b): "Exploring the impact of artificial intelligence: Prediction versus judgment," *Information Economics and Policy*, 47, 1–6.

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

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

Barach, M. A. and J. J. Horton (2021): "How do employers use compensation history? Evidence from a field experiment," *Journal of Labor Economics*, 39, 193–218.

Bertrand, M., E. Duflo, and S. Mullainathan (2004): "How much should we trust differences-in-differences estimates?" *The Quarterly journal of economics*, 119, 249–275.

Brynjolfsson, E., X. Hui, and M. Liu (2019): "Does machine translation affect international trade? Evidence from a large digital platform," *Management Science*, 65, 5449–5460.

Brynjolfsson, E., D. Li, and L. R. Raymond (2023): "Generative AI at work," Tech. rep., National Bureau of Economic Research.

Brynjolfsson, E. and T. Mitchell (2017): "What can machine learning do? Workforce implications," *Science*, 358, 1530–1534.

Brynjolfsson, E., T. Mitchell, and D. Rock (2018a): "What can machines learn and what does it mean for occupations and the economy?" in *AEA papers and proceedings*, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, vol. 108, 43–47.

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

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008): "Bootstrap-based improvements for inference with clustered errors," *The review of economics and statistics*, 90, 414–427.

Cao, X. and D. Zhang (2020): "The impact of forced intervention on AI adoption," *Available at SSRN 3640862*.

Card, D. and J. E. DiNardo (2002): "Skill-biased technological change and rising wage inequality: Some problems and puzzles," *Journal of labor economics*, 20, 733–783.

Cheng, Z., D. Lee, and P. Tambe (2022): "Innovae: Generative ai for understanding patents and innovation," *Available at SSRN 3868599*.

Cockburn, I. M., R. Henderson, and S. Stern (2018): "The impact of artificial intelligence on innovation: An exploratory analysis," in *The economics of artificial intelligence: An agenda*,

University of Chicago Press, 115–146.
COWGILL, B., F. DELL'ACQUA, S. DENG, D. HSU, N. VERMA, AND A. CHAINTREAU (2020): "Biased programmers? Or biased data? A field experiment in operationalizing AI ethics," in *Proceedings of the 21st ACM Conference on Economics and Computation*, 679–681.
ELOUNDOU, T., S. MANNING, P. MISHKIN, AND D. ROCK (2023): "Gpts are gpts: An early look at the labor market impact potential of large language models," *arXiv preprint arXiv:2303.10130*.
FELTEN, E. W., M. RAJ, AND R. SEAMANS (2018): "A method to link advances in artificial intelligence to occupational abilities," in *AEA Papers and Proceedings*, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, vol. 108, 54–57.
——— (2023): "Occupational heterogeneity in exposure to generative ai," *Available at SSRN 4414065*.
GOLDFARB, A. AND D. TREFLER (2018): "Artificial intelligence and international trade," in *The economics of artificial intelligence: an agenda*, University of Chicago Press, 463–492.
HELPMAN, E. AND M. TRAJTENBERG (1994): "A time to sow and a time to reap: growth based on general purpose technologies," .
HORTON, J. J. (2010): "Online labor markets," in *Internet and Network Economics: 6th International Workshop, WINE 2010, Stanford, CA, USA, December 13-17, 2010. Proceedings 6*, Springer, 515–522.
——— (2017): "Price floors and employer preferences: Evidence from a minimum wage experiment," *Available at SSRN 2898827*.
JIN, G. Z. (2018): "Artificial intelligence and consumer privacy," in *The economics of artificial intelligence: An agenda*, University of Chicago Press, 439–462.
LUO, X., M. S. QIN, Z. FANG, AND Z. QU (2021): "Artificial intelligence coaches for sales agents: Caveats and solutions," *Journal of Marketing*, 85, 14–32.
LUO, X., S. TONG, Z. FANG, AND Z. QU (2019): "Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases," *Marketing Science*, 38, 937–947.
NOY, S. AND W. ZHANG (2023): "Experimental evidence on the productivity effects of generative artificial intelligence," *Available at SSRN 4375283*.
OPENAI (2023): "Gpt-4 technical report," *Technical report*.
OVERBY, E., S. A. SLAUGHTER, AND B. KONSYNSKI (2010): "Research commentary—the design, use, and consequences of virtual processes," *Information Systems Research*, 21, 700–710.
PENG, S., E. KALLIAMVAKOU, P. CIHON, AND M. DEMIRER (2023): "The impact of ai on developer productivity: Evidence from github copilot," *arXiv preprint arXiv:2302.06590*.
RESHEF, O. (2023): "Smaller slices of a growing pie: The effects of entry in platform markets," *American Economic Journal: Microeconomics, Forthcoming*.
SCHANKE, S., G. BURTCH, AND G. RAY (2021): "Estimating the impact of "humanizing" customer service chatbots," *Information Systems Research*, 32, 736–751.
SUN, C., Z. J. SHI, X. LIU, A. GHOSE, X. LI, AND F. XIONG (2019): "The effect of voice ai on consumer purchase and search behavior," *NYU Stern School of Business*.
SYVERSON, C. (2004): "Market structure and productivity: A concrete example," *Journal of political Economy*, 112, 1181–1222.
TAMBE, P., P. CAPPELLI, AND V. YAKUBOVICH (2019): "Artificial intelligence in human resources management: Challenges and a path forward," *California Management Review*, 61, 15–42.
TUCKER, C. (2018): "Privacy, algorithms, and artificial intelligence," in *The economics of artificial intelligence: An agenda*, University of Chicago Press, 423–437.
VAN INWEGEN, E., Z. T. MUNYIKWA, AND J. J. HORTON (2023): "Algorithmic writing assistance

on jobseekers' resumes increases hires," Tech. rep., National Bureau of Economic Research.

WEBB, M. (2019): "The impact of artificial intelligence on the labor market," *Available at SSRN 3482150*.

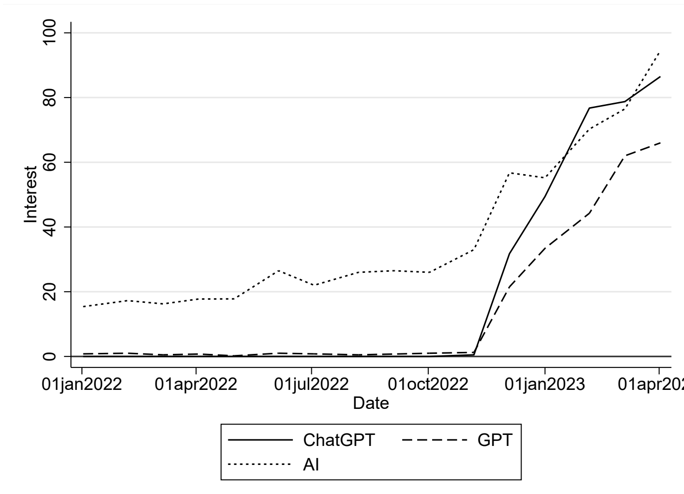
YILMAZ, E. D., I. NAUMOVSKA, AND V. A. AGGARWAL (2023): "AI-Driven Labor Substitution: Evidence from Google Translate and ChatGPT," *Available at SSRN*.

ZHANG, M., T. SUN, L. LUO, AND J. GOLDEN (2021): "Consumer and AI Co-creation: When and Why Human Participation Improves AI Creation." *USC Marshall School of Business Research Paper Sponsored by iORB, No. Forthcoming*.

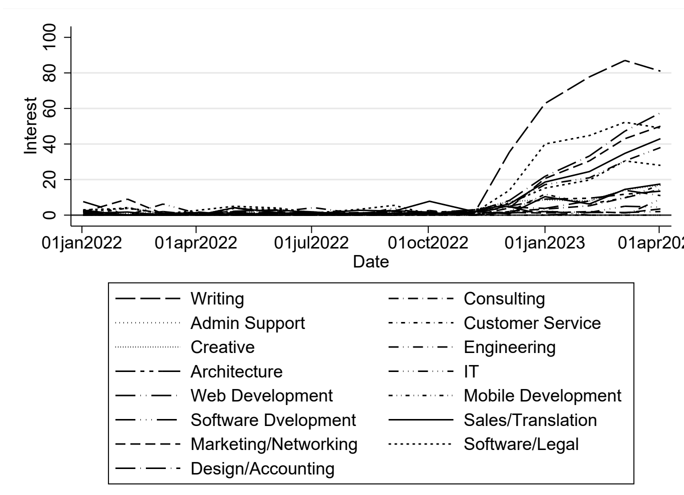
Figures

Figure 1: Descriptive Evidence on Interest in ChatGPT Over Time and Across Types of Occupations

Panel A: Relative Interest in ChatGPT and AI



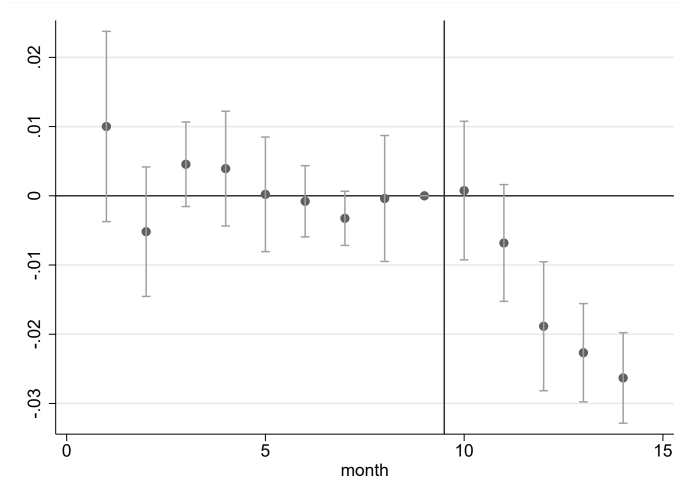
Panel B: Relative Interest in ChatGPT by Types of Occupation



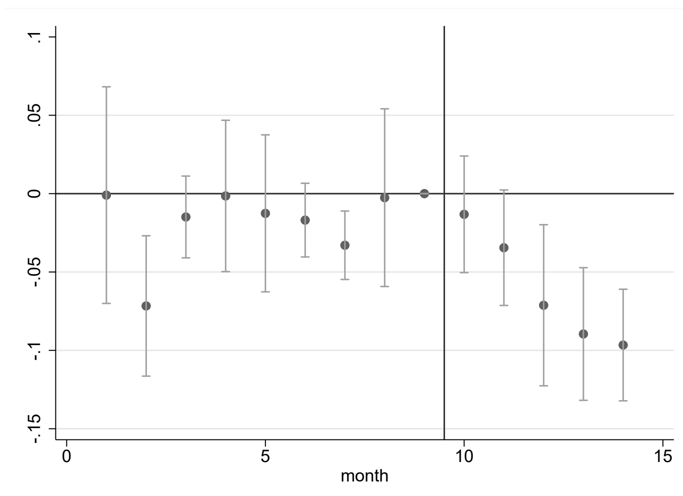
Notes: This figure displays the relative interest in separate search terms on Google Trends. Panel A presents relative interest in three separate search terms. Panel B presents the relative interest over time in nineteen separate search term, where each term represents a domain on the platform and was search together with the word "GPT." For example, "Writing" refers to interest in the search term "GPT Writing."

Figure 2: The Effect of generative AI on Freelancers' Employment on the Platform

Panel A: Monthly Number of Jobs



Panel B: Monthly Revenue



Notes: The figure displays results for monthly jobs and revenue, where the estimate between treatment and control occupations is allowed to vary for each month around the launch of ChatGPT (see description around Equation 2). Each panel also report 95% confidence intervals. Standard errors are clustered at the occupation-level.

Tables

Table 1: Descriptive Statistics

	Mean	SD	P5	P95
No. of Freelancers	92,457	-	-	-
Tenure	3.04	2.98	0	10
US-based	0.16	0.37	0	1
Badge	0.33	0.47	0	1
Success Rate	0.62	0.46	0	1
Hourly Rate	27.9	28.9	5	75
Worked in Month	0.18	0.39	0	1
Monthly Number of Jobs	0.36	1.35	0	2
Monthly Earnings	171.2	1548.5	0	550
Mean Monthly Score	4.82	0.59	4.0	5
Job Skill level	2.12	0.61	1	3

Notes: This table provides the summary statistics of the variables used in the main analysis.

Table 2: The Effect of generative AI on Freelancers' Employment on the Platform

	Total		Extensive Margin	Intensive Margin	
	(1) Num. of Jobs	(2) Income	(3) Worked	(4) Num. of Jobs	(5) Income
Post × Treatment	-0.020***	-0.052***	-0.012***	-0.047***	-0.051
	(0.004)	(0.016)	(0.002)	(0.011)	(0.033)
Observations	1236052	1236052	1236052	171541	174697
# of Clusters	195	195	195	193	193

Notes: This table presents OLS regressions results relating freelancers' employment on the platform to the introduction of ChatGPT in a difference-in-differences design (see description around Equation 1). *Treatment* is an indicator of whether an occupation is substantially affected by the introduction of ChatGPT. The unit of observation is the freelancer-month. The dependent variables are the following: the total monthly number of jobs and income (Columns 1 & 2), an indicator for whether the freelancer had at least one job that month (Column 3), the monthly number of jobs and income conditional on working (Columns 4 & 5). In columns 1, 2, 4, and 5, the dependent variables are inverse hyperbolic sine transformed. All regressions include freelancer and month fixed effects. Standard errors are in parentheses and are clustered at the occupation-level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3: The Effect of generative AI on Freelancers' Employment on the Platform – Heterogeneity by Freelancer Attributes and Performance

	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Monthly Number of Jobs			
Post × T	-0.016***	-0.020***	-0.029***	-0.017***	-0.020***	-0.013**
	(0.004)	(0.004)	(0.007)	(0.003)	(0.003)	(0.005)
Post × T × Var.	-0.007	-0.012*	-0.019**	-0.024***	-0.002	-0.011*
	(0.006)	(0.007)	(0.009)	(0.007)	(0.005)	(0.006)
			Panel B: Monthly Income			
Post × T	-0.031**	-0.074***	-0.077***	-0.054***	-0.045***	-0.010
	(0.015)	(0.016)	(0.027)	(0.010)	(0.013)	(0.018)
Post × T × Var.	-0.036	-0.029	-0.015	-0.087***	-0.028	-0.051
	(0.029)	(0.032)	(0.035)	(0.033)	(0.024)	(0.031)
Observations	1236052	1236052	611375	1236052	1236052	1236052
# of Clusters	195	195	193	195	195	195
Heterogeneity By:	Past Jobs	Past Income	Skill Level	Badge	Success Rate	Hourly Rate

Notes: This table presents OLS regressions results relating freelancers' employment on the platform to the introduction of ChatGPT, examining heterogeneity by freelancers' attributes and past experiences. *Treatment* is an indicator for whether an occupation is substantially affected by the introduction of ChatGPT, and *Var.* is the heterogeneity variable of interest, as indicated below each column. The unit of observation is the freelancer-month. The dependent variable in Panel A is total monthly number of jobs and total monthly income in Panel B. Both dependent variables are inverse hyperbolic sine transformed. Heterogeneity variables are an indicator for above (below) median. All regressions include freelancer and month fixed effects. Standard errors are in parentheses and are clustered at the occupation-level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.